

DEPARTMENT OF COMPUTER SCIENCE
SWANSEA UNIVERSITY

Submitted to Swansea University in fulfilment for the degree

Doctor of Philosophy

Classification and Segmentation of Galactic Structures
in Large Multi-spectral Images

by

Felix Richards

Supervisor: **X. Xie**

Co-Supervisor: **A. Paiement**

June 2022

Copyright: The author, Felix Richards, 2023.
Released under the terms of a CC-BY-NC-ND
License. Third party content is excluded for use
under the license terms.

Abstract

Classification and Segmentation of Galactic Structures in Large Multi-spectral Images

Extensive and exhaustive cataloguing of astronomical objects is imperative for studies seeking to understand mechanisms which drive the universe. Such cataloguing tasks can be tedious, time consuming and demand a high level of domain specific knowledge. Past astronomical imaging surveys have been catalogued through mostly manual effort. Imminent imaging surveys, however, will produce a magnitude of data that cannot be feasibly processed through manual cataloguing. Furthermore, these surveys will capture objects fainter than the night sky, termed low surface brightness objects, and at unprecedented spatial resolution owing to advancements in astronomical imaging. In this thesis, we investigate the use of deep learning to automate cataloguing processes, such as detection, classification and segmentation of objects. A common theme throughout this work is the adaptation of machine learning methods to challenges specific to the domain of low surface brightness imaging.

We begin with creating an annotated dataset of structures in low surface brightness images. To facilitate supervised learning in neural networks, a dataset comprised of input and corresponding ground truth target labels is required. An online tool is presented, allowing astronomers to classify and draw over objects in large multi-spectral images. A dataset produced using the tool is then detailed, containing 227 low surface brightness images from the MATLAS survey and labels made by four annotators. We then present a method for synthesising images of galactic cirrus which appear similar to MATLAS images, allowing pretraining of neural networks.

A method for integrating sensitivity to orientation in convolutional neural networks is then presented. Objects in astronomical images can present in any given orientation, and thus the ability for neural networks to handle rotations is desirable. We modify convolutional filters with sets of Gabor filters with different orientations. These orientations are learned alongside network parameters during backpropagation, allowing exact optimal orientations to be captured. The method is validated extensively on multiple datasets and

use cases.

We propose an attention based neural network architecture to process global contaminants in large images. Performing analysis of low surface brightness images requires plenty of contextual information and local textual patterns. As a result, a network for processing low surface brightness images should ideally be able to accommodate large high resolution images without compromising on either local or global features. We utilise attention to capture long range dependencies, and propose an efficient attention operator which significantly reduces computational cost, allowing the input of large images. We also use Gabor filters to build an attention mechanism to better capture long range orientational patterns. These techniques are validated on the task of cirrus segmentation in MATLAS images, and cloud segmentation on the SWIMSEG database, where state of the art performance is achieved.

Following, cirrus segmentation in MATLAS images is further investigated, and a comprehensive study is performed on the task. We discuss challenges associated with cirrus segmentation and low surface brightness images in general, and present several techniques to accommodate them. A novel loss function is proposed to facilitate training of the segmentation model on probabilistic targets. Results are presented on the annotated MATLAS images, with extensive ablation studies and a final benchmark to test the limits of the detailed segmentation pipeline.

Finally, we develop a pipeline for multi-class segmentation of galactic structures and surrounding contaminants. Techniques of previous chapters are combined with a popular instance segmentation architecture to create a neural network capable of segmenting localised objects and extended amorphous regions. The process of data preparation for training instance segmentation models is thoroughly detailed. The method is tested on segmentation of five object classes in MATLAS images. We find that unifying the tasks of galactic structure segmentation and contaminant segmentation improves model performance in comparison to isolating each task.

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed:  19/11/2022
Felix Richards


This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed:  19/11/2022
Felix Richards

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed:  19/11/2022
Felix Richards

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed:  19/11/2022
Felix Richards

Contents

Abstract	ii
Declaration	iv
Contents	v
Acknowledgments	ix
List of Figures	x
List of Tables	xv
Abbreviations	xviii
1 Introduction	1
1.1 Contributions	3
1.1.1 Training data for automated cataloguing	4
1.1.2 Learnable complex-valued Gabor convolution for robustness to orientation	5
1.1.3 Multi-scale gridded Gabor attention network for segmentation of global contaminants	6
1.1.4 Segmentation of cirrus contamination with deep learning	6
1.1.5 Detection and segmentation of galactic structures in LSB images	7
1.2 Outline	8
2 Background	10
2.1 Galaxy Evolution and Low Surface Brightness	10
2.1.1 Low Surface Brightness Structures	11
2.1.2 Imaging Instrumentation and Observing Strategies	12
2.1.3 Cataloguing	14

2.2	Deep Learning for Image Segmentation	16
2.2.1	Convolutional Neural Networks	16
2.2.2	Image Segmentation	18
2.2.3	Improving Data Efficiency	19
2.2.4	Practical Tricks for Training Deep CNNs	20
2.3	Summary	22
3	Creating an LSB Dataset for Supervised Learning	23
3.1	Introduction	23
3.2	Annotation Tool for Large Multi-spectral Images	25
3.2.1	Annotation Process	26
3.2.2	Annotation Tool Considerations	28
3.2.2.1	Reducing uncertainty in annotations	28
3.2.2.2	World coordinate encoding	29
3.2.2.3	Viewing tool requirements	30
3.2.3	Interactive Viewing of Astronomical Images with Aladin Lite	31
3.2.4	Precise Delineation of Galactic Structures	32
3.2.4.1	Drawing tool	33
3.2.4.2	User Experience Features	36
3.2.4.3	Database design	39
3.3	Annotated Dataset of LSB Images	39
3.3.1	Annotated Dataset	40
3.3.1.1	Images	40
3.3.1.2	Annotations	41
3.3.2	Computing a Consensus	41
3.4	Synthesising Galactic Cirrus for Pretraining	43
3.4.1	Constructing Synthetic Cirrus Images	45
3.4.2	Increasing realness	48
3.5	Conclusion	49
4	Learnable Gabor Modulation in Complex-valued Neural Networks	51
4.1	Introduction	51
4.2	Related Work	53
4.3	Methodology	54
4.3.1	Analytical Modulation of Complex-Valued Networks	55
4.3.2	Learning Analytical Parameters through Backpropagation	57
4.3.3	Cyclic Gabor Convolutions	57
4.3.4	Learnable Gabor Convolutional Networks	59

4.3.4.1	Complex weight initialisation	59
4.3.4.2	Gabor axis considerations	59
4.3.4.3	Invariance vs equivariance	60
4.3.4.4	Projection between \mathbb{C} and \mathbb{R}	60
4.4	Experiments	60
4.4.1	Orientation invariance on MNIST	61
4.4.2	Invariance and equivariance to the dominant orientation of galactic cirri	64
4.4.2.1	Comparing LGCN against a traditional CNN on synthesised cirrus images	64
4.4.2.2	Prediction of cirrus structures in LSB images	67
4.4.3	Boundary Detection on Natural Images	68
4.5	Conclusion	69
5	Multiscale Gridded Gabor Attention for Segmenting Global Contaminants	70
5.1	Introduction	70
5.2	Extending Attention for Segmenting Large Contaminants	72
5.2.1	Multi-scale Gridded Attention	72
5.2.1.1	Background on attention mechanism	74
5.2.1.2	Gridded attention	76
5.2.2	Gabor Attention	80
5.2.2.1	Tri-attention module with orientation-wise attention operator	80
5.2.3	Constructing a Segmentation Model with Attention	81
5.3	Results	83
5.3.1	Segmentation of Cirrus Dust	83
5.3.2	Cloud segmentation in natural images	88
5.4	Summary	89
6	Segmentation of Cirrus Contamination: a Deep Learning Approach	91
6.1	Introduction	91
6.2	Related Work	94
6.3	Training on LSB Images	96
6.3.1	Data	96
6.3.2	Data augmentation and Transfer Learning	97
6.3.3	Adaptive Intensity Scaling	99
6.3.4	Loss function	100

6.4	Comparative Analyses on Proposed Techniques	101
6.4.1	Comparing Strategies Specific to Training on LSB images	102
6.4.2	Ablation Study on Model Modifications	106
6.5	Automated Cirrus Detection on LSB Images	108
6.5.1	Experiment setup	108
6.5.2	Results	110
6.6	Discussion	112
6.7	Summary	117
7	Multi-class Segmentation of Galactic Structures	118
7.1	Introduction	118
7.2	Related Work	119
7.3	Method	121
7.3.1	Mask R-CNN Overview	121
7.3.2	Implementation details	123
7.3.3	Cirrus Subnetwork	123
7.4	Data	124
7.4.1	Dataset	125
7.4.2	Obtaining instance masks	126
7.5	Results	127
7.5.1	Instance Segmentation with Mask R-CNN	128
7.5.2	Panoptic Results	133
7.6	Summary	137
8	Conclusion	139
8.1	Contributions	142
8.2	Future Work	143
	Bibliography	147
	A Appendices	169
A.1	Annotation Database Design	169
A.2	Annotation Tool User Management	170

Acknowledgments

A PhD is either hard and torturous, or you are well supported. Fortunately, I fell firmly in the latter. I am grateful for the guidance of my supervisors, Prof. Xianghua Xie and Dr. Adeline Paiement. They always made time for me, and their combined hard efforts made me smarter and more productive. I am grateful for the support of my collaborators in the MATLAS project, particularly Prof. Pierre-Alain Duc and Elisabeth Sola, who endured many long telecons with me and always provided thoughtful feedback. Thank you to my family, for making me believe I can do anything I set my mind to. Finally, I am grateful for the love of my partner, Laura, who brings so much joy to my life and more importantly looked after me while I wrote this thesis.

List of Figures

1.1	Examples of low surface brightness images.	2
2.1	<i>Left:</i> Diagram of Canada France Hawaii Telescope (CFHT) Megacam's main components, taken from [193]. Note the four spider arms in the top end. <i>Right:</i> Internal reflections cast by a bright star in a MATLAS image taken with CFHT's Megacam.	13
2.2	Example interpretation of a CCD frame by SExtractor, taken from [19]. . .	15
2.3	Visualations of 3×3 convolutions with different hyperparameter setups, taken from [65]. (a) shows a standard convolution; (b) shows a convolution with unit padding; (c) shows a convolution with dilation set to 2.	17
3.1	The annotation tool used to label contaminants and morphological features of NGC0448.	26
3.2	The Aladin Lite tool.	32
3.3	Aladin Lite integrated into the annotation tool.	33
3.4	Buttons used in the drawing tool: B1 activates viewing mode; B2-8 activate different drawing shape tools; B9-12 manage drawn shapes; B13 opens a table of drawn shapes; B14 shows examples of different features; B15 allows the classification of a drawn shape.	34
3.5	Examples of each basic shape in the drawing tool. Each shape is in a "selected" state so that the associated amendment boxes are visible. Note that amendment boxes do not always correspond with bounding boxes, due to how Bezier curves (which ellipses also use in rendering calculations) are calculated from user generated points.	35
3.6	Examples of each complex polygon shape in the drawing tool. Each shape is in a "selected" state so that the associated amendment boxes are visible.	37
3.7	Examples of each complex polygon shape in the drawing tool. Each shape is in a 'selected' state so that the associated amendment boxes are visible.	38

3.8	Entity relationship diagram of the database schema designed for storing annotation tool data.	39
3.9	Annotations of streams on NGC0474 by three users, u_1, u_3 and u_4	44
3.10	Annotations of tidal tails on NGC4270 by four users, u_1, u_2, u_3 and u_4	44
3.11	Examples components used to create synthesised images.	46
3.12	Components generated from a GMM used to shape the cirrus structure.	46
3.13	Example textures used to create the cirrus structure.	47
3.14	The final resulting output and its associated denoising target.	47
3.15	Synthesised and real LSB images arranged side by side for comparison. Note, in the real image, the ghosted halos and saturation trails surrounding stars, and horizontal band containing a comparatively high level of background noise.	48
3.16	Different options to introduce difficulty variations into synthesised LSB images. In this random iteration, the rotation angle of the cirrus was close to 180° , thus the fixed rotation example has a similar orientation but different appearance.	50
4.1	Overview of LGCN, with illustration of the filter modulation and cyclic convolution concepts. The displayed network is tasked with removing cirrus streaks and ghosted halos from the input image. Colour denotes individual feature orientations. Gabor modulation is applied with a range of angles to generate orientation dependent features. In the cyclic operator, each input orientational feature is exposed to every rotation of Gabor modulation, encouraging weight tying across orientation.	55
4.2	Effect of input rotation on MNIST classification accuracy (left) and magnitude of activations in the first modulated layer of the network (right), for different numbers of modulating filters and orientations, and on a subset of 1000 testing samples of MNIST.	62
4.3	Denoising and segmentation results on real and synthesised samples of galactic cirri generated with fixed rotation; randomised rotation; and randomised rotation with stars and telescope artefacts. These are difficult tasks as the striped textures of cirrus regions are easily confused/obstructed with bright diffuse regions and other objects.	65
5.1	Examples of cirrus contaminating different galaxies in various strengths, and their associated annotation. Images are taken from the r-band.	73

5.2	Diagram of the general attention module implemented in this work. Note that with positional attention, a convolutional layer is placed before each initial reshaping operation (green arrows). Matrix multiplication is denoted by \otimes , element-wise addition is denoted by \oplus	75
5.3	Diagram of the guided attention block implemented in this work. Element-wise multiplication is denoted by \odot	76
5.4	Diagram of the proposed gridded attention mechanism. Here number of scales $s = 3$, and common downscaling factor $f = 2$	78
5.5	Different strategies for generating multiscale features, with number of scales $s = 3$, downscale factor $f = 2$ and image size $N = 1024$. (a) Features are taken from intermediate layers; (b) downscaled copies of the input image are created and then each fed into the backbone separately.	79
5.6	Proposed Gabor attention operator. G is number of modulating Gabor filters, N is the product of other axes. Matrix multiplication is denoted by \otimes , and element-wise addition by \oplus . An intermediate correlation result is shown.	81
5.7	Diagram of dual attention versus tri-attention.	81
5.8	Fusion layer and segmentations generated from both attention maps and initial backbone features.	82
5.9	Sky/cloud examples from the SWIMSEG testing dataset (top), corresponding ground truth cloud segmentations (middle), and model predictions (bottom).	89
6.1	Surrounding region of NGC1253 captured in different astronomical imaging surveys.	92
6.2	Cirrus dust of various strengths (top), with uncertain annotations (middle) and predictions (bottom).	94
6.3	Mosaics of augmented images, and their corresponding augmented target masks, generated through our augmentation pipeline.	98
6.4	Examples of different learned intensity scaling transformations on NGC2592/4.103	
6.5	Histograms showing the proportion of actual or predicted cirrus across all testing LSB images. Predictions are taken from models trained with BCE and focal loss.	105
6.6	ROC curves for different consensus loss frameworks with BCE and focal loss functions, on only LSB images containing cirrus.	106
6.7	Comparison of segmentations generated with a patch based method versus a model that segments the entire image in one pass.	108

6.8	Training curves (smoothed) for the proposed model showing how IoU and Dice scores change over training epochs on the training and testing sets. We also fit a logarithmic curve to each plot to help illustrate the convergence trend.	110
6.9	Histogram showing the proportion of cirrus coverage across testing images.	111
6.10	Histograms showing the proportion of predicted cirrus across all testing LSB images, for different prediction techniques.	112
6.11	Typical LSB images with no cirrus contamination. The proposed model predicts zero false positives.	113
6.12	Segmentation predictions on difficult examples with no cirrus coverage. In this figure, we specifically chose examples with regions that present similarly to cirrus contamination, such as large regions of diffuse light in NGC5846 (first row), or large areas of high background levels in UGC03960 (fourth row).	114
6.13	Segmentation predictions on examples with high cirrus coverage. Columns three and four show predictions generated by a single model and an ensemble of models, respectively. Light grey in the prediction map, as in the third row, indicates where the model predicted an uncertain pixel as positive.	115
6.14	The four segmentation predictions with the lowest IoU scores across the testing set. Dark grey in the prediction map, such as in the third row, indicates where the model predicted an uncertain pixel as negative.	116
7.1	Mask R-CNN diagram, based on the first figure of [91].	122
7.2	Distributions of heights, widths and aspect ratios of all target objects in the annotation dataset.	124
7.3	Diagram of the proposed segmentation model, combining a gridded Gabor attention model with Mask R-CNN.	125
7.4	Instance labels for diffuse halos surrounding NGC4281, NGC4277 and NGC4273 before and after processing, with intermediate results.	128
7.5	Precision-recall curves for each object class and associated AP scores over different IoU thresholds. Horizontal axes show recall, vertical axes show precision.	130
7.6	Annotated and predicted objects in NGC6703. Numbers above network predictions represent confidence scores.	131
7.7	Annotated and predicted objects in NGC7710. Numbers above network predictions represent confidence scores. Dark blue belongs to the elongated tidal features class.	132

7.8	Flowchart outlining the implemented human-in-the-loop training protocol.	133
7.9	False positive objects predicted by Mask R-CNN at different epochs throughout the human-in-the-loop training run. Solid colours represent correct predictions that are added into the dataset, opaque colours are rejected predictions.	134
7.10	Precision-recall curves for all classes over different IoU thresholds. Horizontal axes show recall, vertical axes show precision, brighter colours denote higher IoU thresholds.	134
7.11	Comparison of target and predicted labels (PGC050395) on training runs with and without the human-in-the-loop protocol.	135
7.12	Precision-recall curves for each object class and associated AP scores over different IoU thresholds, with the proposed panoptic segmentation model. Horizontal axes show recall, vertical axes show precision.	136
A.1	Different interfaces for user credential verification and registration.	171
A.2	An interactive table showing the user's annotations.	172
A.3	Elements that allow the management of other users by a user of high enough privilege level.	172

List of Tables

3.1	A detailed overview of structures to be annotated in LSB images and associated astronomical definitions consistent with [185].	27
3.2	Detailed overview of basic shapes available to draw in the annotation tool. The interaction of the drawing process is described given two user generated points p and q . The number of vertices used to encode the shape is denoted as $ V $. Euclidean distance is represented as d . Buttons correspond to labels of Figure 3.4.	34
3.3	Detailed overview of polygon drawing tools. The method for calculating shape vertices V is described given a set of user generated points p_0, \dots, p_n . Buttons correspond to labels of Figure 3.4.	36
3.4	Summary of annotated features. The number of annotated features is denoted as d	41
4.1	Classification accuracy on randomly rotated MNIST images.	62
4.2	Classification accuracy on rotated MNIST averaged over 5 splits for different learning strategies of Gabor parameters wavelength λ and scale σ . Rows are divided in the centre to denote whether a single λ and σ is used for all U modulating Gabor filters, or λ and σ are separate for each modulating Gabor filter.	63
4.3	Segmentation IoU (left), denoising PSNR (middle) on synthesised cirri with fixed and randomised orientation, and with stars and telescope artefacts. Segmentation IoU (right) on real cirrus samples in LSB images. *Gabor convolutions of [140] applied to our base model.	66
4.4	Boundary detection results on the BSD500 [6] dataset. *Our parameter restricted implementation. †ImageNet pretrained.	69

5.1	Ablation study of modifications presented in [182]: fusing scales (see Fig. 5.8); generating multiscale features from intermediate layers; and guided attention. Results presented as mean IoU over 5 splits on real and synthesised cirrus samples.	85
5.2	Segmentation scores of different gridded attention models on synthesised and real data. Here s represents the number of scales and f denotes the downscaling factor. Results presented as mean IoU over 5 splits on real and synthesised cirrus samples.	85
5.3	Runtime (seconds) and memory usage (MiB) of multiscale attention calculation for a single batch, for different gridded attention modules and non-gridded attention.	86
5.4	Comparison of attention frameworks: dual attention [77], dual attention with Gabor conv. features, tri attention, tri-attention where channel and positional attention use Gabor conv. features. Results presented as mean IoU over 5 splits on real and synthesised cirrus samples.	87
5.5	Ablation study of the proposed gridded attention and tri-attention, and modifications of previous work [182]. Results presented as mean IoU over 5 splits on three cirrus datasets: large synthesised images (used in previous experiments), small synthesised images (used in Chapter 4), and real LSB images.	87
5.6	Segmentation scores on the SWIMSEG sky/cloud segmentation dataset, comparing the proposed gridded attention and tri-attention against previous works.	89
6.1	Comparison of different intensity scaling layers. Results reported as mean segmentation IoU over 5 splits. *Control model.	102
6.2	Comparison of different augmentation strategies. Results reported as mean segmentation IoU over 5 splits. Minimal augmentation uses rotations and flips, as described in steps 2 and 3 of Section 6.3.2. Contr. represents contrast augmentations. *Control model.	104
6.3	Performance of control model with and without the use of pretraining on synthesised images. Results reported as mean segmentation IoU over 5 splits. *Control model.	104
6.4	Comparison of BCE vs Focal loss functions with different consensus loss frameworks. Results reported as mean segmentation IoU over 5 splits. *Control model.	105

6.5	Ablation study of best performing training strategies. Results reported as mean segmentation IoU over 5 splits. *Control model.	107
6.6	Comparison of attention model modifications: generating multiscale features from intermediate layers; guided attention; use of the proposed gridded attention map; and computing Gabor attention in addition to dual attention. Additional networks are included for comparison. Results reported as mean segmentation IoU over 5 splits. First row represents the control model.	109
6.7	Results for the final network with different prediction generation techniques.	110
7.1	AP ₅₀ and AP ₇₅ scores across different classes from models trained with and without HITL data, evaluated with and without HITL data.	132
7.2	AP ₅₀ scores across different classes from models trained with and without HITL data, evaluated with and without HITL data.	137
A.1	Descriptions of attributes belonging to the shapes relation. Primary key is emboldened.	169
A.2	Descriptions of attributes belonging to the galaxies relation. Primary key is emboldened.	170
A.3	Descriptions of attributes belonging to the annotations relation. Primary key is emboldened, foreign keys are italicised.	170
A.4	Descriptions of attributes belonging to the shapes relation. Primary key is emboldened, foreign keys are italicised.	171

Abbreviations

AI	Artificial Intelligence
AL	Aladin Lite
AP	Average Precision
BCE	Binary Cross Entropy
BN	Batch Normalisation
BSD	Berkeley Segmentation Dataset
CANDELS	Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey
CCD	Charge Coupled Device
CFHT	Canada France Hawaii Telescope
CFIS	Canada-France Imaging Survey
CNN	Convolutional Neural Network
COCO	Common Objects in Context
DCT	Discrete Cosine Transform
DNN	Deep Neural Network
EM	Electromagnetic
FCN	Fully Convolutional neural Network
FITS	Flexible Image Transport System
GAN	Generative Adversarial Network
GCN	Gabor Convolutional Network
GPU	Graphics Processing Unit
GMM	Gaussian Mixture Model
HiPS	Hierarchical Progressive Survey
HITL	Human In The Loop
HTML	HyperText Markup Language
IoU	Intersection over Union
IR	Infrared
LSB	Low Surface Brightness
LGCN	Learnable Gabor Convolutional Network

MATLAS	Mass Assembly of early-Type GaLaxies with their fine Structures
ML	Machine Learning
MNIST	Modified National Institute of Standards and Technology
MS	Multi-scale
NN	Neural Network
NGC	New General Catalogue
ODS	Optimal Dataset Scale
OIS	Optimal Image Scale
PGC	Principal Galaxies Catalogue
PSF	Point Spread Function
PSNR	Peak Signal to Noise Ratio
R-CNN	Region based Convolutional Neural Network
RCF	Rich Convolutional Features
ReLU	Rectified Linear Unit
RGB	Red Green Blue colour model
RoI	Region of Interest
RPN	Region Proposal Network
SDSS	Sloan Digital Sky Survey
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
SWIMSEG	Singapore Whole sky IMaging SEGmentation
UGC	Uppsala General Catalogue

Chapter 1

Introduction

Recent astronomical imaging surveys have uncovered a vast array of interesting objects. As shown in Figure 1.1, advancements in imaging techniques have facilitated the capture of very faint, or low surface brightness (LSB), structures. Better understanding the physical properties of these objects, their past and future is crucial to astronomers as they provide clues to the wider nature of galaxy formation. The study of such structures requires thorough cataloguing and labelling to enable effective statistical analysis. This process typically involves recording information on different structures present in a given image, such as the type of object and its location. A more in-depth cataloguing effort also involves recording the exact spatial properties of structures, such as size and shape, as this information has the potential to allow astronomers to better understand associated physical phenomena. Manually cataloguing structures is a lengthy process and generally will be unfeasible for future surveys producing petabytes of LSB image data. One approach to handle the scale of data is to levy on astronomy hobbyists through community crowdsourcing, from which projects such as Zooniverse [133] have seen great success. While crowdsourcing has done well to fulfill the labelling requirements for surveys such as Sloan Digital Sky Survey (SDSS) [23] and Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) [33], upcoming surveys such as Euclid will produce orders of magnitude more data in comparison. The time required for cataloguing is becoming increasingly infeasible for future surveys with vast amounts of multi-spectral images at unprecedented high resolution.

A promising approach for the complex image processing involved in automated cataloguing of structures in LSB imaging is machine learning. Machine learning research has exploded over the past decade with much focus on deep neural networks, enabled by major hardware advancements in parallel computation. The efficient inference offered by modern machine learning techniques such as neural networks is relevant in astronomy given the vast sample sizes in some datasets. Convolutional neural networks in particular



(a) NGC0448.

(b) NGC7457.

Figure 1.1: Examples of low surface brightness images.

have been recently applied to a variety of tasks on astronomical images [55, 59, 201] including identification of tidal structures in LSB images [20, 200]. While such studies have seen great success at classifying objects, there have been limited attempts to delineate the exact spatial boundaries of objects [18, 31, 69, 87], and to our knowledge this has not been attempted in LSB images. This task of precisely localising an object and predicting the exact spatial location of the object in an image, is termed object segmentation in the computer vision research sphere.

Automated cataloguing of structures in LSB images currently presents multiple challenges, which form the major motivations of this thesis:

- **There is no annotated dataset containing segmentation labels of LSB images, nor a tool to produce such a dataset.** Training modern machine learning algorithms typically requires datasets with example outputs for each input sample, however, there exists no annotated dataset containing 2D labels of structures in LSB images. Furthermore, the creation of such a dataset is not so simple, as annotation of LSB images demands the ability to visually inspect large multi-spectral images and draw shapes over structures, which are features that are not simultaneously present in any public available labelling tool.
- **There is currently a limitation in terms of quantity and quality of available LSB images.** Huge parameter space models such as deep neural networks often require large uniform datasets to attain good generalisation. LSB imaging that boasts both high sensitivity and high resolution is a relatively recent tech-

nology, with only a handful of deep surveys each containing a small sample set in comparison to what is typically necessary for training ML methods. Images also commonly contain artefacting which degrade the quality and thus make training deep neural networks on such images more difficult.

- **LSB imaging instruments detect cirrus clouds which occlude interesting structures.** Due to the sensitivity of modern LSB instruments, scattered light from dust in our galaxy is captured, which presents as wispy cloud-like structures that contaminate images. This contamination greatly impedes analysis of LSB structures, and can appear visually similar to interesting subtle structures, such as material resulting from interactions between galaxies. Distinguishing between weaker cirrus contamination and areas with high background levels can be very difficult in some cases even for domain experts, yet still remains an important distinction for astronomers.

In this thesis we aim to systematically address these challenges, through carefully developing a strategy for data collection and by employing a wide range of machine learning techniques. In particular, a common theme of this thesis will be focusing on how to improve generalisation given the data limitations present in this study. Overcoming this issue will require a multi-pronged approach utilising techniques from many spheres of machine learning research investigating data efficiency, such as transformation invariance and few-shot learning. Finally, we seek to combine lessons learned from tackling these problems into a comprehensive automated cataloguing method, capable of object classification, detection and segmentation.

1.1 Contributions

There is a clear need for automated cataloguing of structures in future surveys, which this thesis aims to address through supervised machine learning techniques. Moreover, a key aim of this thesis is not just to identify the presence of certain structures, but to detect the exact location and shape, allowing further quantitative analysis of structures' spatial properties. This first requires training data so that machine learning algorithms can be trained. In Chapter 3, we develop a tool for collecting 2D annotation labels of structures in LSB images, and present both a real and synthesised dataset of annotated astronomical images. In Chapter 4, we present a modification to the convolutional operator to increase sensitivity to object rotation in images, such as galactic structures which can present in any orientation. In Chapter 5, we utilise attention to create a segmentation model capable of processing large images, and in Chapter 6 attempt to automatically segment cirrus

contamination in LSB images. In Chapter 7, we implement an instance segmentation model capable of simultaneous classification, detection and segmentation of structures in LSB images. This section is dedicated to highlighting the main contributions of this thesis, and how work carried out relates to the motivations detailed in the previous section.

1.1.1 Training data for automated cataloguing

In order to apply supervised machine learning techniques it is necessary to obtain example outputs for corresponding input samples, termed ground truth targets. Attaining generalisation in modern ML algorithms typically requires high quality datasets with a large sample size. To our knowledge, there does not exist any dataset of segmentation labels for LSB structures, which complicates developing an automated cataloguing method. While there exists several tools for annotating 2D labels over images, these do not accommodate the domain specific challenges associated with astronomical images. Namely, images contain multiple channels representing different wavelength spectrum bands, which the annotator must be able to study individually and in combination. Additionally, the location of drawn labels correspond to locations in a real world coordinate system, which should be recorded by the tool. We fill this gap through the development of an annotation tool for LSB images, allowing creation of segmentation targets. The process by which annotators should produce labels using the tool is explicitly detailed, so that bias among annotations is minimised. We then detail a dataset of annotated LSB images produced using our tool. While this dataset serves as an important foundation for training an automated cataloguing method, the number of available LSB images is limited. To this end, we present a method for synthesising images containing features similar to those present in LSB images. This work contains three major contributions.

- We present an online tool for annotating high resolution multi-spectral images. We utilise a popular astronomy image visualisation tool and enable the drawing of complex shapes over structures. The tool is entirely web based, supporting collaboration between multiple users and allowing annotations to easily be stored in a central database.
- The proposed tool is used to collect annotations from four users of 227 images from the MATLAS survey. We define a precise annotation protocol to ensure structures are annotated in a consistent fashion by all users across all images. We discuss methods to combine annotations made by multiple users into a single consensus ground truth, used for training ML models.
- We propose a methodology to generate synthesised samples of galactic cirrus, which

are suitable for pretraining ML models. Multiple noise patterns are carefully chosen and combined to create images with similar properties to LSB images, including structural patterns resembling cirrus clouds. Features learned by training an ML model on these images can be transferred to the target dataset of MATLAS images.

1.1.2 Learnable complex-valued Gabor convolution for robustness to orientation

Structures in astronomical images can present with any angle of rotation, thus it is beneficial to encode some understanding of orientation into the machine learning model. CNNs are inherently equipped with some capability to process translations of objects, i.e. vertical and horizontal shifts, owing to how weights are shared across different locations of the input. This capability, however, does not extend to local or global rotations, which is a major limitation. Such ability to handle rotation variations is typically encouraged through large datasets with orientational augmentation, where samples are re-input to the model after undergoing a manual rotation. Given that data efficiency is a major priority in this thesis due to sample size limitations of LSB structures, we attempt to integrate robustness to variation of orientation in an a-priori fashion. We investigate using Gabor filters, which are analytical filters characterised by rotation sensitivity and frequency localisation, to modify convolutional weights in order to render them more sensitive to orientation. In order to use the full Gabor filter, we use a complex-valued CNN where all layers are modified to support complex-valued arithmetic. The contributions of this work are as follows.

- We present a novel convolutional operator which utilises complex-valued Gabor filters in order to gain sensitivity to orientation. Gabor filters are analytical filters, where rotation and scale can be controlled directly through parameters. We learn these parameters alongside convolutional filters. We refer to the process of modifying convolutional weights with Gabor filters as modulation.
- We present cyclic Gabor convolutions, where modulated weights generated from different Gabor filters are applied to each input feature map. Cyclic Gabor convolutions utilise an iterative process where a set of Gabor filters is applied to convolutional weights, cyclically shifted so that the order of filters is altered, and then reapplied to create a new set of modulated filters.

1.1.3 Multi-scale gridded Gabor attention network for segmentation of global contaminants

The ability to contextualise discriminating features is incredibly important in computer vision. The global surroundings of given features or regions of interest are often key indicators for correct predictions. Convolutions inherently handle local textures well, but features describing global relations are only learned in later layers after multiple successive pooling operations. For processing of contaminants which cover large regions in images, global context is vital for accurate performance. Such structures also often exhibit orientational patterns both locally and globally, and thus standard convolutions are suboptimal. The attention operator has been used to capture longer range dependencies in images, though this is generally has a large memory footprint. We investigate the use of attention to create an memory efficient model capable of processing large images while studying long range dependencies. Further, we seek to integrate the orientation sensitivity offered by Gabor filter modulation, proposed in the previous work, into this attention mechanism. The contributions of this work are as follows.

- We present a novel attention operator utilising Gabor filter modulation where attention is computed with respect to orientation dependent features. Correlations are measured across a new axis representing features dependent on different orientations, allowing the machine learning model to study relationships across orientations.
- We present a multi-scale gridded attention mechanism for computing attention on very large images without downsampling or cropping. Feature maps are generated at different scales and then divided into tiles so that spatial size is standardised. Attention is calculated on each tile and then reassembled to create new feature maps of the original scales.

1.1.4 Segmentation of cirrus contamination with deep learning

Identifying cirrus clouds in LSB images is a major priority for future astronomical imaging surveys. Traditional imaging instruments do not detect cirrus, however the sensitivity of LSB imaging captures light scattered by dust particles within the clouds. Cirrus appears in the foreground of images and occludes interesting structures, impeding statistical analysis of LSB galaxies. In addition to being a contaminant, LSB images of cirrus also provide pictures for high resolution studies of the interstellar medium. While cataloguing of cirrus can currently be performed manually, future surveys producing massive amounts of astronomical image data will require automated methods to record regions affected by cirrus contamination. This is a difficult task as cirrus contamination varies highly

in severity and can be easily confused with other structures or high background levels. This is compounded by the data limitations such as sample size and the high resolution of images. We seek to design a machine learning pipeline for segmenting cirrus clouds which accommodates these domain specific challenges. The contributions of this work as follows.

- We apply the gridded Gabor attention network to the task of segmenting cirrus contamination in LSB images.
- We present an adaptive intensity scaling layer which combines a common astronomical image preprocessing step into the deep neural network, where subtle features are enhanced. Parameters dictating the strength of the scaling operation are learned alongside neural network parameters.
- We propose a loss function for training on probabilistic targets. We coarsely divide probabilities into groups and weight the loss function based on the confidence of the label.
- We apply the proposed pipeline to a novel dataset of annotated cirrus structures in LSB images.

1.1.5 Detection and segmentation of galactic structures in LSB images

There is a need for automated classification of galactic structures in LSB images. Technology advancements in LSB imaging surveys have revealed many interesting structures related to pictured galaxies. For example, tidal features can be detected in LSB images, which are remnants of interactions between galaxies. The presence of such interesting structures and their shapes give clear signs as to how different galaxies are formed. It is thus key that galactic structures in LSB images can be processed and catalogued to facilitate statistical analysis and to further research into how galaxies form and evolve. As is the case for cirrus, cataloguing is currently performed manually by domain experts. With future surveys set to produce orders of magnitude more image data than currently available, investigation into automated processing is a necessity. We explore the use of deep learning to process galactic structures in LSB images. Given that presence and shape of structures is important to astronomical research, we attempt to both detect and segment objects, i.e. the task of instance segmentation. We incorporate lessons learned from previous chapters in handling LSB images and develop an instance segmentation

model capable of making predictions directly from LSB images with no preprocessing. The contributions of this work are as follows.

- We detail the process of preparing MATLAS annotation data to facilitate training instance segmentation models.
- We apply Mask R-CNN on detection and segmentation of localised galactic structures, and investigate the use of a human-in-the-loop training protocol to include correct object predictions which are unannotated into the training dataset.
- We present a conceptually simple panoptic segmentation method which involves combining Mask R-CNN with the attention model of the previous work. This model is used to simultaneously segment localised galactic structures and global contaminants.

1.2 Outline

In this section, we briefly describe each of the remaining chapters of this thesis.

Chapter 2 Background

We detail the prerequisite machine learning concepts related to this thesis, including convolutional neural networks and supervised training. We also provide a background on low surface brightness astronomical imaging and the types of structures analysed throughout this work.

Chapter 3 Creating an LSB Dataset for Supervised Learning

An annotation tool for labelling astronomical images is presented. Annotation labels of 227 low surface brightness images from the MATLAS survey are detailed, and the method of combining labels from multiple users into a single consensus is discussed. A method for synthesising samples of galactic cirrus suitable for enabling transfer learning in deep neural networks is presented.

Chapter 4 Learnable Gabor Modulation in Complex-valued Neural Networks

We present a novel modification to the convolutional layer where kernels are modulated with Gabor filters in order to increase sensitivity to orientational patterns. A complex-valued CNN is implemented with Gabor modulation and applied to problems exhibiting different types of rotational symmetries.

Chapter 5 Multiscale Gridded Gabor Attention for Segmenting Global Contaminants

We present a compute-efficient attention operator for segmentation of global contaminants, where attention is calculated over tiles of different scales. We also propose a

method using attention to measure global orientational dependencies, where Gabor modulated convolutions are used to extract orientational features.

Chapter 6 Segmentation of Cirrus Contamination: a Deep Learning Approach

A pipeline for automated cataloguing of cirrus contamination is presented, involving the attention network of the previous chapter. We discuss multiple domain specific challenges associated with low surface brightness images and propose solutions, including a loss function for training deep neural networks on coarsely probabilistic targets.

Chapter 7 Multi-class Segmentation of Galactic Structures

We investigate the task of panoptic segmentation in low surface brightness images. A conceptually simple model is developed where the attention network of previous chapters is combined with Mask R-CNN to create a model capable of segmentation of localised objects and extended homogeneous structures.

Chapter 8 Conclusion

Final concluding remarks are drawn on the thesis, where the separate contributions are revisited and contextualised among the central goals of this work. Discussion surrounding avenues for future work is provided alongside this summary.

Chapter 2

Background

In this interdisciplinary work, we combine two deep areas of research: machine learning and astronomy. To digest the work in this thesis and fully understand the motivations, it is necessary to set the stage. This chapter details the necessary prerequisite concepts relating to this thesis. In the first half, key concepts and challenges in galaxy evolution are outlined. In the second half, the state of image processing with deep learning is covered.

2.1 Galaxy Evolution and Low Surface Brightness

Understanding the universe has long been a central goal of our species at large. One method of progressing this understanding has been through studying the processes of stars and galaxies. To the naked eye, the observable universe appears as a collection of bright objects, which are mostly stars in our own galaxy, the Milky Way. Galileo's first telescope in the seventeenth century increased viewing capacity, allowing astronomers to study objects that were fainter (less luminous) or further away. Over a century later, further advances in telescope technology enabled Herschel and others to discover that our galaxy was roughly shaped as a disk.

In the early twentieth century, much effort was spent measuring the properties of the Milky Way more meticulously. The work of Leavitt and Pickering [121], revealing the relationship between pulsation period and luminosity of variable Cepheid stars, facilitated accurate measurements of other stars' distances and thus the size and shape of the Milky Way¹. The period-luminosity relation was then later used by Hubble [98] to prove the existence of galaxies other than our own, a major discovery which birthed extragalactic astronomy. Shortly after this, Zwicky [221] discovered a disagreement in some galaxies between calculated velocity, based on observed redshift, and calculated mass, based on

¹Along with disproving heliocentrism.

observed luminosity, where these galaxies moved much faster than could be explained by their observable luminous mass. This finding led to him arguing for the existence of dark matter, matter which does not absorb, reflect or emit electromagnetic (EM) radiation, and thus is non-luminous.

Today, the most widely accepted understanding of the formation of galaxies is modelled by a concept termed as hierarchical merging [46, 208]. In this hierarchical model, smaller galaxies combine through clustering and merging due to gravity, accumulating mass and forming larger galaxies of a new shape and structure, or morphology. Such merger events are a significant driver of galaxy evolution and can cause major changes in galaxy morphology.

2.1.1 Low Surface Brightness Structures

Whereas stars are concentrated sources of light, or "point sources", light from galaxies is spread over a patch of the sky. This light distribution is described in surface brightness, which measures brightness per unit area. The surface brightness of a galaxy directly relates to the density of stars inside it from the observer's perspective, and is thus an intrinsic property of the galaxy. Low surface brightness (LSB) is a term used to describe a brightness level fainter than the night sky.

The surface brightness of the night sky is composed of a variety of objects and structures which vary in intensity and emit different spectral ranges of electromagnetic radiation [124]. These sources include:

Airglow – light emitted due to the glowing of the upper atmosphere.

Zodiacal light – sunlight scattered by dust particles in the Solar System.

Integrated starlight – combined light from low luminosity stars in the Milky Way.

Diffuse galactic light – starlight scattered by dust particles in the Milky Way, also referred to as galactic cirrus.

Extragalactic background light – light from undetected sources outside of the Milky way.

This combination of light sources amounts to a significant disruption in what can be observed in the LSB universe. Further, photon shot noise becomes a relatively larger factor as signal decreases, and has a non-negligible effect on observations of LSB structures. Galaxies with a low surface brightness are thus not well observed, emitting diffuse scattered light which telescope instruments and postprocessing pipelines have traditionally not been optimised to detect.

Large surveys of objects and surveys with higher surface brightness levels than the LSB regime have facilitated large scale statistical studies of the galaxy population, progressing understanding of galaxy evolution. Such analyses, however, have been naturally biased by the incompleteness of LSB object detections. This visibility bias was first noted by Disney [57], who made the analogy of observed galaxies being only the tip of the iceberg, in terms of the number of undetected faint objects, and that only the bright centre of galaxies was pictured. A consequence of this is that understanding of galaxy evolution and driving phenomena are predicated on a subset of the galaxy population, i.e. not including LSB galaxies.

LSB features of detected high surface brightness galaxies offer a wealth of matter that can be studied to further constrain models of the Universe. An example of such components are tidal features, induced by merger events between galaxies, and have been used to provide modelling constraints on dark matter [47]. Tidal features necessarily encode the formation and evolutionary history of associated galaxies. The majority of merger events produce faint tidal features, undetectable in past imaging surveys [62]. The analysis of LSB tidal features is thus key to understanding galaxy evolution. Observing such structures is currently very challenging due to the surface brightness depth required, though observations have been possible through a combination of telescope instrumentation technologies, specialised image post-processing and observing strategies.

2.1.2 Imaging Instrumentation and Observing Strategies

Modern LSB images in the optical and near infrared EM bands are typically captured with charge-coupled devices (CCDs). CCDs are rectangular semiconductor chips with a light sensitive face. Incident photons generate a small electrical charge, due to the photoelectric effect, which is stored in a "potential well". As more photons arrive at the CCD's face, charge accumulates in the well. CCDs contain many wells which correspond to individual pixels. For image capture, the CCD is placed in the focal plane of a telescope which incident photons illuminate, forming an image of the region of the sky that is being viewed. Electrons stored in potential wells are then released and "read out" in a sequential fashion, allowing brightness values to be calculated for each well or pixel. Multiple CCD instruments are often used in combination to either increase the size of an image or reduce noise by averaging over photon shot noise. This can sometimes be the cause of artificially high background levels in isolated sections of an image, where not all CCD read outs are used for this averaging process (due to miscellaneous errors), leaving areas with higher noise levels. Another possible artefact of CCDs encountered in this thesis are saturation trails: potential wells have a limit of charge storage which when reached causes further

electrons to spill over into neighbouring wells. Very bright objects thus can erroneously illuminate nearby pixels, usually in a vertical pattern.

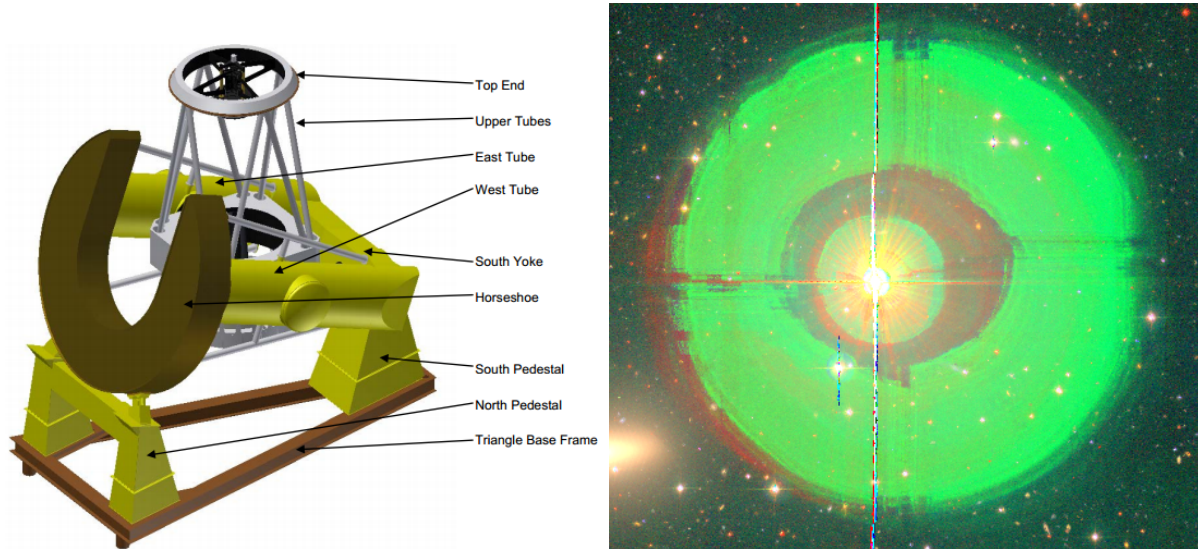


Figure 2.1: *Left:* Diagram of Canada France Hawaii Telescope (CFHT) Megacam's main components, taken from [193]. Note the four spider arms in the top end. *Right:* Internal reflections cast by a bright star in a MATLAS image taken with CFHT's Megacam.

There are several imaging specific challenges related to capturing LSB structures, as systematic instrumentation effects are far more significant when capturing faint surface brightnesses. Careful handling of the point spread function (PSF), which models the diffraction pattern of light emitted from point sources, is important to good performance in LSB imaging. Typical narrow PSFs tightly focus on incident light from point sources though leave extended "wings" which bleed over the focal plane and in LSB imaging can present as more significant artefacts, sometimes appearing visually similar to interesting objects [175]. To mitigate this, there are works integrating subtraction of extended PSF wings into LSB imaging pipelines [12, 183]. It is also important to minimise the amount of internally reflected scattered light reaching the telescope's focal plane. Light arriving on the CCD can reflect back through the telescope, reflect again off internal components of the telescope, and then land on the CCD surface. In practice, this commonly occurs where light reflects off of the telescope's secondary Newtonian mirror and its supporting "spider" arms, leaving a faint shadow and bright cross near imaged stars. This effect can be mitigated by using anti-reflection coatings or minimising the amount of internal supporting structures, such as is done in the Dragonfly telescope array [1].

Atmospheric effects can generate noise in the image, of which the strength and pattern varies with time of day and the area of the sky being imaged. Diffraction and scattering

of light due to the atmosphere can also change the PSF, increasing the extended wing effect. One observational strategy to mitigate against atmospheric and other systematic effects is "dithering", where multiple observations are taken with the viewed patch of sky altered slightly each time (e.g. [195]). Astronomical sources predictably change location in the image, whereas unwanted systematic effects do not, allowing them to be identified and accounted for. Future space-based telescopes, such as Euclid [119], will alleviate atmospheric effects.

2.1.3 Cataloguing

To facilitate statistical studies on celestial objects, it is necessary to record logs of observed objects in a structured manner. Such a record is termed an astronomical catalogue, and contains a tabulation of celestial objects that share some association. Along with an identifying name or code, additional details of objects are recorded, such as size, location and distance. This cataloguing process of logging astronomical information has traditionally been carried out manually by astronomers. Prior to the invention of astronomical telescopes, the largest and most accurate astronomical catalogue by Brahe [28] contained 777 stars. In comparison, the latest release of the Gaia will detail approximately 1.8 billion stars [29, 164], owing to modern imaging technologies and large automation efforts.

There exist numerous catalogues that include galaxies which are referred to in this thesis. Most notable is the New General Catalogue (NGC), compiled in 1888, which along with supplementary updates, known as the Index Catalogues (IC), contains 13226 objects. Many objects detailed in the NGC/IC are still commonly referred to using their NGC or IC numbers, which appears as a four digit number preceded by either NGC or IC. The 1973 Uppsala General Catalogue (UGC) [153] contains 12921 galaxies visible from the northern hemisphere. The Principal Galaxies Catalogue [159] is a collection of 73197 galaxies, published in 1989. PGC is a collation of galaxies from multiple widely used catalogues, such as NGC and UGC. These mentioned catalogues relied on mostly manual processing to obtain tabular information, which for the sample sizes was manageable.

The 1990s were characterised by a major change in astronomical data quantity. Technological advancements brought digital sky surveys, able to capture the sky at an unprecedented scale. To cope with the increased magnitude in sample size, there has been much effort spent in automating the processing steps required to catalogue observations. Source extraction is a key area of this research, which is the process of identifying and extracting information from individual luminous astronomical objects or sources, and involves producing a segmentation mask of sources. SExtractor [19] is a widely used source extraction tool, which uses a pipeline consisting of mostly traditional computer vision

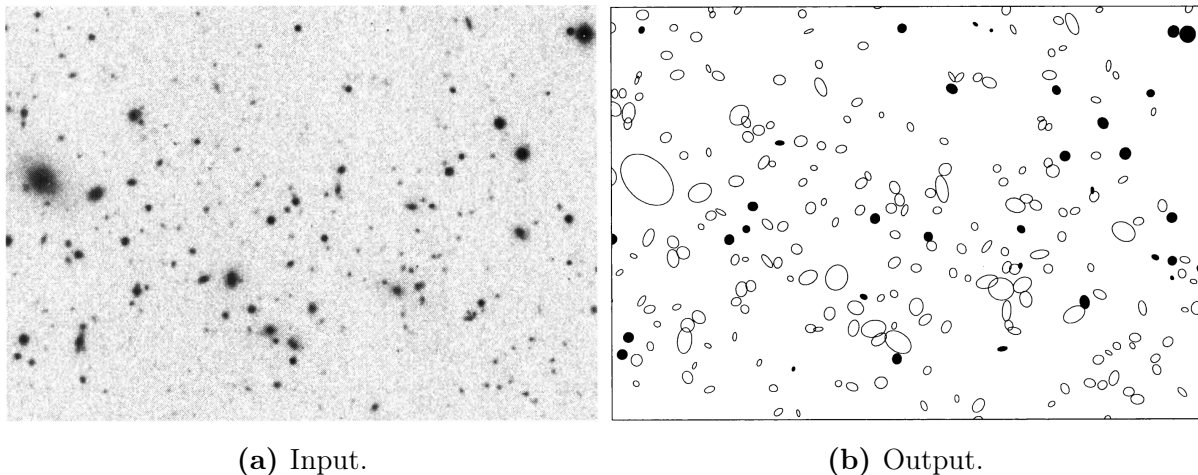


Figure 2.2: Example interpretation of a CCD frame by SExtractor, taken from [19].

techniques, such as Gaussian smoothing and adaptive thresholding. More recent source extraction tools include ProFound [169], NoiseChisel+Segment [3] and MTOBJECTS [192]. These tools are optimised for objects of higher surface brightness, and thus struggle in the LSB regime where signal-to-noise ratio (SNR) tends to be lower. Notably, Prole et al. [163] propose DeepScan, which extends SExtractor using the DBSCAN algorithm [68], for source extraction in LSB images.

More involved cataloguing tasks requiring complex image processing, such as object classification tasks, have naturally led to application of and research into machine learning techniques. Early works in this direction involved using neural networks (NNs) for star-galaxy discrimination [155, 156]. NNs were also applied to galaxy morphology classification [150, 151, 187] and abnormal spectra detection [75, 199]. These networks followed the prototypical NN design, with a single hidden layer formed of a relatively small number of nodes, and the input layer was fed various astronomical parameters known to be discriminative. Other attempts at automation with machine learning alongside these works used variations of decision trees for morphology classification [157, 207]. Following these works, in the 2000s, several works applied: a larger NN with more astronomical input parameters [14]; an ensemble of NNs [15, 16]; and SVMs [100, 161] to morphology classification. Crowdsourcing cataloguing tasks with citizen scientists [133] has also been used for galaxy morphology classification on more recent surveys (e.g. [23, 33]).

More sophisticated techniques are necessary to facilitate cataloguing of modern surveys, due to the unprecedented quantity and quality of astronomical images. This is especially the case for LSB images for which the poor SNR renders traditional machine learning techniques unreliable. Contamination of structures by galactic cirrus compounds this problem, which occludes objects and increases ambiguity of spatial size and shape.

2.2 Deep Learning for Image Segmentation

Much of artificial intelligence research is currently dominated by deep learning studies, owing to advances in neural network architectures and statistical learning. Prior to this, machine learning approaches on image processing tasks involved training smaller models on handcrafted features. Modern deep learning models now incorporate this feature generation step into the model, enabling models to capture exact and optimal features related to the task. Undoubtedly, the most popular computer vision architectures of the past decade are convolutional neural networks (CNNs). In this section, we lay out the CNN architecture, discuss landmark works and detail common strategies for effective and efficient training.

2.2.1 Convolutional Neural Networks

Complex image processing tasks require predictive models with a large learning capacity. This is a problem for standard neural networks, which explode in parameters based on the dimensionality of input data or intermediate features. Processing higher resolution image data with NNs quickly demands models with billions of parameters, making training very computationally inefficient or even infeasible. Such models require large compromises to be made to facilitate training, e.g. heavy image downsampling or reducing the number of network layers, making the use of standard NNs impractical for modern computer vision problems.

Convolutional neural networks greatly improve on parameter efficiency in comparison to NNs on computer vision tasks. The learning capacity of CNNs can be controlled through various means: breadth is altered through channel and kernel size; depth is altered through the number of layers. CNNs are also inherently well equipped to "understand" discriminative patterns in images. This is first due to the approximate translation invariance offered by convolutional layers, where features are extracted similarly irrespective of their location in the image [117]. Secondly, the structure of CNNs captures a descriptive range of low-level image statistics [196]. This is a consequence of the fact that convolutional kernels are typically of a small size, e.g. 3×3 pixels, and convolutional layers are used sequentially in combination with downsampling, allowing images to be broken into a hierarchy of descriptive components. This parameter efficiency combined with hardware advances in parallel processing through modern GPU devices has enabled feasible training of large CNNs capable of difficult image processing tasks [41, 117].

The prototypical CNN design consists of several repeated blocks placed in succession to extract a hierarchy of discriminative features. This block consists of several operations which each perform a key function in the CNN: the convolutional layer, pooling layer, and

activation layer. We note that pooling and activation layers commute so their order has no impact on network output, though pooling followed by activation is more computationally efficient as the activation layer acts on downsampled features. We now detail each of these convolutional block components.

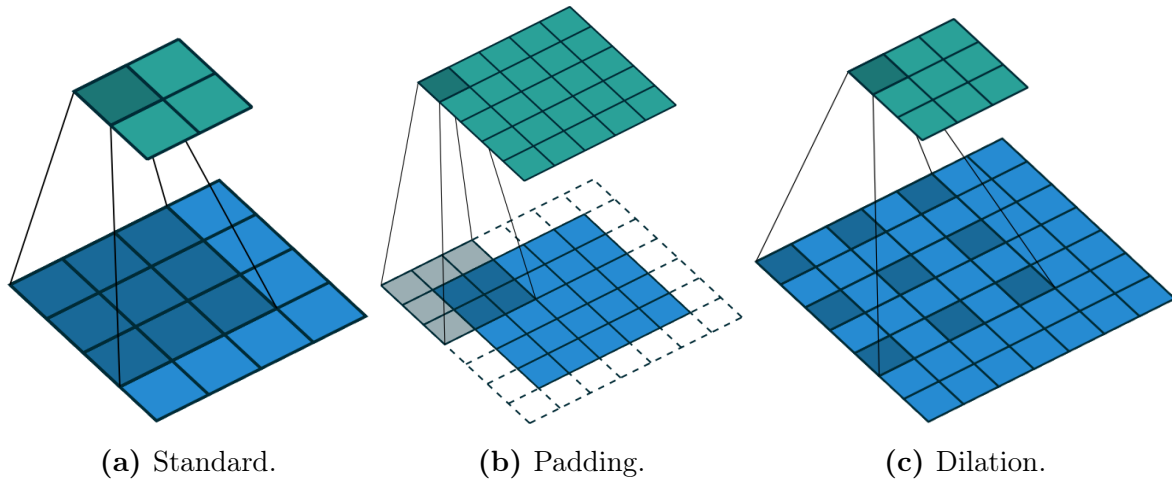


Figure 2.3: Visualizations of 3×3 convolutions with different hyperparameter setups, taken from [65]. (a) shows a standard convolution; (b) shows a convolution with unit padding; (c) shows a convolution with dilation set to 2.

Convolution – The convolutional layer is typically placed first, where a collection of kernels with predefined fixed size are convolved over input data. Values of the kernel, or weights, are learnable through backpropagation, allowing convergence onto an (hopefully) optimal kernel configuration. Other than kernel size, the behaviour of the layer can also be controlled through: stride, which forces the convolution to skip over a given interval of pixels; dilation, which expands the range of a convolving kernel; and padding, which artificially adds pixels along the boundaries of input. The convolution computation produces a collection of features which can be used for further processing.

Pooling – A downsampling pooling layer follows second, which reduces the dimensionality of input features while preserving discriminative information. The pooling operation slides a kernel over the input and performs some downsampling strategy over each data "window". Two common pooling strategies are max pooling, which takes the maximum pixel value in each window, and average pooling, which takes the mean pixel value in each window. An important property of the pooling layer is that it introduces local invariance into the network, as the maximum value of a set of pixels is unaffected by the ordering of such pixels. Thus, the layout of pixel values in a window does not affect the output.

Activation – The final third layer in the block is a nonlinear activation function which transforms feature values into an "activation" value where higher values should denote

a stronger feature response. While a linear activation can be used (i.e. no activation), patterns in images are usually nonlinear. Given that all other layers in CNNs are typically linear, it is desirable to introduce a nonlinearity that can capture such patterns. Traditionally, the sigmoid or hyperbolic tangent functions were used as nonlinearities, which respectively map input onto values between 0 and 1, and -1 and 1. A drawback of both functions which becomes noticeable in larger models is saturation, where high magnitude values all map onto a small neighbourhood around the limits of the function. For example, $\tanh(3) \approx \tanh(30)$. Once a neuron becomes saturated, large changes in weights correspond to negligible changes in outputs thus making backpropagation ineffective due to vanishing gradients. A common choice of activation function which solves the saturation problem is the rectified linear unit (ReLU) [79, 152], which unchanges positive values and sets negative values to zero, i.e. $f(x) = \max(0, x)$. While negative values effectively saturate, such weak responses can be thought of as irrelevant to the learning problem. One gentle approach to alter this scenario is the leaky ReLU [142], which linearly suppresses negative values rather than setting them to zero.

Following the application of several successively chained convolutional blocks, resulting low-dimensional features must be transformed into predictions. A common next step is to apply a fully connected layer. Whereas convolutions focus on local textural patterns, a motivation of using fully connected layers on the low-dimensional features is to facilitate learning of global patterns, as relationships between individual features in a feature map can be expressed. Prediction then involves a choice of activation layer depending on the computer vision task. In regression problems, linear layers are often used. In classification problems, sigmoid and softmax are commonly used for binary and multiclass data, respectively.

2.2.2 Image Segmentation

The goal of segmentation is to classify exact regions in an image that belong to a semantic class. Whereas standard classification tasks involve prediction of a probability distribution output per input sample, segmentation requires prediction of a probability distribution per pixel. This distribution can cover an arbitrary number of classes: binary segmentation refers to the scenario with two classes, positive and negative; and semantic segmentation refers to the scenario with more than two classes.

Early works investigating image segmentation with CNNs applied a "sliding window" approach, where the image is divided into many smaller patches and classified [40]. Patch predictions are then recombined to obtain a segmentation prediction. This approach is computationally expensive and discards significant contextual information.

Modern CNN-based segmentation methods use the entire image as input, and output a corresponding entire segmentation map. Long et al. [138] propose fully convolutional networks (FCNs), where final fully connected layers are replaced with convolutions of unit kernel size, allowing input images to have arbitrary size. Lower dimensional features are also upsampled with transposed convolutions combined with features from intermediate layers. Ronneberger et al. [172] build on this with U-Net, a symmetric fully convolutional network where each convolutional layer informs the opposing transposed convolutional layer during upsampling. This method of transferring information across multiple layers is referred to as using "skip connections". Chen et al. [35] investigate the use of dilated convolutions to expand the receptive field of convolutions to increase contextual information, and later use conditional random fields for boundary refinement [34]. Zhao et al. [218] propose the pyramid pooling module, which applies multiple convolutions of different kernel sizes to low dimensional features before upsampling into a segmentation. More recent background works related to this thesis are reviewed throughout individual chapters.

To perform comparative studies among different methodologies it is necessary to be able to quantify predictive performance. The most simple way of doing this is by measuring pixel accuracy, or the proportion of correctly predicted pixels. Such a metric, however, is poorly suited to datasets with class imbalances. Two popular image segmentation metrics in computer vision literature are Intersection over Union (IoU) and the Dice (or F1) score. IoU measures the ratio over overlap between positive ground truth regions and positive predicted regions:

$$\text{IoU}(A,B) := \frac{|A \cap B|}{|A \cup B|} \quad (2.1)$$

where A is the ground truth label, and B is the predicted segmentation. Dice score is defined as the harmonic mean between precision and recall:

$$\text{Dice}(A,B) := \frac{2|A \cap B|}{|A| + |B|} \quad (2.2)$$

2.2.3 Improving Data Efficiency

A downside of the high learning capacity of deep CNNs is that they require large amounts of data to train. Features can only be learned that perform well on the training set, and unseen samples that are sufficiently unlike training samples are likely to be misclassified. Without a large dataset of suitable quality, CNNs are extremely prone to overfitting. There is thus a strong motivation to explore the use of techniques which improve the data

efficiency of CNNs and reduce overfitting.

Data augmentation is one such technique that can greatly improve model generalisation [117, 122, 181]. Augmentation seeks to reduce overfitting by artificially expanding the training dataset. This is achieved by applying some transformation to training samples, exposing the model to desired variations. Typically, augmentations take the form of geometric transformations: translations, rotations or flips; and element-wise transformations: brightness, contrast, saturation, hue adjustment or noise addition. This is especially desirable when symmetries or variations exist in the dataset. For example, galaxies can appear in an image at any location, orientation or reflection, thus geometric augmentations can induce approximate invariance to these transformations in the trained model.

Another paradigm for mitigating overfitting effects is reusing network weights from other tasks. This process is referred to as transfer learning, or pretraining. A model that is trained on a large dataset of good quality learns a wide range of descriptive features. These features can be "transferred" over to a new task, and used as an initial starting point. This process allows training deep models on small datasets, provided there exists a large labelled dataset available for pretraining. Transfer learning can be beneficial even between datasets that do not appear visually similar, as starting training from a set of descriptive learned features discourages overfitting.

2.2.4 Practical Tricks for Training Deep CNNs

Optimisation algorithm – The choice of optimisation algorithm has a direct impact on the characteristics of model convergence in the form of computational cost and better optima. Backpropagation involves optimisation of some loss function with respect to network weights through a gradient descent style algorithm. Plain gradient descent is a poor compromise between these factors as backpropagation only happens once for each forward pass of the entire dataset. Stochastic gradient descent is a proven adaptation of this, where backpropagation is applied for each (randomly ordered) training batch. With SGD, convergence is faster due to a compounding effect of more optimisation steps which outweighs gradient "noise" introduced by smaller sample size per step [26]. A popular optimisation algorithm which empirically improves on convergence speed is Adam [111], which builds on work combining SGD with adaptive moment estimation [64]. Adam uses momentum, where backward passes utilise the exponential weighted average of the current batch and past batches in an epoch.

Optimisation algorithms typically contain several hyperparameters. Learning rate controls the strength of each backpropagation update: common values in the literature are between 10^{-3} and 10^{-5} . Larger values allow fast but poor convergence, and the oppo-

site for smaller values. Weight decay applies a regularising effect to kernel weights which grow large magnitudes and cause network instability, where small changes in network input correspond to large changes in output [167]. The L2 norm of weights, multiplied by the weight decay hyperparameter (typically values vary between 10^{-4} and 10^{-7}), is added into the loss function so that weights are optimised to be small. Finally, adaptive moment estimation based optimisation algorithms, such as Adam, require setting a momentum hyperparameter, which controls the weighting of past accumulated gradients in the exponential average.

A practical trick often used in conjunction with an optimisation algorithm is learning rate scheduling. This involves adjusting the learning rate during training to be larger at the beginning of training, and smaller at the end. This trick exploits easy early gains and readjusts later to extract micro performance increases by settling into minima. Several strategies exist for learning rate scheduling. Step scheduling divides learning rate by some factor after every interval of some number of epochs, e.g. half every 25 epochs. Time-based scheduling operates similarly but after every given time-step. Exponential scheduling multiplies learning rate by a small coefficient after every epoch, referred to as the learning rate decay.

Weight Initialisation – Careful weight initialisation is important for stable training. A proper weight initialisation strategy must avoid exponential growth or decay of feature response magnitudes due to successive layer chaining. A standard practice is to encourage magnitudes to have unit variance. This is usually achieved by randomly initialising weight values with a Gaussian distribution of zero mean, and variance depending on the method. A widely used initialisation strategy is known as He initialisation [88], where variance is derived based on the number of input channels n or output channels \hat{n} , multiplied by the convolutional kernel width K . To preserve magnitude variance during the forward pass, one sets variance to $\frac{2}{nK^2}$, and to preserve magnitude variance during the backward pass, one sets variance to $\frac{2}{\hat{n}K^2}$.

Batch Normalisation – Another strategy to mitigate against training instabilities in large models is batch normalisation (BN) [103]. Maintaining unit variance across layer outputs can be difficult in larger models where, in addition to its own weights, the output distribution of a layer is heavily dependent on the distribution of previous layers which constantly change due to backpropagation. BN normalises layer inputs in an attempt to mitigate against this effect, though more recent work has shown that its success is largely attributed to different factors. Santurkar et al. [176] demonstrate that BN has a smoothing effect on the loss landscape, making gradients more predictable and stable. Bjorck et al. [22] find a synergistic effect between batch normalisation and more aggressive learning rates, thus BN enables faster convergence by avoiding or "stepping over" sharp

local minima.

Finally, there are two commonplace practical techniques for numerical stability not specific to deep learning. Backpropagation often involves dividing by quantities, e.g. consider the gradient of a square root. As such quantities become small and tend to zero, gradients explode and overflow, effectively breaking training. The frequency of these scenarios is reduced by choosing mathematical operations which avoid divisions, such as using mean squared error as loss rather than root mean square error. One seemingly unavoidable scenario is the computation of sigmoid or softmax in classification problems. This, however, can be mitigated using the "log-sum-exp" trick where the fact that the loss function immediately follows sigmoid/softmax is exploited. Using a log likelihood loss function such as cross entropy enables this scenario, and sigmoid/softmax and the loss function can be combined into a single numerically stable function. Another trick which mitigates overflow in unavoidable division scenarios by simply adding a small constant to the denominator, preventing division by zero.

2.3 Summary

In this section, we detailed the necessary prerequisite information for dissemination of this thesis. A brief history of galaxy evolution and low surface brightness research was set out, before covering astronomical instrumentation and its surrounding challenges. We detailed popular galaxy catalogues, and discussed how machine learning techniques have been used to assist with cataloguing tasks. A background to deep learning and convolutional neural networks was then documented. Finally, we provided discussion on several practical techniques, commonly mentioned in the machine learning literature, that are applied throughout this thesis.

Chapter 3

Creating an LSB Dataset for Supervised Learning

In this chapter, the process of creating a dataset of LSB images for supervised machine learning is detailed. For this thesis' study of exact spatial detection of galactic structures, 2D segmentation labels of all structures are required for each LSB image. As high resolution LSB images are a relatively recent advancement in astronomy, there exists little annotated data to be used as training targets. To accommodate supervised learning, it is thus necessary to create a dataset of LSB images with corresponding 2D annotation labels categorised by different types of structures. This chapter presents a tool for creating said annotations and the exact protocol by which annotations are performed, as well as a method for synthesising LSB images for pretraining CNNs.

3.1 Introduction

An inherent requirement of supervised learning is the need for corresponding ground truth target labels for each input training data. While there exists a few examples of ML segmentation techniques applied to astronomy images, these methods either use unsupervised graph-based techniques [31, 87], train on automated masks [69] created with SExtractor [19], or train on purely synthesised images [18]. Supervised segmentation of galactic structures is a relatively unexplored area of research, and to the author's best knowledge no extended dataset of segmentation labels for LSB structures exists. It is therefore necessary to create an annotated dataset suitable for training supervised ML algorithms.

Astronomical images present several challenges for precise 2D annotation, in comparison to natural images. Namely, images must cover a large region of the sky so that the

annotator can study surrounding regions of interest, and be of high enough resolution so that local scale features are not lost. Images are also multi-spectral, with which each associated wavelength band must be able to be studied individually and in combination. Given the extra uncertainty introduced through pixel-wise labelling, i.e. boundaries or even the presence of structures may be ambiguous, it is important for annotators to be able to collaborate on annotations. Finally, the location of regions of the sky captured by an image is represented by real world coordinates, and each pixel can be mapped to a specific coordinate or location of the sky. 2D annotations of astronomical images thus also occupy a world coordinate space, which is of interest to any study considering the astronomical nature of images and their annotations. Additionally, world coordinates of the sky are a parameterisation of spherical space, which must be taken into account when projecting images and annotations onto the user interface, which inherently exists in 2D Cartesian space.

A purpose designed tool is necessary to effectively and efficiently annotate LSB images. While there exists several public annotation tools, such as CVAT [178] or LabelMe [173], the ability to efficiently and precisely draw 2D labels on high resolution multi-spectral astronomical images is not supported, to our knowledge. Such tools also do not natively support world coordinate mapping or non-Cartesian projection. We present a tool for the creation of 2D annotations on astronomical images which considers the aforementioned factors. We design a drawing tool that enables users to quickly draw and amend complex shapes. This drawing tool is integrated into a popular astronomical image visualisation application, Aladin Lite [24], allowing users to properly inspect multi-spectral images through zooming and panning. Aladin Lite also handles spherical projection and stores world coordinate information of images, which we exploit to encode user generated 2D annotations in world coordinates. The annotation tool presented in this chapter is entirely web based, supporting collaboration from multiple users while annotations can be stored on a central server.

Discrepancies in annotations due to contradictory ideas of how annotations should be performed, termed recall bias, is an important factor to consider in managing the process of gathering annotations of data. Defining exactly what objects should be annotated and how is key to ensure annotations attempt to describe the same intrinsic properties of every image. It is vital that annotators have consistent definitions of categories of objects, and each delineate objects in the same manner so that any trained algorithm is not ‘confused’ by different annotators using different annotation strategies. This is not to say that each user’s annotation must be identical, but that their idea of how to annotate is consistent. This bias can be introduced by one person annotating with different methods across different images, and/or multiple users each using a different

annotation method. For example, one annotator delineating a large background feature may draw around an occluding foreground feature while another may not. We thus clearly define the astronomical structures to be annotated and detail their annotation protocol.

An annotated dataset of LSB images created using the designed tool is presented. For the annotations to be suitable for training supervised ML algorithms, annotations must be as accurate as possible. A common strategy to minimise uncertainty in annotations is to make sure every sample is annotated multiple times, thus making it easier to recognise anomalies through some statistical analysis. We collect annotations made using the presented annotation tool by four users of 227 LSB images from the MATLAS survey. As there are multiple annotations per image, each made by a different user, a method of combining these annotations into a single consensus annotation must be used. We review different strategies for combining 2D labels, and justify the use of a weighted majority voting method which takes the expertise of each user into account for combination.

To mitigate against dataset size limitations, it is common to first train ML models on datasets other than the pertaining data, a practice referred to as pretraining. Modern ML techniques such as CNNs require vast amounts of data to train into reliable inference systems. Currently the amount of LSB images is very limited, with the MATLAS survey containing around 200 and NGVS containing approximately another 200. While networks are most commonly pretrained on large related benchmark datasets such as ImageNet [51], it is possible to train on any data provided it contains some features in common with the target dataset. There are numerous examples in the application of ML to astronomy where models are either pretrained or entirely trained on real data [58, 60, 141] or even on synthesised data [18, 80, 160]. As there exists a small amount of suitable LSB data, it would be beneficial pretrain on synthesised images containing features resembling those exhibited in LSB images.

The rest of this chapter is organised as follows. In section 3.2, the creation of a tool that allows users to characterise different structures in multi-spectral images and precisely draw shapes representing their spatial profile is documented. The exact annotation process is detailed in Section 3.3, including the method for labelling each type of structure and the consensus protocol for duplicate annotations. Section 3.4 presents a method for synthesising LSB images, of which the resulting images can be used for pretraining CNNs.

3.2 Annotation Tool for Large Multi-spectral Images

In this section, a purpose designed tool for annotating LSB images is presented, illustrated in Figure 3.1. The nature of astronomical images presents challenges for 2D annotation. For natural images, segmentation labels are typically created by an annotator who man-

usually draw the envelope with predefined shapes over a static image. However, accurate annotation of the sky requires more information than a single static image can provide. The purpose designed tool must accommodate domain specific challenges while prioritising ease of use for expert annotators. There is a clear benefit to designing the tool as a web application, namely facilitation of collaboration and simultaneous annotation by multiple users while centralising the storage of images and user annotations. We first detail the exact protocol by which annotators delineate LSB images. The technical and user interface design decisions are then discussed.

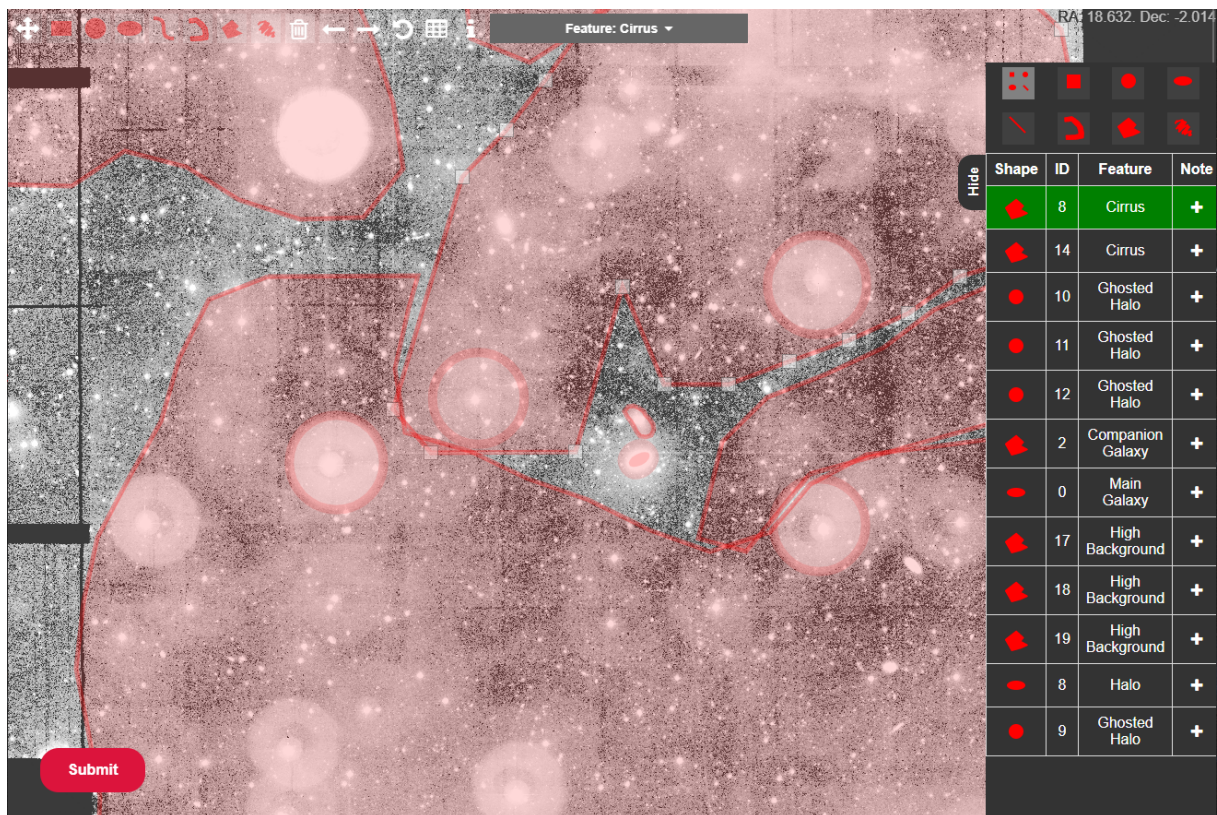


Figure 3.1: The annotation tool used to label contaminants and morphological features of NGC0448.

3.2.1 Annotation Process

In this study, the aim of each annotator is to precisely delineate the boundary of every galactic structure visible in a given image. In addition, annotators are to delineate features that are relevant to the study of these structures, such as contaminating objects. This list of structures includes stellar structures and contaminants. For clarity, we detail the full list of objects to be annotated in Table 3.1.

Feature	Description
Main galaxy	The central galaxy, including luminous features such as spiral arms.
Halo	The surrounding diffuse light of the main galaxy captured by LSB imaging.
Tidal tail	Stellar material propelled from the main galaxy during a major merger taking an elongated antennae-like geometry.
Plume	Stellar material propelled from the main galaxy during a major merger that does not exhibit typical tidal tail geometry.
Stream	Stellar material propelled from a progenitor other than the main galaxy, such as a companion galaxy.
Shell	Concentric circular sector-shaped features presenting typically in groups.
Companion galaxy	A nearby massive galaxy likely to be involved in a tidal interaction with the main galaxy.
Ghost halo	Artificial circular regions surrounding bright stars caused by reflections internal to the measuring instrument.
Cirrus	Dust clouds near or occluding the main galaxy presenting as diffuse structured regions often with a filamentary texture.
High background	Regions other than cirrus where background levels are noticeably high such that visualisation is impeded.
Instrument artefact	Any image artefacts other than ghost halos.
Satellite trail	The light trail left by a passing satellite.

Table 3.1: A detailed overview of structures to be annotated in LSB images and associated astronomical definitions consistent with [185].

We create a set of shape drawing tools and purposely restrict the shapes that can be used depending on the type of LSB structure being delineated. Annotators are able to choose from a predefined selection of shapes which they can use to draw over a suspected object in an LSB image. Different candidate objects in LSB images are assigned different shape tools which should be used to annotate them. Classes of LSB structures typically exhibit similar geometry. By creating a protocol where different categories of objects are delineated with certain shapes, variation in annotations or recall bias due to drawing style is minimised.

A summary of the exact use cases for each shape drawing tool is described in both Tables 3.2 and 3.3. Galaxies and thus their associated diffuse halos typically present with an elliptical geometry, thus are assigned the ellipse shape tool. Artificial ghost halos surrounding bright stars occur due to a predictable mechanism involving internal reflections within the measuring instrument and present exclusively as circles: these features are assigned the circle drawing tool. Streams, tails and plumes each present as a curved column of varying width. We assign the snake drawing tool these features as this allows the user to trace the path of the fine structure and then adjust the width at different sections of the shape. Edges of concentric shells follow a simple curve which can be modelled by

a low order polynomial curve, thus we assign the line drawing tool to these structures. Finally, it should be noted that structures occasionally present in geometries that cannot be captured by the assigned shapes. We allow annotators to annotate any feature with the region and freehand polygon drawing tools at their own discretion, to account for non-typical structures.

A good selection of shape types for annotators to draw over structures is important for the drawing tool. The shapes should be easy to draw and have predictable geometries to enable efficient annotation. Consideration must be given to the typical geometries of structures which will be annotated as this is useful prior knowledge which can be exploited to make annotation easier. For example, galaxies often present with an elliptical boundary, thus the ability to draw an ellipse is helpful, whereas a triangle fits no commonly expected geometry of any presenting fine galactic structure.

The final factors considered to mitigate against recall bias in this study are how object boundaries are decided and how occlusions are handled. In image analysis, an isophote refers to sections of an object with equal or typically approximate brightness. This concept is useful as it separates an object and its surrounding region into various sections of brightness, allowing object boundaries to be more clearly defined. In cases where there is an occluding object, the object's outer isophote may be disturbed, making this separation less clear. We first define the region of an object to be delineated along as the object's outer isophote, however, in the case of occlusion the annotator should delineate along a rough approximation of what the boundary would be. While this process is not perfect, it is in most cases reasonably accurate as objects often have predictable geometries. For further clarity, in the case where there is an occluding object that does not disturb the outer isophote, this definition of annotation is not affected. That is to say that the annotator should not remove annotated regions where an occluding object is enclosed by the target object's envelope.

3.2.2 Annotation Tool Considerations

Here we provide specifications of the tool so that reliable and useful annotations of LSB images can be generated. We then discuss requirements that the tool must fulfil to allow the user to sufficiently inspect astronomical images and considerations to increase quality of user experience.

3.2.2.1 Reducing uncertainty in annotations

Segmentation labels introduce a new dimension of variance/uncertainty into the annotation process. Semantic segmentation labelling inherits from object classification the same

uncertainty in exact categorisation, however there is also uncertainty in the geometric envelope of the 2D mask. Specifically in the case of LSB images, the boundary of an object is often highly ambiguous making it difficult to precisely outline the shape of a structure. To mitigate against uncertainty in ground truth labelling it is common to involve multiple experts in the annotation process so a consensus can be reached. It is key that the annotation tool recognises and logs different people using the annotation tool so that annotations can be separated by user, and so that a user does not make duplicate annotations of the same image. In addition, the tool should allow collaboration between users through the ability to compare each other's annotations. Finally, there should exist a functionality to separate users based on expertise, so that annotations made by users with more experience can be treated as more reliable in a systematic fashion.

3.2.2.2 World coordinate encoding

For the annotations to be useful to astronomers the annotations must be encoded in world coordinates so that drawn shapes align correctly with annotated structures. It is crucial that during annotation, the user changing their field of view should not disturb the real world coordinate location of annotations. This is both the case to improve user experience and to ensure that correct real world coordinate encoding of drawn shapes will align annotations with any astronomical image survey, enabling a statistical analysis of annotations on a variety of surveys.

The exact world coordinate encoding of drawn shapes must be carefully designed. It is necessary to first choose how to encode the rendered shape's rasterisation into a manner which is compatible with conversion to real coordinates. While it is possible to simply store the rasterisation itself and convert and store each pixel's location in world coordinates, this would result in slow performance as any change to the field of view requires world coordinate conversion of all pixels of all drawn shapes. This choice would also have significant data transfer overhead when an annotation is uploaded to the server. Though the user annotates on a Cartesian grid, their browser page, the annotated shapes' world coordinate encoding must exist on the curvilinear approximation of spherical space. In particular, polar distortion during the encoding process is a key concern, where the underlying distance in spherical space can be largely different between two sets of points on the Cartesian projection. This effect becomes severe at declination outside ± 60 deg, where at a reasonable field of view for annotation, the Cartesian distances between two points at the top of a browser page represents a much different curvilinear distance than two points at the bottom of the browser page. As a general rule, parameterisations of shapes involving distance are avoided where possible, such as using the width and height of a rectangle.

3.2.2.3 Viewing tool requirements

Contextual information from regions surrounding suspected structures is necessary: the user must be able to properly inspect large areas of surrounding regions. Proper visual inspection is dependent on local scale features as well as global scale. A direct consequence of these requirements is that images must cover a large region of the sky at high resolution. To facilitate these factors, the tool should allow the user to pan and zoom around the image, allowing them to easily change their field of view. In addition, the region of sky covered by an image is encoded in world coordinates, which should be visible to the user.

Structures appear differently across the wavelength spectrum, thus the user must be able to inspect all available wavelength bands. A given survey contains multiple images of the same region each capturing a different wavelength band. It is vital that the user is able to switch between these images in order to assess the presence and structural properties of a suspected object. In many cases the combination of two or more wavelengths is also helpful for visual inspection, for example a composite band of the difference between two bands or even an RGB mapping of three bands or composite bands. The annotation tool should facilitate inspection across wavelength bands, and while retaining the world coordinate location of the user's field of view.

To ensure a quality user experience it is important that data transfer is made efficient wherever possible. As previously discussed, LSB images requiring annotation by this purpose designed tool cover large regions of the sky at high resolution. In practice, images can approach sizes of 10000×10000 pixels and occupy up to 800MB. Pyramidal image formatting is an effective solution to efficiently transfer large high resolution images. In short, the base image is downsampled proportionally to the number of pixels in the user's field of view, before file transfer. For example, suppose the user's field of view is an entire 1080p screen containing 1920×1080 pixels and the user wishes to view an entire image of resolution 10000×5000 , then the image can be downsampled by an approximate factor of 5 with a minor loss of detail, reducing the amount of data to be transferred by 25 times. As the user changes their field of view through panning and zooming, different sections of the pyramid formatted image are loaded so that the region and its associated downscaling factor are dynamically changed. To continue the example, suppose the user zooms into the image to view a region of 2000×1000 pixels, then this section of the image is transferred with minimal downscaling, so the user can study the local scale features of the image while only having transferred the exact section of the image that is needed.

3.2.3 Interactive Viewing of Astronomical Images with Aladin Lite

To fulfil the requirements for proper visual inspection, we integrate Aladin Lite (AL) [24], an interactive visualisation tool for astronomy images, into the annotation tool. A benefit of choosing to create the annotation tool as a web application is that third party components can easily be integrated. Integration through embedding is relatively simple: web applications can be encoded as a single HTML element which can be plugged in to any location of a web page. Such embedding is directly supported by AL, allowing the application to be integrated and customised to fit the overall requirements of the annotation tool. The ability to zoom and pan across an image is provided by AL, where the user can click and drag to pan and use the mouse’s scroll wheel to zoom in and out. Real world coordinates of the displayed image are tracked as the user changes their field of view. AL also offers the ability to convert between pixel space and real world coordinate space. This functionality makes it a suitable choice for integration, as how annotated shapes are rendered requires the exact pixel location of points in the user’s field of view.

Aladin Lite exclusively works with images of pyramidal formatting, ensuring data transfer is minimised. AL requires that images are encoded as Hierarchical Progressive Surveys (HiPS) [71], a multi-resolution data structure for astronomical images. There exists tools for converting commonly used astronomical image formats such as FITS into HiPS, allowing custom surveys to be generated and displayed in AL. AL also provides the ability to switch between surveys without refreshing the web browser page or disturbing the user’s real coordinate field of view. This means that if a HiPS image is generated for each wavelength band or composite band for an LSB image requiring annotation, then the user can easily switch between bands through AL. Furthermore, HiPS support RGB mapping where a combination of bands/composite bands can be mapped to colour space, which can then be displayed on AL.

To maximise user experience, the configuration of Aladin Lite is customised upon integration so that only parts that are pertinent to the annotator remain in view. Figure 3.2 shows the default layout and configuration of Aladin Lite. Of the elements provided in Aladin Lite’s interface, the annotation tool only retains the real coordinate display, centre reticle, manual zoom buttons and survey selector button. The ability to change image surveys is of high importance, so is retained though it is positioned next to a drawing/viewing toggle button which deactivates the AL’s zooming and panning and activates the drawing tool. Real world coordinates and the centre reticle are kept as they provide key positional context to expert annotators. The manual zoom buttons are kept to ensure the tool’s compatibility to mice without scroll wheels. The annotation tool, and



Figure 3.2: The Aladin Lite tool.

thus AL’s visualisation, covers the entire user’s browser page to maximise the amount of image data present on the user’s screen. Interface buttons are kept opaque so that pixels underneath buttons are still visible. Text displays such as the real world coordinates are rendered using font colouring opposite to the underlying pixels, to ensure that text is readable despite the colour of underlying objects. The exact configuration of AL in our annotation tool is displayed in Figure 3.3

3.2.4 Precise Delineation of Galactic Structures

Creating a dataset of segmentation labels outlining the envelopes of galactic structures requires thorough planning to ensure that labels are sufficient for the required use cases. Any astronomy focused analysis of the annotations requires that the world coordinate locations of annotations is somehow saved along with the geometry of drawn shapes. In addition, the shapes that an annotator is able to draw and the exact functionality of the shape drawing interaction must be designed so that annotators can precisely and efficiently outline fine galactic structures. Finally, upon being uploaded to a central server, annotations must be organised into a coherent data structure.

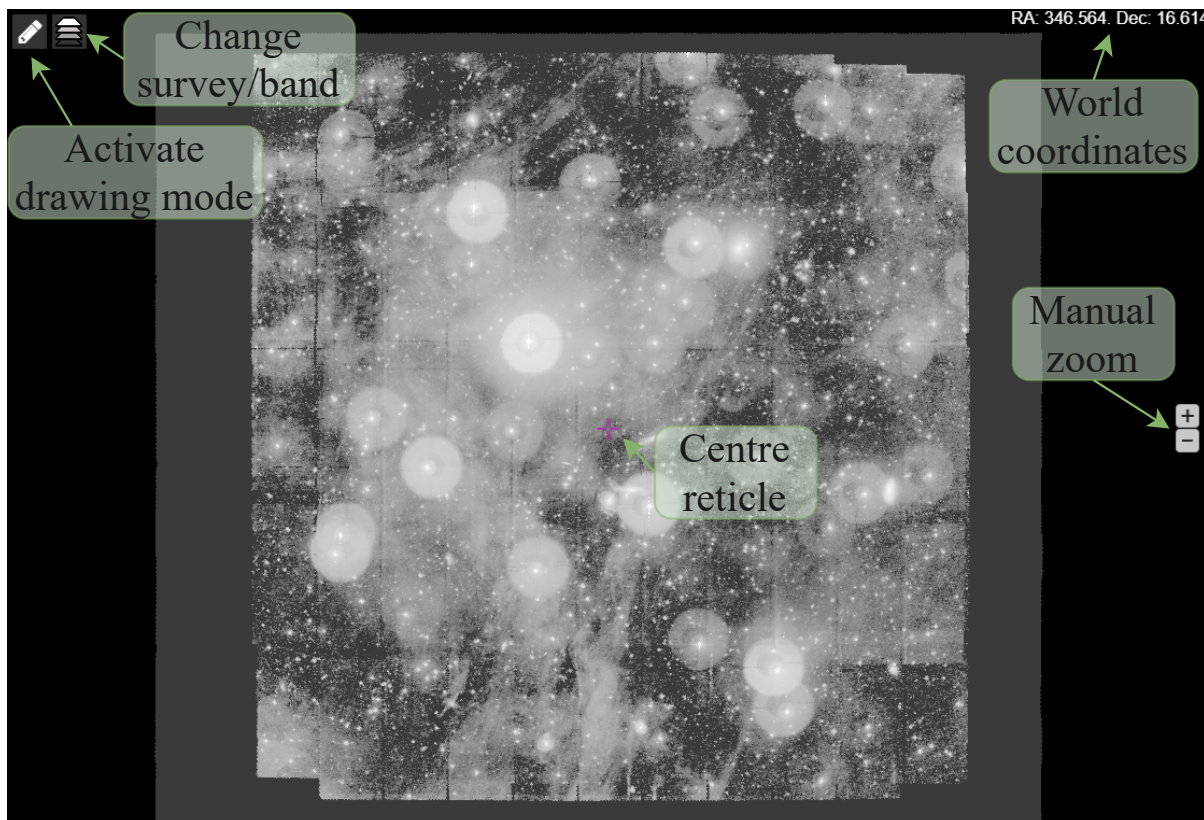


Figure 3.3: Aladin Lite integrated into the annotation tool.

3.2.4.1 Drawing tool

To enable annotators to delineate objects in astronomical images, we design a drawing tool allowing shapes to be drawn over images. In the drawing tool we model shapes mostly as polygons, i.e. a collection of vertices joined with straight lines with the enclosed area filled in. There is a toggle enabling the user to switch between the drawing tool and viewing tool. Drawn shapes remain rendered regardless of the mode toggled. Further, panning and zooming in the viewing tool accordingly adjusts the position and size of drawn shapes. There exists a selection of buttons for different shape types and annotation management features in the top left of the drawing tool, as shown in Figure 3.4 with buttons labelled as B1 to 15, with visual characteristics such as size and opacity similar to buttons in the viewing tool. Actions performed by each button can also be activated by hotkeys; hovering over a button displays the hotkey required to activate the button.

A collection of basic shapes form the foundation of the drawing tool. The geometry of many galactic structures can be parameterised by simple shapes such as circles and ellipses. These shapes are detailed in Table 3.2 and illustrated in Figure 3.5. The vertices used to encode rectangles, circles and ellipses represent the corners and edge midpoints

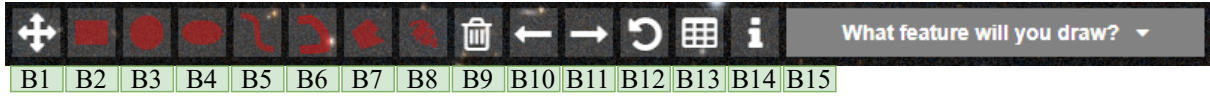


Figure 3.4: Buttons used in the drawing tool: B1 activates viewing mode; B2-8 activate different drawing shape tools; B9-12 manage drawn shapes; B13 opens a table of drawn shapes; B14 shows examples of different features; B15 allows the classification of a drawn shape.

Shape	$ V $	Interaction	Use case	Hotkey	Button
Rectangle	8	A rectangle is formed with p, q as opposing corners.	Contaminants	R	B2
Circle	8	A circle is formed with centre p and radius $d(p, q)$.	Ghosted halos	C	B3
Ellipse	8	An ellipse is formed inside the bounding rectangle with p, q as opposing corners.	Galaxies Diffuse halos	E	B4
Line	4	A straight line is formed between p and q . Two intermediate vertices at a third and two thirds of the line's length can then be altered to curve the line.	Shells	L	B5

Table 3.2: Detailed overview of basic shapes available to draw in the annotation tool. The interaction of the drawing process is described given two user generated points p and q . The number of vertices used to encode the shape is denoted as $|V|$. Euclidean distance is represented as d . Buttons correspond to labels of Figure 3.4.

of the rectangle/bounding rectangle. Lines are modelled with a third order Bezier curve containing four vertices, p_0, p_1, p_2, p_3 , where any point along the curve is given by $P(t) = (1-t)^3 p_0 + 3(1-t)^2 t p_1 + 3(1-t)t^2 p_2 + t^3 p_3$, with $t \in [0, 1]$ as a parameter that represents normalised distance along the curve. Encoding shapes with only vertices ensures that curvilinear projection is not distorted at extreme declinations, as described in the previous section.

The ability to draw complex polygon shapes ensure that any object can be annotated. While basic shapes parameterise a large range of objects, objects often do not fit these geometries either because their geometry is non-typical or their geometry is inherently more complex. The annotation tool provides three ways to draw polygons with an unlimited number of vertices, detailed in Table 3.3 and illustrated in Figure 3.7. After polygon vertices have been generated from the selected drawing tool, the enclosed space is filled similar to basic shapes.

All shapes are drawn by the user with a click and drag approach. This involves the

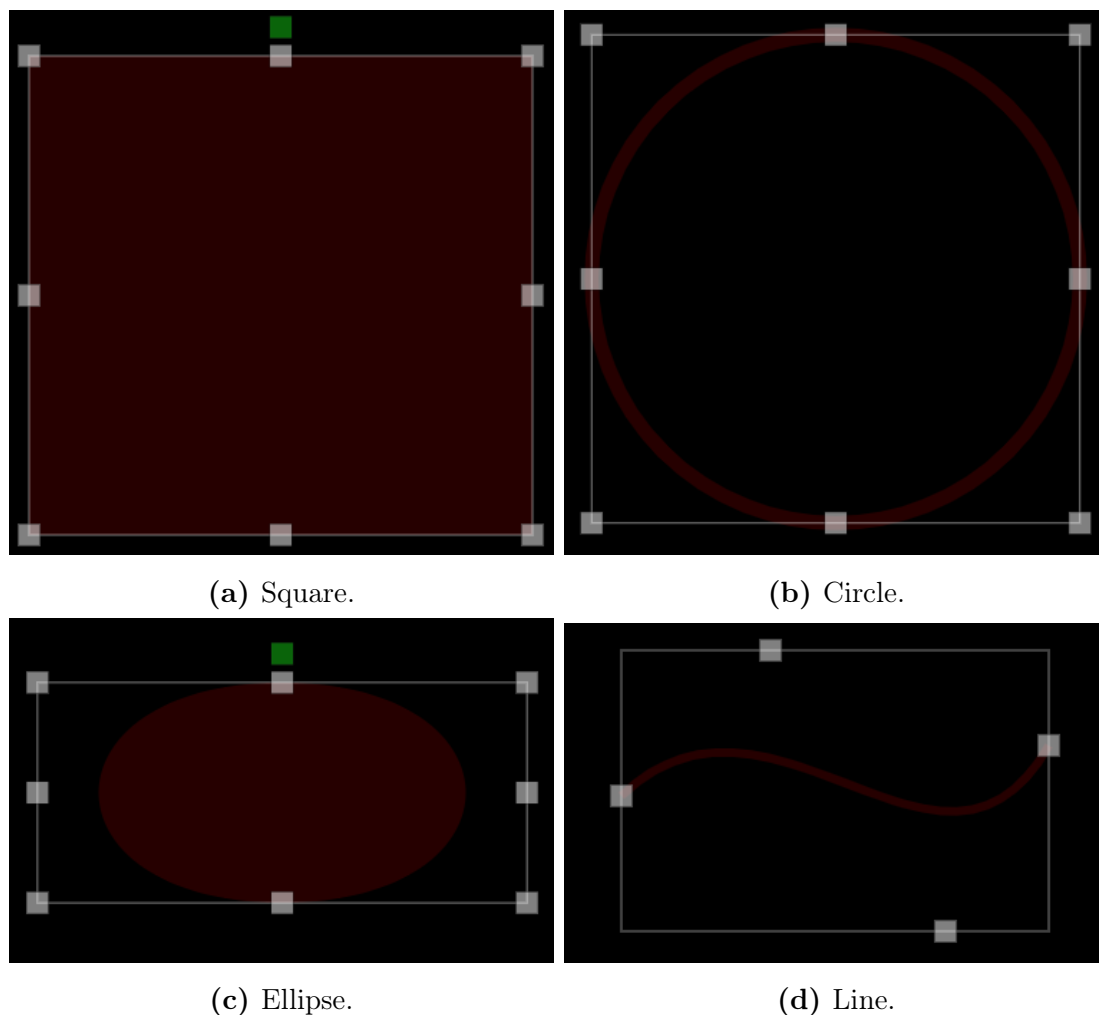


Figure 3.5: Examples of each basic shape in the drawing tool. Each shape is in a "selected" state so that the associated amendment boxes are visible. Note that amendment boxes do not always correspond with bounding boxes, due to how Bezier curves (which ellipses also use in rendering calculations) are calculated from user generated points.

user clicking down at point p on the image, dragging their cursor to control the shape, and releasing the click at point q to finish the shape. The interaction process for basic shapes is described in Table 3.2. User interactions for each polygon tool follow a similar process. A trail of points $p_i = p_0, \dots, p_n$ is saved following the user's cursor, where $p_0 = p$ and $p_n = q$. Intermediate points p_i for $1 \leq i \leq n - 1$ are saved if the distance between the cursor and the previous point p_{i-1} covers 25 pixels. Each tool then calculates vertices as described in the 'Interaction' column of Table 3.3.

Each generated vertex is displayed to the user as an 'amendment box'. Amendment boxes can be clicked and dragged to change the location of the shape's underlying vertex, thus allowing the user to edit the shape after it has been drawn. In addition, some shapes

Shape	$ V $	Interaction	Use case	Hotkey	Button
Snake	$2n$	Vertices are calculated as 10 pixels away from each point p_i in each direction perpendicular to the direction from p_i to p_{i-1} .	Streams Tails Plumes	S	B6
Region	n	Each point p_i corresponds to a vertex of the polygon, with p_0 and p_n joined to enclose the shape.	Any	A	B7
Freehand	$\leq n$	The convex hull of all points p_i is computed.	Any	F	B8

Table 3.3: Detailed overview of polygon drawing tools. The method for calculating shape vertices V is described given a set of user generated points p_0, \dots, p_n . Buttons correspond to labels of Figure 3.4.

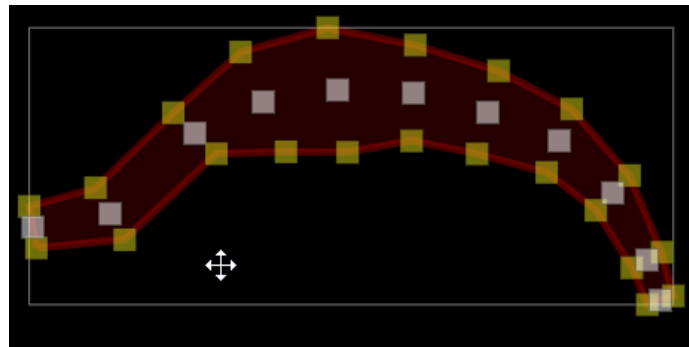
have special amendment boxes allowing the user to change multiple vertices through some parameterisation. Ellipses and rectangles contain a green amendment box drawn above the highest mid-point vertex, which rotates the entire shape around its centre. Snakes contain multiple yellow amendment boxes allowing the user to change the width of a section of the shape. Figures 3.5 and 3.7 show some example shapes with amendment boxes displayed.

In this study, we are not only interested in the delineation of structures, but also the type of each structure present in a given image. It is important that this classification information is retained along with the drawn shapes, so that statistical analysis of the resulting annotated segmentations by type of structure is possible. When a shape is selected the user is able to assign a type of structure to it, via a dropdown list of possible objects which the user can use to classify an annotated object, shown in Figure 3.4 as button 15.

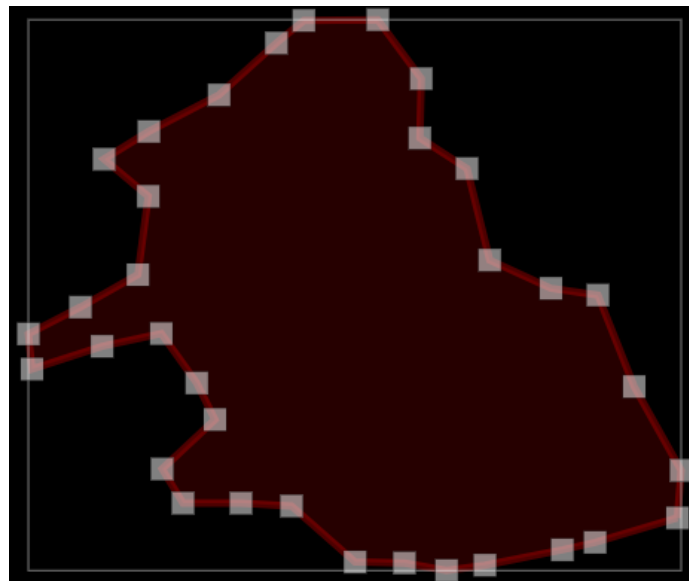
When the user has completed their annotation, it is possible to submit and upload it to the central server. Submission requires that all drawn shapes have been classified: there is a ‘not sure’ classification option in the case where a user is confident there is a structure but is unsure of the classification. This submission process uploads each shape’s vertices in both world coordinates and pixel coordinates at time of submission along with their classification and note (if added), and details of the user that performed the annotation. By storing user details in addition to the annotation, it is possible to later weight annotations differently depending on the expertise of the user that performed it.

3.2.4.2 User Experience Features

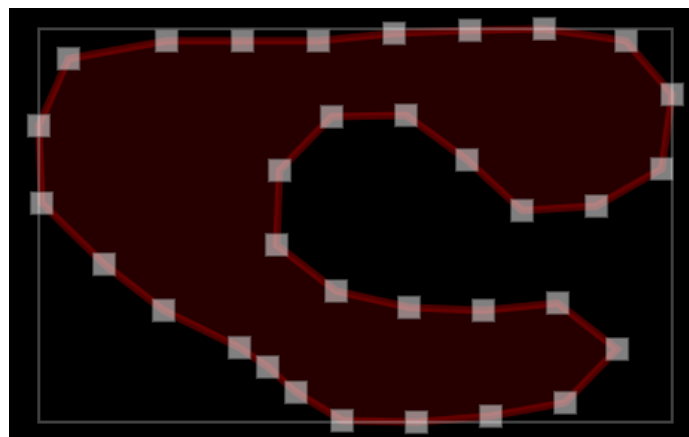
To improve user experience, there are additional functionalities that make editing annotated shapes easier. Amendment boxes appear after a shape has been drawn but disappear when the user deselects the shape. This process occurs either when a location other than



(a) Snake.



(b) Freehand.



(c) Region.

Figure 3.6: Examples of each complex polygon shape in the drawing tool. Each shape is in a "selected" state so that the associated amendment boxes are visible.

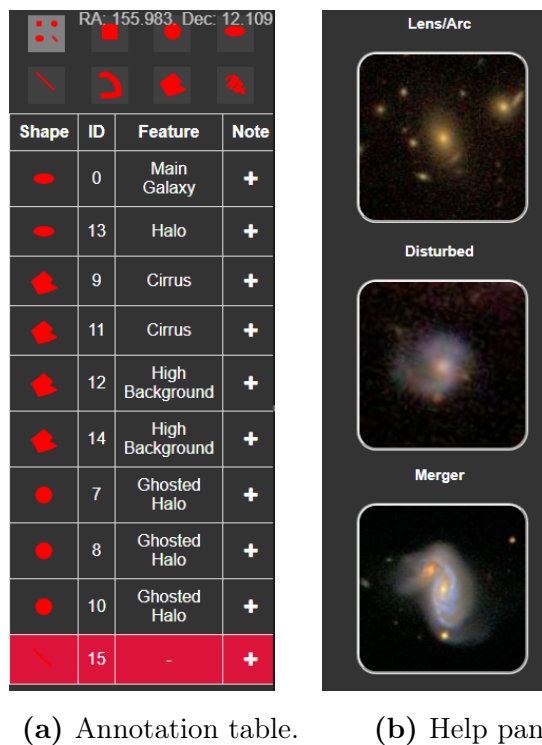


Figure 3.7: Examples of each complex polygon shape in the drawing tool. Each shape is in a ‘selected’ state so that the associated amendment boxes are visible.

the shape’s bounding box is clicked or the viewing tool is activated. A shape can be re-selected by clicking anywhere in the shape’s enclosed space, causing amendment boxes to reappear. A shape can be clicked and dragged to move the entire shape, if selected the entire bounding box can be dragged, and if unselected anywhere in the shape’s enclosed space can be dragged. If selected, an entire shape can be deleted. Throughout the annotation process, a stack of user operations is saved, allowing undo and redo of all operations.

In addition to the already discussed features, there exist several features for managing drawn shapes and providing assistance. A help panel can be revealed (see Figure 3.7b), containing example images of different objects. A table of annotation can be revealed which lists all drawn shapes, shown in Figure 3.7a. The table allows the user to write a note about object which is saved into the annotation record. Selection of shapes in the drawing tool can also be performed through the annotation table by clicking a shape’s corresponding row. The user can also filter the table’s listed shapes by shape type. Finally, if a shape has been drawn but not classified, yet the user attempts to submit the annotation, the shape is highlighted red in the annotation table, as shown in Fig 3.7a.

3.2.4.3 Database design

Data produced by the annotation tool must be organised in some fashion that ensures data integrity and reduces data redundancy. It is vital that the ontological separation of data is sufficiently designed so that unintended and unwanted changes to data, for example through some processing error, are minimised. We organise data using a relational database model, where data is divided into tables which may be connected through some relation. The quality of organisation can then be enforced through normalisation. We design the database schema to conform to Boyce-Codd normal form. We separate data into four relations: **users**, **galaxies**, **annotations** and **shapes**, as is shown in Figure 3.8. Briefly, users and galaxies are connected to annotations by a one-to-many relation, as a galaxy can be annotated multiple times by multiple users. Then, annotations are connected to shapes by a one-to-many relation, as a annotation is composed of multiple shapes. Exact implementation details can be found in appendix A.1

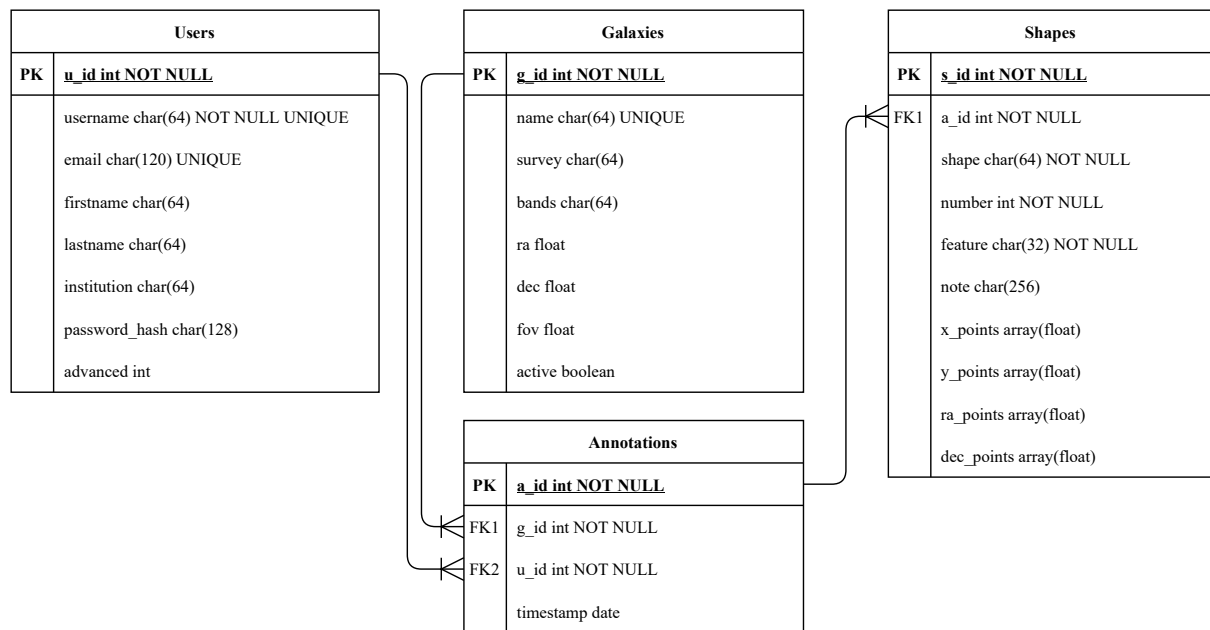


Figure 3.8: Entity relationship diagram of the database schema designed for storing annotation tool data.

3.3 Annotated Dataset of LSB Images

In this section, a dataset of annotated LSB images for training supervised machine learning algorithms is detailed. Firstly, it is important that annotators have a consistent approach to annotation to mitigate against recall bias. We discuss the exact annotation process

used to classify and delineate LSB images. Second, we detail properties of the annotated LSB dataset generated using the annotation tool presented in this chapter. Finally, typical supervised ML training protocols make use of a single target sample per input sample. We describe the process used to combine annotations from multiple users into a single consensus annotation.

3.3.1 Annotated Dataset

In this section, we describe the dataset of LSB images and associated collected annotations that are used in this thesis. As previously mentioned, training supervised ML models capable of automating the classification and segmentation process requires both input and ground truth examples. We first detail the images to be used as input. Following this, the collected annotations and their characteristics are presented and discussed.

3.3.1.1 Images

In this project we use images from the Mass Assembly of early-Type GaLaxies with their fine Structures survey (MATLAS [61]), captured by the Canada-France-Hawaii Telescope’s (CFHT) MegaCam instrument. These images prior to preprocessing cover a world coordinate region of $1^\circ \times 1^\circ$ with a resolution of 0.187 arcsecond¹ per pixel, where the centre of each image targets a galaxy with suspected LSB fine structures. Preprocessing then involves removing outer sections of the images suffering from large gaps due to instrument artefacts and downsizing the images by a factor of 3 in order to enhance fainter structures, resulting in an image of approximate average size 6000×6000 . Captured images are multispectral, with a given region being captured in up to three wavelength bands, g , r and i . However, typically most images are composed of just the g and r bands, corresponding to wavelength sections around 464nm and 658nm, respectively. In total, in this thesis we use 227 MATLAS images of which there a combination of mostly multispectral images and occasionally images of a single spectrum band. Specifically, in total there are 199 g -band, 186 r -band and 68 i -band spectral images. In addition, for annotation an intensity scaling transformation is applied to the image where fainter pixels are increased relative to brighter pictures, in order to further enhance fainter structures. Specifically, inverse hyperbolic sine scaling is applied to the images according to $x' = \text{arcsinh}(x)$, where x' and x represent the scaled output and input images, respectively.

¹Equal to 1/3600 of a degree.

Feature	d	Images with $d \geq 1$
Main galaxy	655	227
Halo	605	216
Tidal tail	172	46
Plume	103	43
Stream	111	47
Shell	217	36
Companion galaxy	549	130
Ghost halo	2657	224
Cirrus	255	59
High background	915	219
Instrument artefact	15	9
Satellite trail	18	13

Table 3.4: Summary of annotated features. The number of annotated features is denoted as d .

3.3.1.2 Annotations

Using the annotation tool presented in this chapter, a total of 655 annotations of MATLAS LSB images were generated. These annotations include a total of 6573 drawn features, of which an overview is shown in Table 3.4. Annotations were produced by four annotators of varying expertise, which we denote as users 1 and 2, who are experts, and users 3 and 4, who are non-experts. Users 1 and 4 annotated all 227 images, user 3 annotated 183 images and user 2 annotated 18 images. We note that user 2 has contributed a far smaller number of annotations in comparison to other users, which may cause a small subset of annotations to contain different labelling characteristics than the rest of the dataset. However, the increased confidence in these annotated labels, due to a larger pool of annotators, likely outweighs any negative effect on the dataset caused by this potential bias. All structures were consistently annotated as described in Section 3.2.1. A detailed analysis of the annotations can be found in [185], which additionally includes annotations made using the tool on images from surveys other than MATLAS.

3.3.2 Computing a Consensus

In order to train supervised machine learning models, it is necessary to decide how LSB images with multiple annotations made by different users are handled by the training protocol. Each single annotation has an associated uncertainty in various factors, such as the categorisation or the delineation of a feature. Making use of all available annotations is incredibly important as this mitigates against human error made during the annotation process, thus reducing the uncertainty. A naive approach to fulfilling this goal where

the model is trained on each input-annotation pair may be detrimental to performance due to contradicting examples, in addition to being computationally inefficient. There have recently been efforts to extend the supervised learning training loop to use multiple annotations per input as target data, referred to as learning from a crowd. Yan et al. [214] develop a probabilistic model to learn the ground truth from multiple annotators, where the expertise level of different annotators is also learned. Albarqouni et al. [5] integrate a similar approach into an active learning framework, where uncertain labels are fed back to annotators to ‘double-check’. Rodrigues and Pereira [170] propose a ‘crowd layer’ which learns annotator weightings through an expectation-maximisation algorithm. While these works offer promising approaches to handling multiple ground-truths in an adaptive fashion, all efforts are limited to data with just classification labels.

In datasets with 2D annotated labels from multiple humans per sample [6, 9, 10, 42, 134, 146], the most common approach to combine annotations into a single consensus is through majority voting [115, 134, 137]. The FreeSurfer method [73, 74] and STAPLE method [204] both propose expectation-maximisation algorithms to compute a single consensus annotation from multiple 2D annotations. Both methods have been used extensively in medical imaging literature [52, 102, 202], however comparative studies have shown that such more complex approaches do not reliably generate more accurate segmentation consensuses than majority voting [11, 134, 174]. This is particularly the case in data where there is a hierarchical relationship between label classes [147]. Given these factors, we combine all available annotations for a given image into a single consensus annotation using a weighted majority voting framework. The resulting consensus is then represented as a probability map image where a high probability pixel corresponds to a pixel where annotators concur on this annotated pixel.

Prior to the consensus process, all annotations are converted into pixel-wise binary masks m_i , with identical resolution to the associated annotated image. The consensus c for a given image is then defined as the weighted average of these mask,

$$c = \frac{\sum_{i \in I} w_i m_i}{\sum_{i \in I} w_i}, \quad (3.1)$$

where I denotes the available annotations for a given image. While most shapes are either polygons or enclose relatively large areas such as circles and ellipses, curved lines fall outside this property which must be considered. For these such shapes where there is little overlap in the annotated regions, we simply take the intersection of all annotations as a compromise. In this consensus framework, pixels with values greater than 0.5 are defined as a majority consensus.

The exact combination of annotations into a single consensus must be carefully chosen.

The varying expertise of each user must be considered in the consensus, as less experienced annotators may be more likely to make incorrect annotations. This is achieved by first assigning a weight to each user, w_i where a larger weight represents a more experienced annotator. The majority of images in our dataset (169) were annotated by one expert and two non-experts, 40 were annotated by one expert and one non-expert, 14 were annotated by all annotators, and 4 were annotated by two experts and one non-expert. From these combinations, we establish a set of conditions to decide whether a pixel is set as positive:

- In the case where an image was annotated by one expert, a pixel is positive if
 - it is marked by the expert, or
 - it is marked by two non-experts.
- In the case where an image was annotated by two experts, a pixel is positive if
 - it is marked by two experts, or
 - it is marked by an expert and a non-expert.

We then use a set of weights which satisfies these scenarios, i.e. for each scenario, respectively,

$$w_1 \geq w_3 + w_4 \quad w_3 + w_4 \geq w_1 \quad w_1 + w_2 \geq w_3 + w_4 \quad w_1 + w_3 \geq w_2 + w_4 \quad (3.2)$$

The first two equations naturally lead to an equality, with $w_1 = w_3 + w_4$. Assuming experts each have equal weights, and non-experts each have equal weights, it can be seen that we must set $w_1 = w_2 = 2$ for experts and $w_3 = w_4 = 1$ for non-experts. Example annotations and their consensuses are shown in Figures 3.9 and 3.10.

3.4 Synthesising Galactic Cirrus for Pretraining

Modern machine learning algorithms such as CNNs require massive amounts of data in order to generalise well on a given task. For example, the cornerstone work of Krizhevsky et al. [117] trains on ImageNet [51], consisting of 1.2 million images. A major challenge in applying deep learning techniques to LSB data is the limited sample size: in order to effectively tackle the problem, it is necessary to address this limitation. A common approach in mitigating against limited dataset size when using modern ML techniques is that of pretraining for transfer learning. When dealing with a target dataset limited in size, pretraining describes the process of training a model on a dataset related in some way to the target dataset. This pretrained model thus learns features that are relevant to

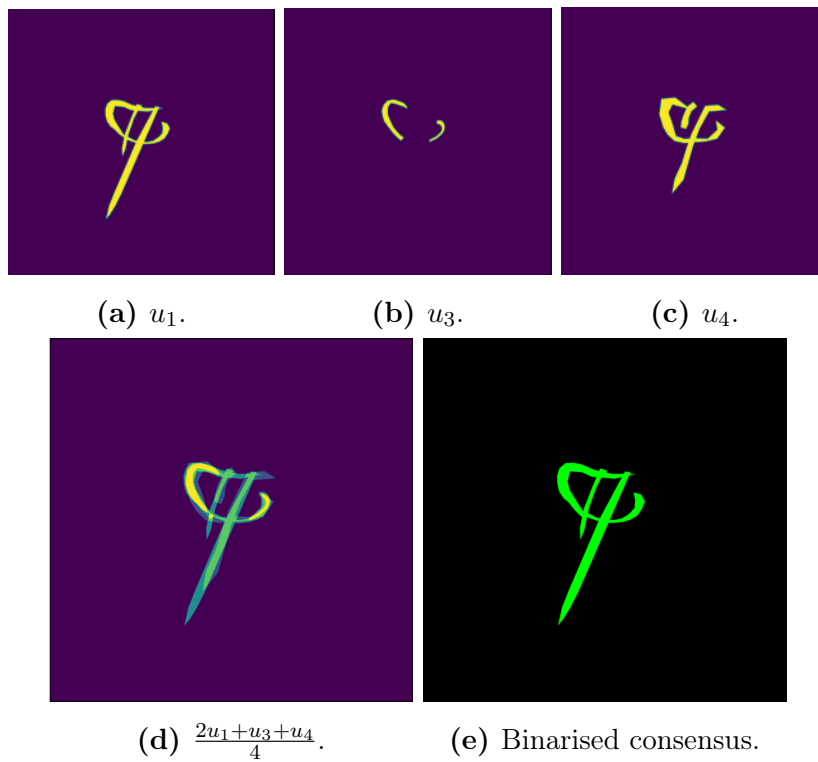


Figure 3.9: Annotations of streams on NGC0474 by three users, u_1 , u_3 and u_4 .

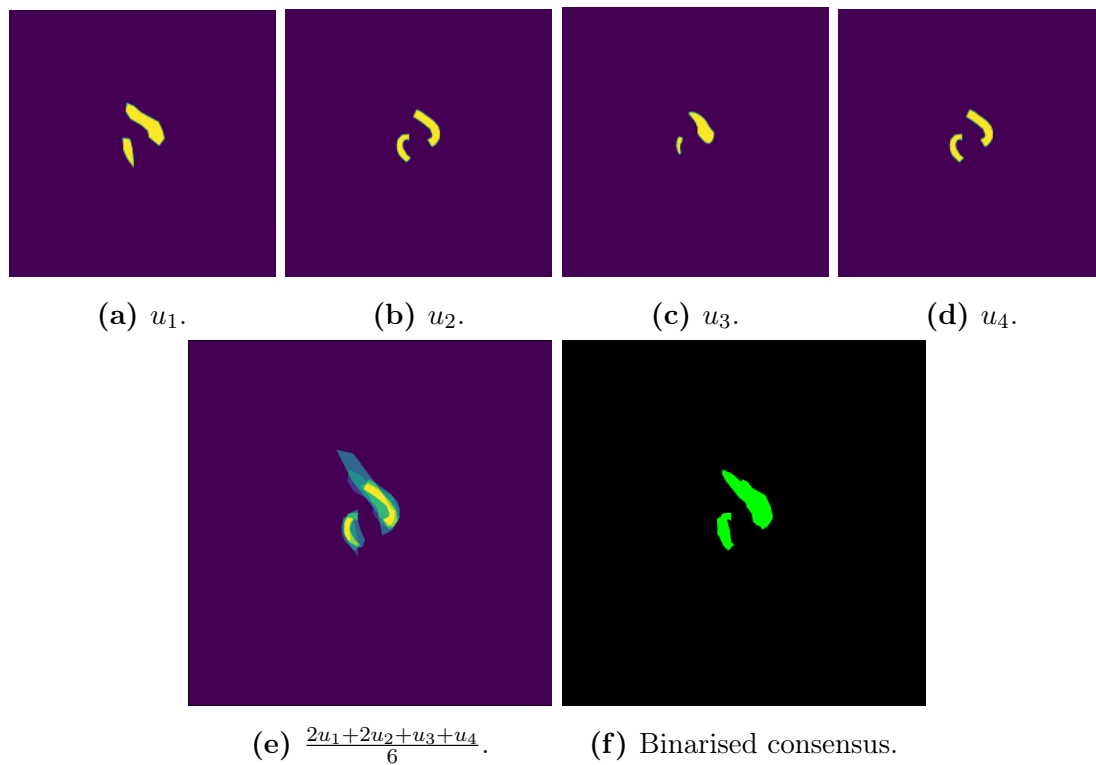


Figure 3.10: Annotations of tidal tails on NGC4270 by four users, u_1 , u_2 , u_3 and u_4 .

the target task, and can then be trained on the target dataset to either relearn the final mapping of these features to output, or use these features as an initialisation point. This process of transferring related learned features is described as transfer learning.

LSB images are commonly contaminated with galactic cirrus structures, which the trained ML model must be able to properly process despite limited sample size. Typically astronomical images in the optical spectrum do not contain cirrus structures as galactic cirrus does not emit these wavelengths. However, the sensitivity of LSB imaging captures scattered light from dust in cirrus clouds, even in the optical band. In future surveys such as Euclid massive amounts of cirrus will be uncovered relative to what is currently seen in optical wavelength images. The ability for ML models to handle contamination of LSB images by galactic cirrus is vital, though currently samples are limited. For example, Table A.3 shows that only 59 MATLAS images contain galactic cirrus, of which an even smaller portion contain strong cirrus contamination.

In this section we present a synthesised dataset of galactic cirrus images for pretraining ML models. We synthetically replicate structural patterns present in real LSB images, resulting in images with similar properties to the target LSB dataset. By combining suitable noise models with carefully chosen parameters, it is possible to create a synthesised dataset with arbitrary resolution and sample size. ML models can be trained on this dataset in order to learn a foundation of features necessary to process LSB images and in particular galactic cirrus, which can be transferred to the target dataset.

3.4.1 Constructing Synthetic Cirrus Images

In order to create images exhibiting discriminative features similar to real images of galactic cirrus, multiple noise patterns are combined. All synthesised images are formed from at least three parts, background B , cirrus C , and bright regions R , shown in Figure 3.11. This dataset contains two problem scenarios, a segmentation scenario where the goal of an ML model is to segment the cirrus structures, and a denoising scenario where the ML model must remove the cirrus structures while preserving underlying objects.

The background B attempts to create a canvas with a non-zero background level and numerous faint background objects, as shown in Figure 3.11a. To achieve this, pixels are drawn from a Gaussian distribution $B^* = \mathcal{N}(\mu, \sigma) = \mathcal{N}(0, 0.01)$ and then normalised between 0 and 1. Following this pixels are inverted according to $B = 1 - B^*$.

The cirrus component C , shown in Figure 3.11b, contains textured cloud shapes with smooth boundaries. The cloud shapes C_{shapes} (see Figure 3.12a) are produced by a 2D Gaussian mixture model (GMM) with 13 randomly located components with standard deviation proportional to the image resolution. A binary mask M (see Figure 3.12b)

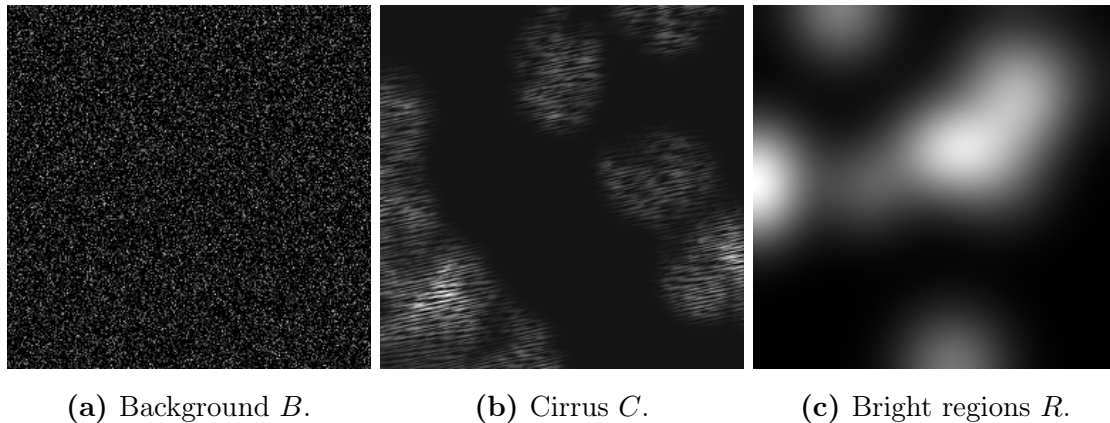


Figure 3.11: Examples components used to create synthesised images.

extracted from this GMM by setting values greater than 10^{-5} as 1 and others as 0. The cirrus texture is created by combining a cloud texture C_{cloud} (Fig. 3.13a) and a streak texture C_{streak} (Fig. 3.13b), both generated from Perlin gradient noise of varying frequencies as shown in Figure 3.13c. These components are combined according to

$$C = (C_{\text{cloud}} + \frac{C_{\text{cloud}}C_{\text{streak}}}{4})C_{\text{shapes}}\mathcal{B}(M), \quad (3.3)$$

where \mathcal{B} (see Figure 3.12c and 3.12d) is a blurring function which convolves a Gaussian kernel with size proportional to the image resolution. The binary mask M forms the segmentation target for these synthetic cirrus structures.

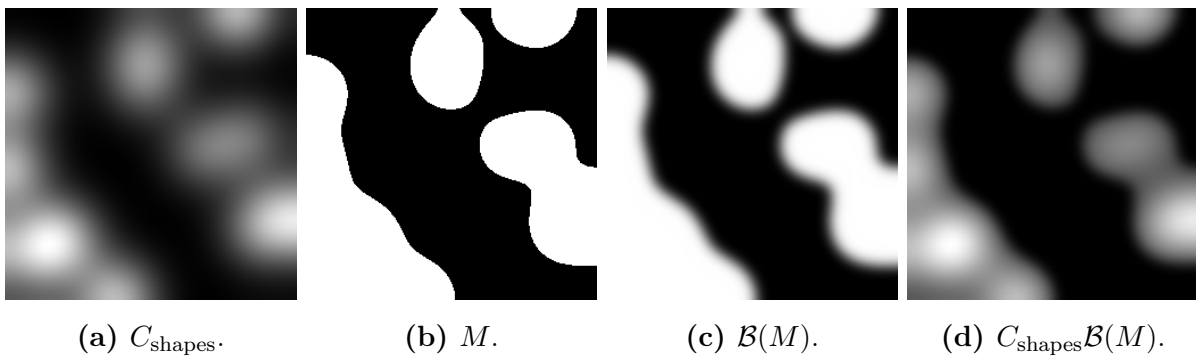


Figure 3.12: Components generated from a GMM used to shape the cirrus structure.

Bright regions R , shown in Figure 3.11c, are smooth isotropic bright regions resembling regions of diffuse light surrounding low surface brightness galaxies. These are created from a GMM with a similar process to the cirrus case. In this GMM however, the number of modes varies uniformly between 3 and 17, and standard deviations of each Gaussian mode are varied randomly by $\pm 20\%$. This randomisation increases the difficulty of separating bright regions from cirrus structures, and aims to guide an ML model to focus on the

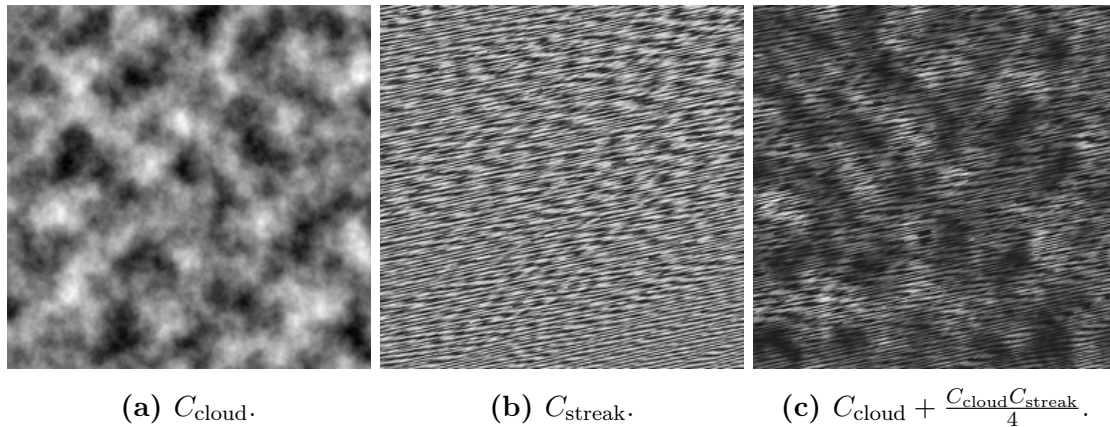


Figure 3.13: Example textures used to create the cirrus structure.

texture of the cirrus rather than the envelope or intensity level relative to the background.

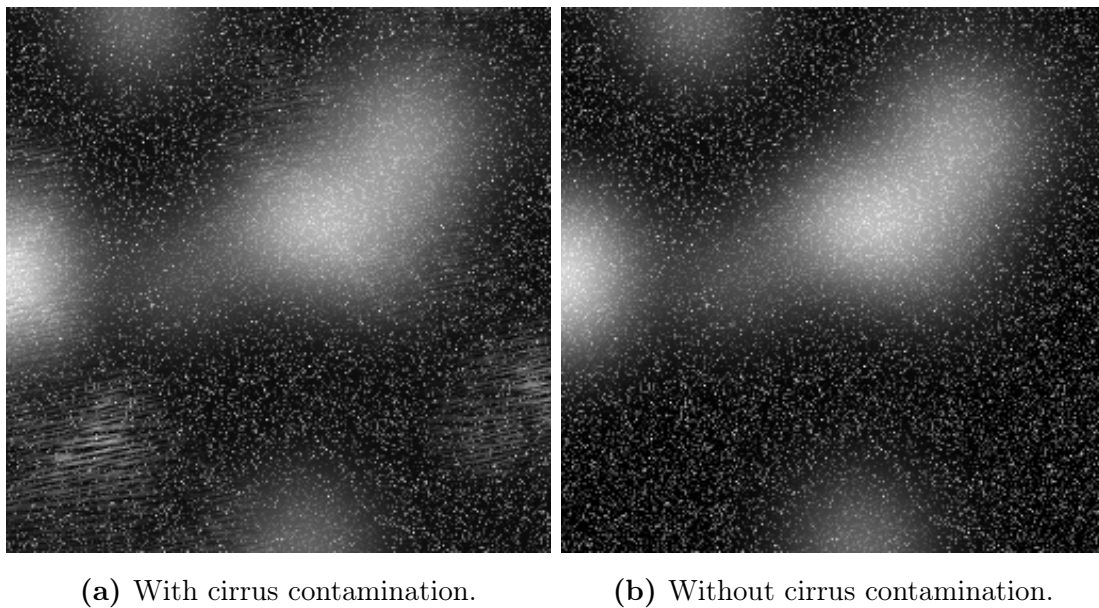
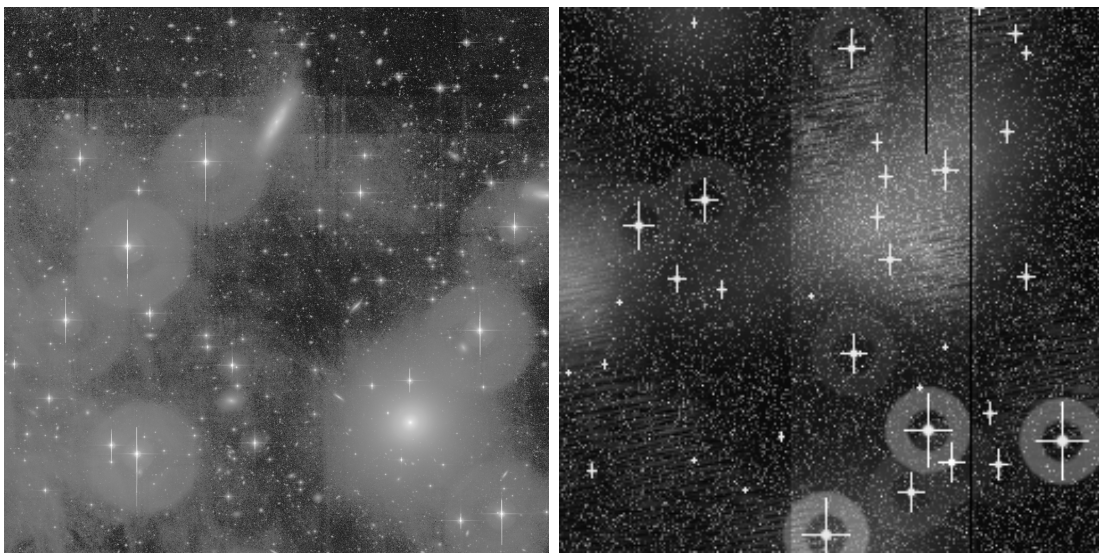


Figure 3.14: The final resulting output and its associated denoising target.

Finally, all parts are combined according to $\gamma B + \gamma C + R$ with a target not contaminated with cirrus set to $\gamma B + R$, before normalisation between 0 and 1, where γ balances bright regions vs. background and cirrus regions. All figures throughout this chapter of synthesised cirrus images use $\gamma = 0.4$. An example of the final output and its associated target is shown in Figure 3.14. As images are composed with noise distributions in a procedural fashion, extending images to arbitrary resolution is trivial. In addition, this synthesis method allows components of the image to be easily changed for example in intensity or texture, meaning that the dataset can be adapted to best suit the specific LSB image experiment.

3.4.2 Increasing realness

We design the dataset to have variations of differing realness, in order to enhance pre-training benefits. Real LSB images contain a variety of objects non related to this thesis' study, which the ML model must be able to properly process. By introducing the model to such objects in this synthesised data, the model may be able to generalise more effectively to real data. An added benefit of these variations is the ability to perform comparative analysis between different ML models. Ablating portions of the dataset before training will demonstrate characteristics of the trained model.



(a) A real LSB image containing NGC0532 (b) The combination of all difficulty variations on a synthesised LSB image.

Figure 3.15: Synthesised and real LSB images arranged side by side for comparison. Note, in the real image, the ghosted halos and saturation trails surrounding stars, and horizontal band containing a comparatively high level of background noise.

We implement several options (see Figure 3.16) to change the realness of the synthesised data:

1. The first option fixes orientation of cirrus clouds to be horizontal across the image, as shown in Figure 6.1a, allowing the evaluation of the ability to handle rotation.
2. The second option introduces star-like objects with telescope ghosted halo artefacts (i.e. bright transparent halos around each bright spots simulating stars). These star-like objects are created from a sharp Gaussian profile approximating a point source, where the standard deviation of each star's Gaussian profile is randomly slightly varied to ensure variation. A synthetic halo resembling a telescope artefact

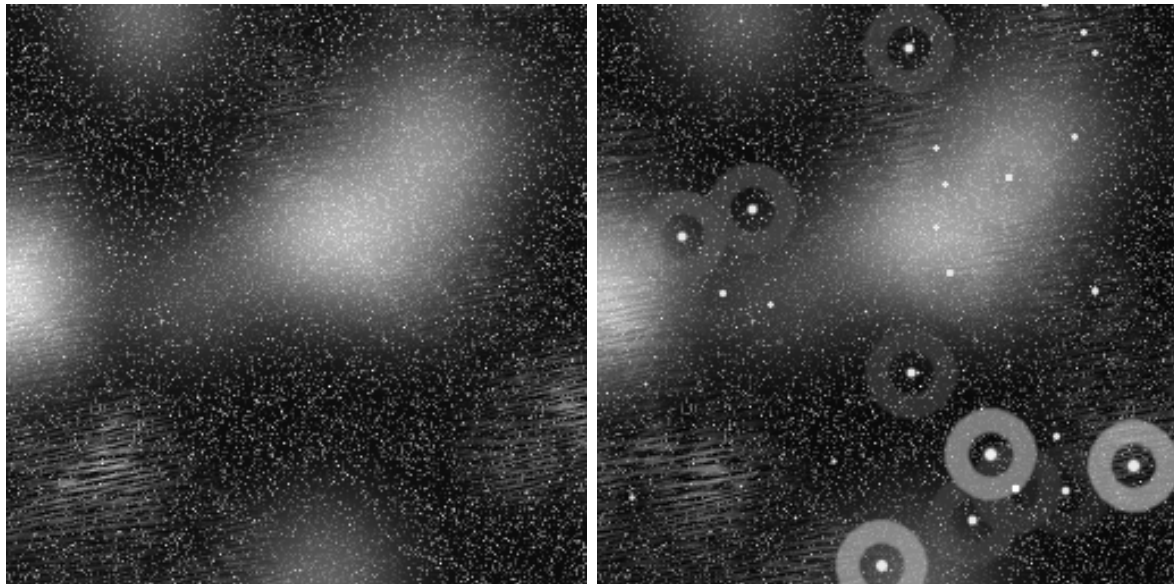
is then added around each star, and is created from a circle of fixed radius and width and with a uniform brightness proportional to the star’s associated Gaussian standard deviation. In the denoising scenario, in addition to removing cirrus, the ML model must also remove synthetic halos from stars. An example of this effect is shown in Figure 6.1b.

3. The third option creates gaps in the image to simulate an instrument artefact where in the process of tiling different captured regions of the sky into one image, there exists no picture of an area between tiles. This option additionally adds a ‘saturation’ effect where pixels containing bright objects such as stars saturate neighbouring pixels. An example of these synthesised effects is shown in Figure 3.16c.
4. In the fourth option further artefacts are added to the image where sections of the image contain higher background levels than the surrounding image, as shown in Figure 3.16d.

The combination of options 2, 3 and 4 results is illustrated alongside a real contaminated LSB image in Figure 3.15.

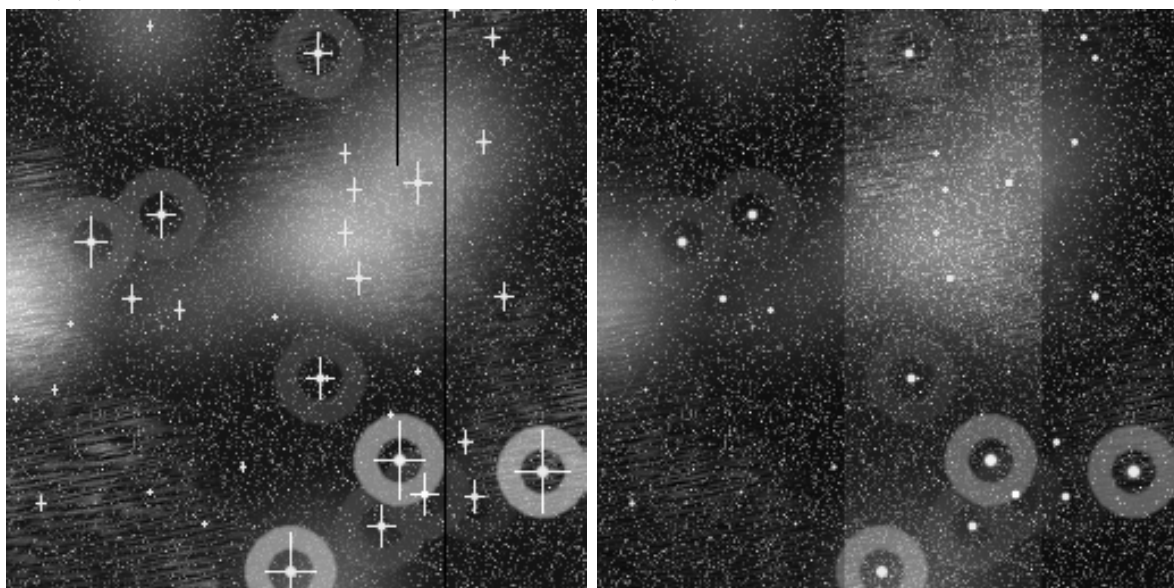
3.5 Conclusion

Supervised ML algorithms such as CNNs are capable of complex image processing, though training such algorithms requires target examples for every input sample. In this chapter, we presented an annotation tool that allows astronomers to precisely delineate the envelope of objects in large multispectral astronomical images. We detailed a dataset of 6573 annotated objects that was generated with this tool on LSB images from the MATLAS survey, and discussed how annotations by multiple users are combined for training ML algorithms. Finally, to address the limited sample size of the detailed annotated dataset, a dataset of synthesised images with features similar to LSB images was presented for pretraining ML models.



(a) Fixed horizontal cirrus orientation.

(b) Artificial stars with ghosted halos.



(c) Artificial gaps and saturated pixels.

(d) Artificial high background sections.

Figure 3.16: Different options to introduce difficulty variations into synthesised LSB images. In this random iteration, the rotation angle of the cirrus was close to 180° , thus the fixed rotation example has a similar orientation but different appearance.

Chapter 4

Learnable Gabor Modulation in Complex-valued Neural Networks

We shift attention from the previous chapter and investigate integrating orientation robustness, a useful ability for processing LSB images. Robustness to transformation is desirable in many computer vision tasks, given that input data often exhibits pose variance. While translation invariance and equivariance is a documented phenomenon of CNNs, sensitivity to other transformations is typically encouraged through data augmentation. We investigate the modulation of complex valued convolutional weights with learned Gabor filters to enable orientation robustness. The resulting network can generate orientation dependent features free of interpolation with a single set of learnable rotation-governing parameters. By choosing to either retain or pool orientation channels, the choice of equivariance versus invariance can be directly controlled. Moreover, we introduce rotational weight-tying through a proposed cyclic Gabor convolution, further enabling generalisation over rotations. We combine these innovations into Learnable Gabor Convolutional Networks (LGCNs), that are parameter-efficient and offer increased model complexity. We demonstrate their rotation invariance and equivariance on MNIST, BSD and a dataset of simulated astronomical images of Galactic cirri.

4.1 Introduction

We enable learning of approximate orientation invariance and equivariance in convolutional neural networks (CNN). Datasets in various domains often exhibit a range of pose variation (e.g. scale, translation, orientation, reflection). CNNs are inherently equipped to handle translation invariance, but remedies for other symmetries often involve large models and datasets with plenty of augmentation. This inability to properly adapt to

transformations such as local/global rotations is a major limitation in CNNs.

An important distinction is that of equivariance versus invariance. For a network to be equivariant, it should be robust to variation in pose and be able to carry over transformations of the input to transformed features and output. For tasks where output is dependent on these transformations, network invariance alone is suboptimal as transformation information is discarded, by definition. For example, in ultra deep astronomical imaging, the scattered light from foreground Galactic cirrus contaminates and occludes interesting Low-Surface Brightness (LSB) extragalactic objects. These cirrus clouds exhibit orientation dependent features: segmenting cloud regions is a problem requiring invariance, as orientation of cloud streaks does not necessarily affect the geometry of the cloud’s envelope. On the other hand, removing occluding clouds, which is crucial to studying background LSB galaxies, is a denoising problem that requires robust and descriptive equivariant features.

Numerous works have been published alongside CNN research attempting to integrate forms of rotation invariant and equivariant feature learning in an a-priori fashion. Approaches typically generate rotation dependent responses by one of the following strategies: 1) learning orientations by constructing filters from a steerable basis [44, 206, 211], 2) rotating convolution filters/input by preset angles [118, 144, 145], or 3) introducing orientation information through analytical filters [140, 220]. A significant drawback of the former type is that it introduces significant computational overhead [38]. In the second category, the rotation process imposes the use of interpolation which results in artefacts for any rotation outside of the discrete sampling grid. This is overcome in the latter category by using analytical filters with an inherent rotation parameter. Orientations are static in [140, 220], similarly to the second category, however there is no inherent limitation of analytical filters preventing them from having learnable orientation parameters. There is thus a need for a dynamic orientation sensitive architecture that can accurately adapt to the input’s transformation. We address this need in this work using Gabor filters, analytical filters that are parameterised by orientation, scale and frequency among other variables. Furthermore, Gabor filters are differentiable with respect to their parameters, meaning that these parameters can be learned through steepest descent style algorithms.

Contributions - In this paper we propose Learnable Gabor Convolutional Networks (LGCN), a complex-valued CNN architecture highly sensitive to rotation transformations. We utilise adjustable Gabor modulation of convolutional weights to generate dynamic orientation activations. By learning Gabor parameters alongside convolutional filters we achieve features that are dependent on exact angles with no interpolation artefacts. Moreover, there is no explicit constraint on convolutional filters, allowing a diverse feature space that adapts to the degree of rotation equivariance required. We extend the modulation

approach used in [140] to complex space, enabling use of the full complex Gabor filter and exploiting the inherent descriptive power of complex neurons. Further, we build on this and propose a convolutional operator where Gabor filter modulation is cyclically shifted, inspired by group theory CNNs [17, 43, 44, 56, 206], allowing propagation of orientation information throughout a forward pass in an equivariant manner.

4.2 Related Work

Numerous methods have been developed in an attempt to integrate transformation invariance in an a-priori fashion. Prior to CNN popularity, the use of hand-crafted features such as SIFT [139] and Gabor filters [7, 86] was explored to generate rotation/scale invariant representations. A widely adopted technique in deep learning is to augment transformations into a dataset [39, 117, 180]. This brute force approach introduces new samples to prompt the model to learn this new range of transformations. Models with learned invariance through augmentation require a very large parameter space to capitalise on data augmentation, and still may generalise poorly to unseen transformations.

There has been much work recently on encoding symmetries into CNN architectures. Early efforts utilised pooling over transformed responses, e.g. siamese networks [212], training-time augmentation [108, 184], parallel convolutional layers [55, 56], kernel-based affine pooling [81], and image warping [92, 106, 107, 128]. Specifically in the last few years there has been a surge of interest in rotation equivariant architectures. Authors have been able to formulate CNNs entirely from principles of group theory and thus construct modified operators and/or constrain filters [43–45]. Bekkers et al. [17] employ bi-linear interpolation to enable any regular sampling of the continuous group of 2D rotation. Similarly [144] and [145] utilise copied and rotated filters, but pool over the produced activation maps. While interpolation allows rotation by exact angles (as in [17, 144, 145]) it introduces artefacts for angles outside of the discrete sampling grid. In [105] residual blocks are combined with principles of steerable bases to learn approximate equivariance. Worrall et al. [211] allow exact orientation representations while overcoming dependence on interpolation by constraining filters to the family of complex circular harmonics: there is a clear demonstration of complex neurons encoding rotational information which justifies our usage of complex CNNs for rotation equivariance. Finzi et al. [72] are able to construct group equivariance without steerable filters by constructing filters as parameterisations of Lie algebra. Similarly, Weiler et al. [206] present a CNN architecture with learnable steerable filters, and derive a generalised weight initialisation method for steerable basis coefficients. Using a formulation of steerable filter architectures, [205] proposes a general framework for equivariant networks under any combination of rotation,

reflection or translation. We draw inspiration from the cyclic shifting group convolutions commonly used in group theory based CNNs [17, 43, 44, 56, 206], and propose a similar operation for rotation generalisation without requiring derivation from group theory and reducing computational overhead.

Analytical filters have made a resurgence in many deep learning contexts. Specific to transformation invariance, analytical filters parameterised by rotation are fast and can extract orientational features dependent on exact angles, overcoming interpolation artefacts. Several approaches replace convolutional weights with wavelet filters [30, 66, 179]. Wavelets are also applied to inputs of standard convolutional layers in a preprocessing fashion [78]. In [203] authors present a framework for convolutional weight modulation, achieving enhanced filters with binarised weights. Zhou et al. [220] exploit rotation parameterisation of discrete Fourier transforms to extract orientation information, modulating standard convolutional filters with a filter bank of rotated analytical filters. Luan et al. [140] implement a similar approach but opt to use Gabor filters, demonstrating that they are more robust to rotation and scale transformations. In [109] wavelet filter hyperparameters are learned in an end to end fashion for spectral decomposition through wavelet deconvolutions. We combine lessons learned from [109] with [140] to construct Gabor filter modulation with learnable parameters.

4.3 Methodology

Our LGCN achieves sensitivity to rotation transformations through adjustable Gabor modulation of convolutional weights. In the architecture defined below, modulation parameters are learned alongside convolutional filters. Having separate modulation and convolution parameters keeps backpropagation simple. Given that convolution filters are not explicitly constrained as in other methods attempting to overcome transformations, the result is a larger space of possible features. Fig. 4.1 illustrates the general structure of LGCNs, providing an overview of the concepts proposed throughout this section. An important development of this approach is that parameters belong to complex space, allowing both real and imaginary parts of analytical filters to be utilised. Given that frequency response filters are often designed over complex space, this enables a variety of modulation options.

LGCNs can consider several orientations simultaneously, which are finely tuned to the task being solved. LGCNs are able to achieve activations dependent on arbitrary continuous rotations with no interpolation artifacts and without using steerable filter bases. With modulation, LGCNs increase model complexity at little cost to the parameter size. In our case Gabor filters are calculated with orientation θ and wavelength λ , meaning they

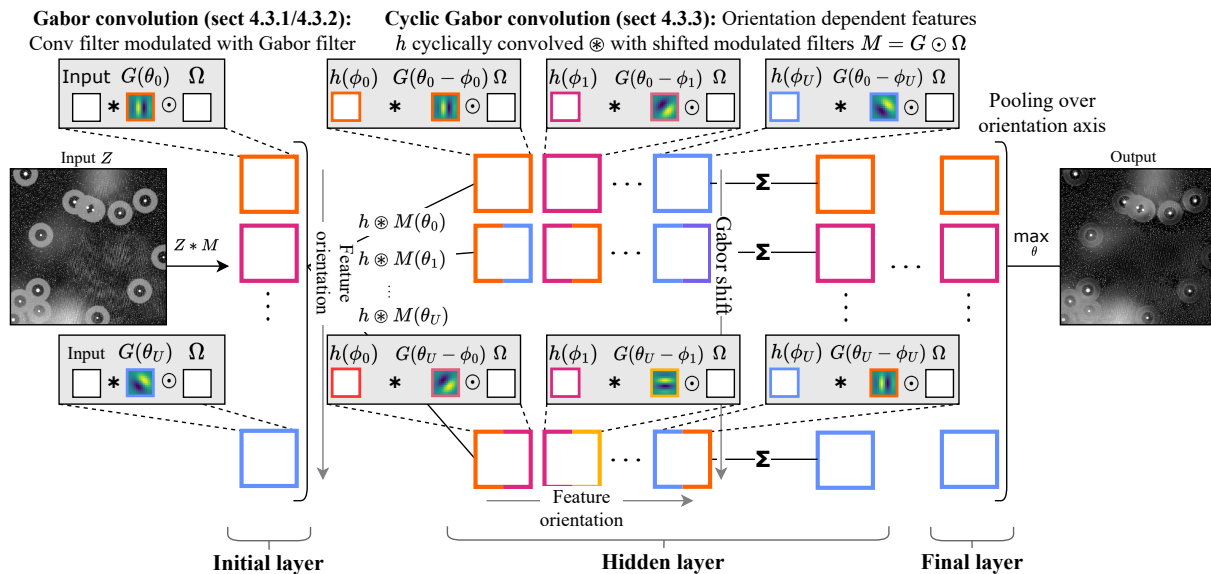


Figure 4.1: Overview of LGCN, with illustration of the filter modulation and cyclic convolution concepts. The displayed network is tasked with removing cirrus streaks and ghosted halos from the input image. Colour denotes individual feature orientations. Gabor modulation is applied with a range of angles to generate orientation dependent features. In the cyclic operator, each input orientational feature is exposed to every rotation of Gabor modulation, encouraging weight tying across orientation.

can generate a new feature channel with only two parameters. Finally, LGCNs utilise a novel convolutional operator where Gabor filters are cyclically shifted during modulation, enabling propagation of orientation information across layers and thus facilitating learned invariance and equivariance.

4.3.1 Analytical Modulation of Complex-Valued Networks

In order to enable compatibility with a wide range of analytic frequency response filters, we construct complex-valued CNN layers. As described in [194] we construct complex numbers by encoding real and imaginary parts as separate real valued elements. A complex convolutional weight tensor can be written as $\Omega = A + iB$, where A and B are stored internally as real tensors. Complex arithmetic is then simulated with appropriate real operations acting on these elements. For a complex valued input $H = X + iY$, convolution is computed as:

$$Z = \Omega * H = (A * X - B * Y) + i(B * X + A * Y) \quad (4.1)$$

where $*$ denotes the convolution operator. For nonlinearities, we use the complex ReLU proposed in [8], $\mathbb{C}\text{-ReLU}(Z) = \text{ReLU}(|Z| + b) \frac{Z}{|Z|}$ with b a real-valued bias term. We also implement complex analogues of batch normalisation, given in [194], and average spatial pooling, trivially given by considering the average of real and imaginary parts separately. Implementing CNNs with complex-space algebra can also be viewed as a form of regularisation, where weights are inherently tied together in pairs. Such a weight tying constraint results in a model directly encouraged to analyse phase information in features, which may be beneficial for encoding orientational information [194], while also likely benefitting from the standard paradigm of reduced overfitting through regularisation.

For an analytical filter $\Phi^P = \Phi_{\text{Re}}^P + i\Phi_{\text{Im}}^P$ with D parameters $p_d \in P = \{p_0, \dots, p_{D-1}\}$ we extend the convolutional modulation presented in [203] to complex space. The modulation of convolutional kernel $\Omega^c = A^c + iB^c$ of channel c with discretised filter Φ^P is given by $M^{c,P} = \Phi^P \odot \Omega^c$, where \odot represents complex element-wise multiplication. Output from convolution with the modulated filter is then given by $Z^{c,P} = M^{c,P} * H$, or for each pixel at coordinates (s, t) :

$$z_{s,t}^{c,P} = \sum_{k=1}^K \sum_{l=1}^K m_{k,l}^{c,P} h_{s+k,t+l} \quad (4.2)$$

This construction of modulated filters can be viewed as a collection of filter banks, where both the underlying kernels (via convolutional filter Ω) and frequency sub-bands (via Φ) are learnt. The complete filter bank has dimensions $2 \times C_{\text{out}} \times C_{\text{in}} \times U \times K \times K$, with C_{out} and C_{in} the number of output and input convolutional channels respectively, U the number of modulating filters, and K the convolution kernel size. Any given filter is obtained by modulating a convolutional filter W^c of channel c with analytical filter Φ^{P^u} . A significant advantage of this formulation is that a filter bank of U filters is created from a single canonical filter, meaning that encoding of transformation representations requires little computational overhead. Adjusting parameters through backpropagation requires calculating the gradient of a differentiable loss function L with respect to p_d :

$$\frac{\partial L}{\partial p_d} = \frac{\partial L}{\partial M^{c,P}} \frac{\partial M^{c,P}}{\partial p_d} = \sum_{k=1}^K \sum_{l=1}^K \frac{\partial L}{\partial m_{k,l}^{c,P}} \frac{\partial m_{k,l}^{c,P}}{\partial p_d} \quad (4.3)$$

$$= \sum_{k=1}^K \sum_{l=1}^K \left(\sum_{s=1}^N \sum_{t=1}^N \frac{\partial L}{\partial z_{s,t}^{c,P}} \frac{\partial z_{s,t}^{c,P}}{\partial m_{k,l}^{c,P}} \right) \frac{\partial m_{k,l}^{c,P}}{\partial p_d} \quad (4.4)$$

$$= \sum_{k=1}^K \sum_{l=1}^K \frac{\partial m_{k,l}^{c,P}}{\partial p_d} \sum_{s=1}^N \sum_{t=1}^N \frac{\partial L}{\partial z_{s,t}^{c,P}} h_{s+k,t+l}. \quad (4.5)$$

Thus the only constraint on choice of analytical filter Φ is that it is differentiable with respect to parameter p_d . In the following subsection we compute the above derivative in the scenario where only a subset of parameters are learned.

4.3.2 Learning Analytical Parameters through Backpropagation

In this paper we modulate with Gabor filters similarly to [140], which are feature detectors characterised by rotation sensitivity and frequency localisation: $G(\lambda, \theta, \psi, \sigma, \gamma)_{k,l} = e^{-\frac{k'^2 + \gamma^2 l'^2}{2\sigma^2}} e^{i(\frac{2\pi}{\lambda} k' + \psi)}$ with $k' = k \cos \theta + l \sin \theta$ and $l' = l \cos \theta - k \sin \theta$. Two major differences with [140] is that we work with complex-valued networks, and we learn the parameters of the filters while they were fixed to static orientations in [140]. A significant advantage of Gabor filters in comparison to Fourier related methods such as DCT is that they are not constructed from a sinusoidal basis, meaning that discontinuous patterns, such as edges, can more easily be represented. We fix (hyper)parameters other than orientation θ and wavelength λ : $G(\lambda, \theta, 0, \frac{1}{\sqrt{2}}, 1)$ as in [140] which demonstrated that this provides sufficient expressivity while simplifying computation. Though we choose to modulate with the well-documented Gabor filters due to orientation and frequency parameterisation, it is possible to modulate with a variety of complex analytical filters with this approach.

Thus the modulated filter $M^{c,P}$ can be written as $M^{c,P} = G^P \odot \Omega^c$. We evaluate $\partial m_{k,l}^{c,P} / \partial p_d$ at pixel k, l in the context of Gabor filter modulation for both parameters. Given that $\theta, \lambda \in \mathbb{R}$ we treat $M^{c,P}$ as a function of the real and imaginary parts separately (k, l indices omitted for readability):

$$\frac{\partial m^{c,P}}{\partial \theta} = a'^c \frac{\partial G_{\text{Re}}}{\partial \theta} + b'^c \frac{\partial G_{\text{Im}}}{\partial \theta} = \frac{2\pi}{\lambda} e^{-(k^2+l^2)} l' [-a'^c \sin(\frac{2\pi}{\lambda} k') + b'^c \cos(\frac{2\pi}{\lambda} k')] \quad (4.6)$$

$$\frac{\partial m^{c,P}}{\partial \lambda} = a'^c \frac{\partial G_{\text{Re}}}{\partial \lambda} + b'^c \frac{\partial G_{\text{Im}}}{\partial \lambda} = \frac{2\pi}{\lambda^2} e^{-(k^2+l^2)} k' [a'^c \sin(\frac{2\pi}{\lambda} k') - b'^c \cos(\frac{2\pi}{\lambda} k')], \quad (4.7)$$

where $a'^c = a^c + b^c$, $b'^c = a^c - b^c$. Backpropagation $\frac{\partial L}{\partial p_d} = \frac{\partial L}{\partial M^{c,P}} \frac{\partial M^{c,P}}{\partial p_d}$ can now be calculated, enabling learning of Gabor filters' parameters alongside convolutional weights. Accordingly, parameters are updated by $\theta' = \theta - \eta \frac{\partial L}{\partial \theta}$ and $\lambda' = \lambda - \eta \frac{\partial L}{\partial \lambda}$, with η denoting learning rate.

4.3.3 Cyclic Gabor Convolutions

In intermediate layers we implement cyclic convolutions to further increase rotation information without increasing parameter size and utilise the additional feature channels generated by Gabor modulation. We exploit the cyclic property of finite subgroups of 2D rotation transformations to create convolutional filters based on all permutations of

orientation and canonical filters. By sharing all weights across every orientation, the underlying canonical filters further generalise over rotations. This is analogous to how filters are exposed to all translations in standard CNNs to encourage generalisation over translations. This propagation of rotation dependence directly facilitates equivariance, in contrast to non-cyclic Gabor convolutions which must pool over the orientation axis per layer.

Note that this cyclic framework does not require analytical filters that are steerable, only that filters can be parameterised by rotation. That is to say, it is not a requirement that filters meet the criteria of linear steerability according to Freeman et al. [76]. In particular, we demonstrate that rotational weight sharing through cyclic shifting can be achieved with Gabor filters, which are not steerable. Networks constructed from steerable filters where basis coefficients are learned in place of convolutional kernels inherently and explicitly limit the filter space – whether this is a downside or an optimal regularisation to achieve rotation equivariance is yet to be shown. In comparison, learnable modulation with Gabor filters implicitly regularises filter space.

The cyclic convolution design we propose takes inspiration from group convolutions presented in [43]. Specifically, cyclic Gabor convolutions utilise the shifting operation used in group convolutions defined over 2D roto-translations [43, 206]. As the modulation transformation cannot be used to form a symmetry group, we do not derive computation using the group framework. However, using the orientation sensitivity of the Gabor filter we implement a similar resulting feature composition, enabling rotational weight sharing without requiring a proof for strict equivariance.

With $h^c(\theta)$ denoting channel c and orientation θ of the previous layer’s activation map, for a single orientation and output channel, cyclic convolution \circledast is computed as:

$$z^{\hat{c}}(\theta) = \sum_{c=1}^{C_{\text{in}}} \left[h^c \circledast M^{\hat{c}c} \right] (\theta) \quad (4.8)$$

$$= \sum_{c=1}^{C_{\text{in}}} \sum_{\phi \in P} \left[h^c(\phi) * M^{\hat{c}c}(\theta - \phi) \right] \quad (4.9)$$

$$= \sum_{c=1}^{C_{\text{in}}} \sum_{\phi \in P} \left[h^c(\phi) * (G(\theta - \phi) \odot \Omega^{\hat{c}c}) \right]. \quad (4.10)$$

P is the set of U orientations that are used to generate Gabor filters. Note this formulation allows any size of P . In order to keep implementation efficient and avoid recalculating Gabor filters for all permutations of learned orientations we keep $\phi \in P$ as the original angles. This choice allows filters to be reused with a cyclic shift of the orientation components per different output orientation θ .

4.3.4 Learnable Gabor Convolutional Networks

The framework presented above allows learnable modulation to be added into any convolutional layer, making the method very versatile. There are some considerations to take into account however, which we discuss in this section.

4.3.4.1 Complex weight initialisation

LGCNs operate over complex space, requiring weight initialisation to be rethought. Principles of He weight initialisation [88] no longer hold given that $\text{Var}(\Omega) \neq \text{Var}(A) + i\text{Var}(B)$, i.e. real and imaginary parts cannot be initialised independently. We use Trabelsi’s generalisation of He’s strategy over complex space [194], setting $\text{Var}[|\Omega|] = \frac{4-\pi}{2n_{\text{in}}}$ with n_{in} denoting the number of input units. The phase is then uniformly distributed around the circle. It is worth noting that He’s derivation is specific to the traditional ReLU, using the result that for a given input X_l to a layer l , and previous output Y_{l-1} : $\text{E}[X_l^2] = \frac{1}{2}\text{Var}[Y_{l-1}]$. This holds for traditional ReLU, $X_l = \max(0, Y_{l-1})$, as Y_{l-1} has zero mean and a symmetric distribution which is essentially split along its axis of symmetry. However with \mathbb{C} -ReLU, for $b < 0$, Y_{l-1} is no longer divided along the axis of symmetry. For this reason we simply initialise the biases of \mathbb{C} -ReLU layers to zero.

The choice of initialisation for modulation parameters is largely dependent on the choice of analytical filter, and should be influenced by the function’s domain and the roles of individual variables. For initialisation of Gabor parameters, as discussed in Section 4.3.2, we fix phase shift ψ , aspect ratio γ and scale σ in order to simplify computation. Given that wavelength is a non-negative quantity we initialise λ with mean $3\sqrt{U}$ and variance $\frac{\sqrt{U}}{4}$ as per [88], and verified that training is stable. This choice of initialisation also avoids spatial aliasing of the Gabor filter for all kernel sizes (i.e. 3×3 or larger) at network initialisation. As the filter is sampled more than twice per phase, the signal is adequately captured, as per the Nyquist-Shannon sampling theorem. For orientation θ , in the real case there is no benefit of using the full interval of rotations due to evenness, however in the complex case the oddness of the imaginary part causes orthogonal filters for θ with differing sign. For this reason we initialise θ uniformly around the full circle.

4.3.4.2 Gabor axis considerations

Though the ability to create enhanced filters from a single canonical filter has advantages of parameter efficiency and weight-tying, it leaves the network prone to dimensionality explosion. This can be controlled using one or more of three approaches depending on the problem at hand: adjusting the number of convolutional channels C depending on the dataset’s feature complexity; adjusting the number of modulating filters U based on

the dataset’s pose variation; and max pooling along the orientation axis i.e. over the modulating filters for each pixel of the bank of modulated feature maps. The latter operation has the additional advantage of focusing the attention of the network on (local) dominant orientations, which is a particularly useful feature for orientation invariance.

4.3.4.3 Invariance vs equivariance

There is a clear relation between pooling technique and invariance versus equivariance. Preserving only the strongest orientation response discards low response representations and disentangles features, this is however at the cost of encouraging invariance to local rotations rather than equivariance. In practice, invariance is achieved through pooling after each hidden layer over the feature orientation dimension or the Gabor shift dimension, for convolutions and cyclic convolutions, respectively – see Fig. 4.1.

4.3.4.4 Projection between \mathbb{C} and \mathbb{R}

Finally, since data used in this paper is real, we set the imaginary part of inputs to zero. Following the first layer, due to the nature of complex algebra, imaginary parts are no longer zero-valued. This has the effect of the first layers of the network learning a beneficial imaginary projection of the input, alongside feature learning. Some works [177, 194] opt to include a preprocessing step to estimate the imaginary part though we found this had a detrimental effect on performance. For real classification, final complex feature maps must be projected back onto real space. We experimented with several projection methods such as complex linear layers and using magnitudes, but found empirically that simply concatenating real and imaginary values into fully connected linear layers performed best. We hypothesise that concatenation removes the regularisation caused by complex algebra, and that this is beneficial in the classification layer due to the comparatively small number of neurons versus previous convolutional layers.

4.4 Experiments

In this section we validate our learnable modulation formulation, showing that learning analytical filter parameters leads to improved accuracy on both artificial and real data. Initially, LGCNs are evaluated on variants of MNIST [122] containing rotated samples, where we evaluate the network’s learned invariance. In the next section we compare invariance and equivariance in both a standard CNN and a learnable Gabor modulated CNN, where networks process synthesised and real samples of galactic cirri. All experiments throughout this section were run using a single NVIDIA GTX 1080 Ti.

4.4.1 Orientation invariance on MNIST

MNIST [122] (CC BY-SA 3.0 license) is a standard benchmark for transformation invariance because of its simplicity, interpretability and vast array of variants. The dataset contains 60000 images of handwritten digits, where the typical task is to classify the drawn digit from 0-9. In the context of transformation invariance, the digits undergo some transformation, e.g. in our case a rotation. The network’s classification performance of digits after image transformations have been applied indicates the robustness of the network to the applied transformations. We apply a random rotation between $[0, 2\pi)$ to yield a rotated MNIST, and train with 5-fold validation. Our baseline classification architecture is similar to that used in [43, 206, 211], with three blocks of increasing channels, representing a hierarchy of feature complexity. Each block contains two learnable Gabor modulated convolutional layers with a kernel size of 3×3 followed by max pooling along the orientation axis and average spatial pooling. We use no cyclic Gabor convolutional layers, but these may be included in future experiments. In the final block, features are pooled globally so that a given activation contains one complex value per feature channel. We then concatenate real and imaginary parts into a single vector and use three (real valued) fully connected layers for classification. The Adam optimiser [111] is used for network training, starting at a learning rate of 0.001 and then decaying with an exponential schedule by 0.9 every epoch. L2 weight regularisation is also enforced with a penalty of 10^{-7} .

Exploration of rotation invariance in the feature maps – The number of modulation filters U has a direct effect on the network’s ability to capture rotation dependent features. We vary this parameter and investigate its effect on network’s performance and the learned features of the first layer, which are the most directly affected by low-level geometrical transformations of the input image. For this first experiment, we train networks with $U \in \{1, 2, 4, 8, 16\}$. We measure and compare response magnitudes (measured as the ratio between the average magnitudes of input and output activations) between original and rotated samples, for all rotations in the (discrete) range $[0, 360^\circ]$, for each network (Fig. 4.2 right). Though response magnitude varies slightly, this may be largely due to interpolation artefacts caused by rotation of the input samples. Nonetheless, the pattern remains predictable throughout the rotation interval with decreasing amplitude for increasing U , indicating that the number of modulating filters has a direct impact on rotation invariance. We also measure classification accuracy as a function of rotation for 1000 samples from the MNIST test set for each network (Fig. 4.2 left). The small difference in accuracy between $U = 1$ and others indicates that even a little orientation information is helpful in generating intra-class rotation-invariant features that remain

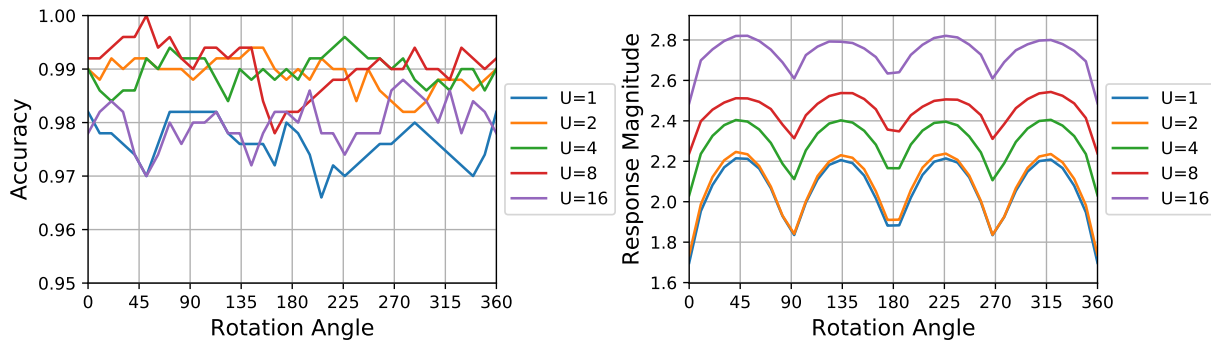


Figure 4.2: Effect of input rotation on MNIST classification accuracy (left) and magnitude of activations in the first modulated layer of the network (right), for different numbers of modulating filters and orientations, and on a subset of 1000 testing samples of MNIST.

LGCN (proposed)	Complex static	Real learnable	GCN4 [140]	ORN8 [220]	CNN
0.9950	0.9915	0.9911	0.9890	0.9888	0.9718

Table 4.1: Classification accuracy on randomly rotated MNIST images.

inter-class separable. At $U = 16$ there is a detrimental saturation of orientations possibly due to the model becoming too complex for the dataset size and task. Optimal performances are reached for U between 2 and 8, with LGCN being not very sensitive to the exact value of this hyperparameter.

Evaluation of the individual modifications to [140] – We evaluate the performance improvements from our two modifications to [140] individually, namely the use of complex-valued filters and of learnable Gabor orientation parameters. In these experiments we apply these modifications both in turn and jointly to the model of [140]. The channel sizes for each LGCN variant were adjusted so that the total parameter size is at most equal to all of the compared models: for complex models this required halving the number of feature channels. The final results, shown in Tab. 4.1, show that both modifications improve classification accuracy, demonstrating the additional feature expressivity afforded in comparison to standard CNNs. All variants also outperform GCNs which use real and static Gabor modulation, with an absolute error difference of 0.6 for LGCNs, showing the benefit of our method’s changes over the previous work. The combination of modifications leads to a large performance increase that may indicate a synergy between the two approaches. One possible explanation for this is that the complex Gabor filter provides smoother gradients with respect to θ and λ , as opposed to only the real part. We will test this hypothesis in future work.

Effect of adjusting learning strategies of Gabor parameters – We investigate the effect of learning Gabor parameters other than θ , and study how changing initialisation methods impacts the model’s performance. The default training configuration for LGCNs uses fixed σ set at π , and learnable λ initialised using a normal distribution with mean $3\sqrt{U}$ and variance $\frac{\sqrt{U}}{4}$. We experiment with fixing or learning wavelength λ and scale σ and study how different configurations affect LGCNs. In addition we compare additional weight initialisation strategies for these variables. For wavelength initialisation we apply: fixed $\lambda = 3$; normal distribution with unit mean and unit variance (not adjusting for U); and uniform distribution between $[-1.5, 1.5]$. For scale, in addition to fixed $\sigma = \pi$, we initialise with a normal distribution with mean equal to π and quarter variance, and enable backpropagation. Finally, we repeat these experiments with only one λ, σ for all modulating Gabor filters per layer.

We train parameter restricted models with varying Gabor parameter learning strategies on rotated MNIST for 30 epochs with 5-fold validation, and record the average performance over all splits. Results are shown in Table 4.2. While initialising with a normal distribution $\lambda = \mathcal{N}(3\sqrt{U}, \frac{\sqrt{U}}{4})$ and fixing $\sigma = \pi$ achieves the highest average performance, there is no clear strategy for either variable that remains best with the other variable strategy changed. Notably, there is a performance decrease when aliasing of the modulating Gabor filters is forcibly introduced by initialising wavelength λ from a uniform distribution with bounds $[-1.5, 1.5]$. In further tests it was noticed that in this scenario λ values do not recover from this range of aliasing even after training for >100 epochs. For this experiment we conclude that given parameters are not in an aliasing range, LGCNs are not particularly sensitive to learning strategy of wavelength λ and scale σ .

		$\lambda = \mathcal{U}(-1.5, 1.5)$	Fixed $\lambda = 3$	$\mathcal{N}(3, \frac{1}{4})$	$\mathcal{N}(3\sqrt{U}, \frac{\sqrt{U}}{4})$
Separate λ, σ	Fixed $\sigma = \pi$	0.9672	0.9702	0.9686	0.9692
	$\sigma = \mathcal{N}(\pi, \frac{1}{4})$	0.9690	0.9704	0.9693	0.9678
Single λ, σ	Fixed $\sigma = \pi$	0.9684	0.9707	0.9698	0.9713
	$\sigma = \mathcal{N}(\pi, \frac{1}{4})$	0.9673	0.9699	0.9707	0.9685

Table 4.2: Classification accuracy on rotated MNIST averaged over 5 splits for different learning strategies of Gabor parameters wavelength λ and scale σ . Rows are divided in the centre to denote whether a single λ and σ is used for all U modulating Gabor filters, or λ and σ are separate for each modulating Gabor filter.

4.4.2 Invariance and equivariance to the dominant orientation of galactic cirri

We validate the benefit of modulation by applying LGCNs to a domain demanding robust orientation-sensitive features. We demonstrate that modulation not only enables the network to learn invariance and equivariance, but aids the network’s ability to generate features unaffected by local disturbances. For these experiments we analyse samples of galactic cirrus clouds – astronomical objects with striped quasi-textures exhibiting clear dominant orientations, as shown in Fig. 4.3 – as they allow the design of experiments that assess both orientation invariance and equivariance separately and the comparative robustness between different models. These images are very challenging, exhibiting overlapping semi-transparent objects, including foreground cirrus with oriented patterns, background objects (e.g. galaxies) with vastly different textures and intensities, and telescope artefacts.

4.4.2.1 Comparing LGCN against a traditional CNN on synthesised cirrus images

In this experiment we evaluate performance on various datasets composed from the synthesised images of cirrus structures described in Section 3.4. We design the dataset to have three variations of increasing realness. The first variation possesses only cirrus clouds with constant orientation and bright regions (see Fig 6.1a); the second randomises cirrus orientation (see Fig 3.14); finally the third introduces star-like objects with telescope halo artefacts (i.e. bright transparent halos around each bright spots simulating stars, see 6.1b). These star-like objects are created from a sharp Gaussian profile approximating a point source, where the standard deviation of each star’s Gaussian profile is randomly slightly varied to ensure variation. A synthetic halo resembling a telescope artefact is then added around each star, and is created from a circle of fixed radius and width and with a uniform brightness proportional to the star’s associated Gaussian standard deviation. Each synthesised dataset contains 300 samples: 160 for training, 40 for validation (for 5-fold validation) and 100 for testing.

We create a U-Net [172] style architecture in both standard form and with Gabor modulated convolutional layers, where skip connections are combined via summation (as in [165]) rather than concatenation. To enable comparison, we create four variants of this network: one with plain convolutions; one with complex-valued convolutions, denoted C-CNN; one with static real Gabor filter modulation as in [140]; and one with learnable complex Gabor modulation with cyclic convolutions. These networks are tasked with first segmenting the cirrus clouds, and secondly removing clouds and artefacts (if applicable).

The complex filters of \mathbb{C} -CNN and LGCN naturally require twice the convolutional filter parameters. We ensure a fair comparison by adjusting channel sizes accordingly, thus keeping total parameter size of the two networks roughly equal. For the denoising task we do not utilise orientation pooling so that orientation information is preserved and equivariance is encouraged, as per discussion in 4.3.4, and experimental verification. Results are presented in Tab. 4.3 with IoU metric for segmentation and peak signal to noise ratio (PSNR) for denoising.

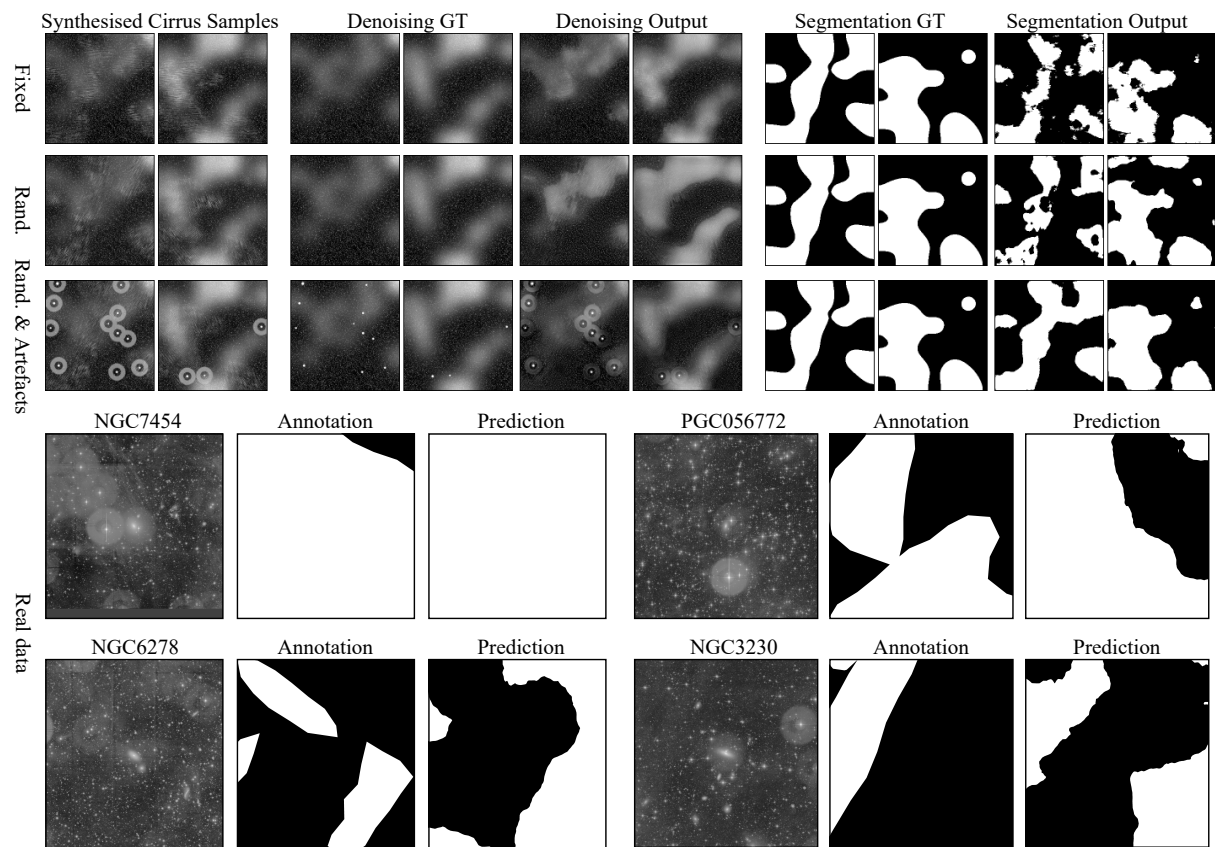


Figure 4.3: Denoising and segmentation results on real and synthesised samples of galactic cirri generated with fixed rotation; randomised rotation; and randomised rotation with stars and telescope artefacts. These are difficult tasks as the striped textures of cirrus regions are easily confused/obstructed with bright diffuse regions and other objects.

In the segmentation case, given that rotation of cirrus texture does not affect the cloud’s envelope, this is a problem where invariance is beneficial. The denoising problem requires equivariance, as isolating the cloud from the detailed background is dependent on the orientation of its streaks. In the first dataset, CNN, \mathbb{C} -CNN and GCN performances are close to LGCN’s, generating fine segmentations of the cirrus clouds with few missed regions, and similarly good denoising. For the second, with orientation variation, GCN

	Fixed	Rand.	Rand & artefacts	Fixed	Rand.	Rand & artefacts	Real cirrus
Base model	0.914	0.882	0.806	26.3	25.3	23.5	0.673
\mathbb{C} base model	0.918	0.905	0.839	26.7	25.6	24.5	0.679
GC model*	0.923	0.920	0.875	26.7	26.2	24.8	0.674
LGCN	0.925	0.918	0.898	27.4	27.1	25.4	0.715

Table 4.3: Segmentation IoU (left), denoising PSNR (middle) on synthesised cirri with fixed and randomised orientation, and with stars and telescope artefacts. Segmentation IoU (right) on real cirrus samples in LSB images. *Gabor convolutions of [140] applied to our base model.

performs marginally best in the segmentation case but its performance drops for denoising. On the other hand, LGCN maintains a quite stable performance for denoising (and also for segmentation) on this dataset. This difference in behaviour may be explained by the use of cyclic convolutions in LGCN that better preserve orientation information due to rotational weight-tying across layers. CNN and \mathbb{C} -CNN performances start to fall behind, with relative differences of 4.1% and 1.4% IoU, respectively, for segmentation and 7.1% and 5.9% PSNR for denoising. This separation becomes much larger in the final experiment on the most complex data exhibiting overlapping textured regions and localised objects, with LGCNs outperforming GCNs in both tasks by 2.6% and 2.4%, \mathbb{C} -CNN by 7.0% and 3.7%, and plain CNNs by 11.4% and 8.1%. The affect of randomising rotations and even introducing telescope artefacts makes little difference to LGCN’s performance for segmentation, demonstrating its strength in generating rotation invariant features that are robust to local disturbances. We see that denoising performance is stable with randomised rotation, indicating equivariant encoding produced by the modulated layers. While performance drops for the third dataset, due to artefacts introducing strong variations locally, LGCN still outperforms other models by a larger margin than without artefacts, showing that feature robustness is exhibited in the equivariant case. The results demonstrate that both the use of complex numbers and the modulation of filters are beneficial. We note that the margin between CNN and \mathbb{C} -CNN is significantly the largest on this dataset variation requiring robustness, compared to other dataset variations. LGCN combines both augmentations and cyclic convolutions for further improved results. Visual analysis of the network outputs (see Fig. 4.3) indicates a possible overfitting for the first two datasets with no telescope artefacts for both the segmentation and denoising tasks, which results in a more difficult generalisation and poorer (visual) quality on test data. This issue may be due to these two simpler scenarios requiring simpler models and/or fewer training steps, and it will be investigated in future work.

4.4.2.2 Prediction of cirrus structures in LSB images

We task LGCNs with segmenting cirrus clouds in optical telescope images, demonstrating the effectiveness of our method on a real world problem. The dataset used in this experiment is composed from a subset of the dataset described in Section 3.3.1, using only images that contain annotations of cirrus. This dataset contains 48 expert annotated images of approximate size 5000x5000, with two channels representing the g and r bands. Of the 48 images, we use 32 for training, 8 for validation (for 5-fold validation) and 8 for testing. Across the entire dataset 55% of pixels are labelled as cirrus contaminated. Networks are trained over 300 epochs on random crops of size 512x512, which are then downscaled by a factor of two. To mitigate against the limited sample size, we augment data with random flips and 90° rotations, and pretrain networks on an extended version ($N=1200$) of the synthesised dataset.

We also train a standard CNN, a \mathbb{C} -CNN, and a GCN [140] for comparison as in the synthesised data experiment, fixing parameter size to be roughly equal. In comparison to the synthesised images, real cirrus regions often exhibit much fainter textures and orientation is more subtle and can vary slightly globally. In addition, training labels may not be fully reliable, due to the difficulty in annotating precisely the borders of cirri – there is an inherent uncertainty associated with each annotation, especially due to the ambiguous nature of the cirrus cloud boundary, so several experts may disagree on the exact location of borders –, and due to the limited number of available expert annotators – in our dataset, 2 to 3 annotators annotated each image, but for simplicity in this proof-of-concept, we worked only with the annotations of the single most experienced expert, making the simplified assumption that their annotation corresponds to the ground-truth. These factors, in combination with more severe artefacts, background noise, and small training size, make the dataset incredibly challenging. Results are shown in the last column of Table 4.3: LGCNs achieve an IoU of 71.5%, with an absolute increase of 4.2% over standard CNN, 3.6% over complex CNN and 4.1% over real-valued static Gabor modulation without cyclic convolutions. Notably, GCN barely surpasses the base model and is outperformed by \mathbb{C} -CNN, suggesting that only static rotation sensitivity is not sufficient on more challenging datasets, a finding which is supported by results from the previous experiment of synthesised images. LGCNs significantly outperforms compared methods, demonstrating the ability of proposed augmentations to generate robust orientation sensitive features, even on data with extreme contamination. Given that the class balance is 55%, the problem is very difficult, and although this absolute increase represents a significant performance improvement, more progress may be achieved by considering e.g. new architectures to be augmented by our methods, or a multi-scale

approach, and a more complete dataset with consensus annotation from several experts.

4.4.3 Boundary Detection on Natural Images

We demonstrate the general applicability of learnable Gabor modulated convolutions on the Berkeley Segmentation Dataset (BSD500) [6, 146]. This task requires the ability to learn equivariant features in the scenario where there is no dominant global orientation, and the network must handle high variations in local feature orientation dependence. The dataset contains natural images of size 321×481 in both portrait and landscape, with 200 training samples, 100 validation samples and 200 testing samples. Each image has associated with it several ground truth labellings produced by different annotators.

We replicate the pipeline of one of the highest performing methods, RCF [137], and replace convolutional operators with learnable Gabor modulated convolutions. This methodology uses a pretrained VGG16 [181] based architecture, taking ‘side’ outputs from each convolutional block that represent coarser scale edges as network depth increases. These side outputs are then fused together with a 1×1 convolutional layer. The final prediction is then computed as the average between all side outputs and the fused output. We denote our modified implementation as LGCN-RCF: an additional Gabor convolutional layer is used to create an orientation channel; all convolutional layers apart from the fusion layer are replaced with cyclic Gabor convolutions; orientation features are pooled prior to side output. We train for 250 epochs with 3-fold validation. Results are shown in Table 4.4, using the optimal dataset scale (ODS) and optimal image scale (OIS) metrics defined in [6]. LGCN-RCF achieves 0.727 ODS and 0.747 OIS, which is a strong result considering in each epoch we train on one random augmentation per image, as opposed to other methods which use the entire range of augmentations per image (due to lack of compute and time, thus care is to be taken when comparing results). We note that LGCN-RCF only marginally improves on the results of H-Net [211] despite containing roughly 15x more learnable parameters, speaking to the parameter efficiency of steerable filter methods. As previously mentioned, such methods have a larger cost in terms of training time and runtime memory usage, limiting their ability to scale to larger parameter sizes and thus higher performance. Our method also significantly outperforms our implementation of RCF [137] with parameters restricted to match LGCN-RCF, demonstrating the benefit of modulating with complex Gabor filters on tasks with natural images.

Table 4.4: Boundary detection results on the BSD500 [6] dataset. *Our parameter restricted implementation. †ImageNet pretrained.

	Kivinen et al. [114]	DexiNed [162]	RCF* [137]	H-Net [211]	LGCN-RCF	RCF† [137]
ODS	0.702	0.728	0.707	0.726	0.727	0.806
OIS	0.715	0.745	0.720	0.742	0.747	0.823
# params	-	4.41M	1.80M	0.12M	1.88M	14.84M

4.5 Conclusion

We presented a framework for incorporating adaptive modulation into complex-valued CNNs. This framework was used to design an orientation robust network with convolutional layers using Gabor modulated weights, where complex convolutional filters and Gabor parameters are learned simultaneously. A cyclic convolutional layer was proposed to retain rotational information throughout layers and encourage equivariance. Our architecture is able to generate unconstrained representations dependent on exact orientations, without interpolation artefacts. We validated this empirically for three use cases, with experiments designed to test properties of both invariance and equivariance to orientation. We first verified that LGCNs are able to effectively produce rotation invariant features on the rotated MNIST dataset. An ablation study was performed to assess in turn and in combination the effect of two proposed augmentations to GCNs [140], namely using complex-valued weights and learning parameters of modulating Gabor filters. Secondly, we carried out experiments on a purpose designed dataset of varying difficulty. The architecture’s modulated layers were able to create fine segmentations in synthetic and real images despite local disturbances. The presented LGCN architecture achieved strong denoising scores in comparison to standard CNNs, even on contaminating cirrus cloud structures with randomised orientation. Clear performance improvements were observed for both use cases, demonstrating the effectiveness of the augmentations. Thirdly, we applied an LGCN architecture to boundary detection in natural images and achieved strong metrics in comparison to other non-pretrained methods. The successful augmentation of three different architectures also demonstrates the general applicability of our method, and it may be applied to more complex DNNs and application scenarios in the future.

Chapter 5

Multiscale Gridded Gabor Attention for Segmenting Global Contaminants

In this chapter, we address the challenge of segmenting global contaminants in large images. The precise delineation of such structures requires ample global context alongside understanding of textural patterns. CNNs specialise in the latter, though their ability to generate global features is limited. Attention has been used to measure long range dependencies in images, capturing global context, however this incurs a large computational cost. We propose a gridded attention mechanism to address this limitation, greatly increasing efficiency by processing multi-scale features into tiles with smaller resolution. We also extend ideas of Chapter 4 and present a novel way to utilise Gabor filter modulation to encourage orientation sensitivity over larger scales. We measure correlations across features dependent on different orientations of underlying modulating Gabor filters, in addition to channel and positional attention. We present segmentation results on both synthetic and real images containing cirrus samples.

5.1 Introduction

Global context is vital in vision: scenes are described with focus on key descriptive regions, such as grass or sky, as well as through objects. This is especially relevant when processing contaminants covering large regions in images, such as clouds [85] in remote sensing images and in solar imaging, [70], or cirrus clouds in LSB imaging. Multi-scale (MS) CNNs were proposed aiming to increase global context in CNNs, e.g. [89, 218], though contextual understanding in convolutions remains limited to the few final convolutional layers where the kernel begins to span longer ranges after multiple successive pooling operations. In this scenario, drastic downscaling is required to achieve more global receptive fields.

Attention has been proposed to model long range dependencies in data [13, 197]. In image data specifically, attention measures correlations between feature positions and channels [77]. This is achieved by computing, for each pixel, a weighted sum of all other pixels according to feature similarity. While attention has been effective in capturing global context, this pairwise operation has a huge computational footprint. Positional attention has squared complexity in relation to image size, which is barely manageable on popular image datasets with modern GPUs. The compute cost of attention is typically reduced by downscaling feature maps. However, sacrificing texture for gained context may not be a worthwhile compromise for some vision tasks, as severe downscaling significantly erodes local textures often to the detriment of model performance.

Orientalional information is also a valuable discriminator in identifying classes of objects. Textures are intrinsically composed of orientation dependent patterns, and thus understanding of orientations has been shown to increase performance on a variety of segmentation tasks e.g. [105, 118, 140, 206]. Such methods generate feature sets dependent on different orientations in order to decompose orientational patterns within textures, which can then be analysed separately and in combination by follow network layers. Semantic classes may also exhibit inter-dependence between orientations over global ranges, such as contaminants where texture varies over longer distances. While the cited works combined with MS architectures can capture this longer range dependence, there is again a compromise lying in the convolutional operator’s weakness in capturing global dependencies.

In this chapter, we investigate the task of segmenting global contaminants in large images, such as cirrus dust in LSB images. In the case of cirrus, pollution can be difficult to spot, ranging from a slight change in background intensity to total occlusion of galaxies, as shown in Figure 5.1, making cirrus localisation challenging. Even in clean regions of images, background intensity levels vary across the image, thus it is necessary to consider the entire image ($>5000 \text{ px}^2$) to maximise global contextual information. Local textural patterns are also necessary discriminating properties of cirrus, which presents as a wispy texture often with filamentary structures sharing a common orientation. This requirement to analyse both local and global information within the image is common for large contaminants, especially when the contaminant at a glance may be confused with a different semantic class, such as clouds cover versus snowy terrain in remote sensing images.

We propose a gridded MS architecture (Section 5.2.1) that addresses the computational cost of attention. Furthermore, gridded attention introduces a multi-scale aspect with no added (and even reduced) cost, while MS tends to be computationally expensive. We divide features of different scales into tiles of smaller constant size. Positional and

channel attention is computed on these tiles to assess both local and global context in an efficient manner, before reassembling tiles into a final attention map. A closely related work is [182] where attention is also applied to each scale of MS features, but considering whole feature maps in each attention module which results in very high computing costs. Additionally, we present a novel attention operator using orientation (Section 5.2.2), for improved sensitivity to textures including the filamentary patterns of cirri. We utilise Gabor modulated convolutions detailed in Chapter 4 to generate features dependent on different angles. Attention is then computed across these angles, measuring correlations between orientation dependent features.

5.2 Extending Attention for Segmenting Large Contaminants

In this section, we present enhancements of attention modules for effective and computationally efficient segmentation of large contaminants. These enhancements may be applied to various attention modules, and we demonstrate them on [77]. Likewise, different backbone CNNs may be used, and we demonstrate on two different backbones in our experiments. We first detail our approach for generating a hierarchy of features from different spatial scales using dual attention [77]. Second, a novel attention operator for studying correlations between orientational information is proposed. Finally, we provide a detailed overview of our deep learning architecture.

5.2.1 Multi-scale Gridded Attention

Accurate identification of contaminants requires comparison to surrounding regions. Cirrus structures, for example, can be very difficult to identify and differentiate from other contaminants such as image artefacts. Various cirrus regions and their associated annotations are shown in Figure 5.1. While cirrus can present as uniform bright regions, bright areas near a source-like object is can also be diffuse light from a galaxy, rather than cirrus pollution. To reliably spot areas of cirrus pollution it is important that the model is able to study the image at multiple levels of field of view. This is much like how a human annotating the cirrus region would zoom in to study textures and zoom out to see surrounding regions. This idea is referred to as context in computer vision, where classification of a region is assumed to have a dependency on neighbouring regions. Attention [13, 197] has been proposed as a method to analyse contextual relations in images. Attention maps can be computed with respect to position, in order to measure correlations between different

regions, and with respect to feature channels, in order to measure correlations between different learned features [77, 209].

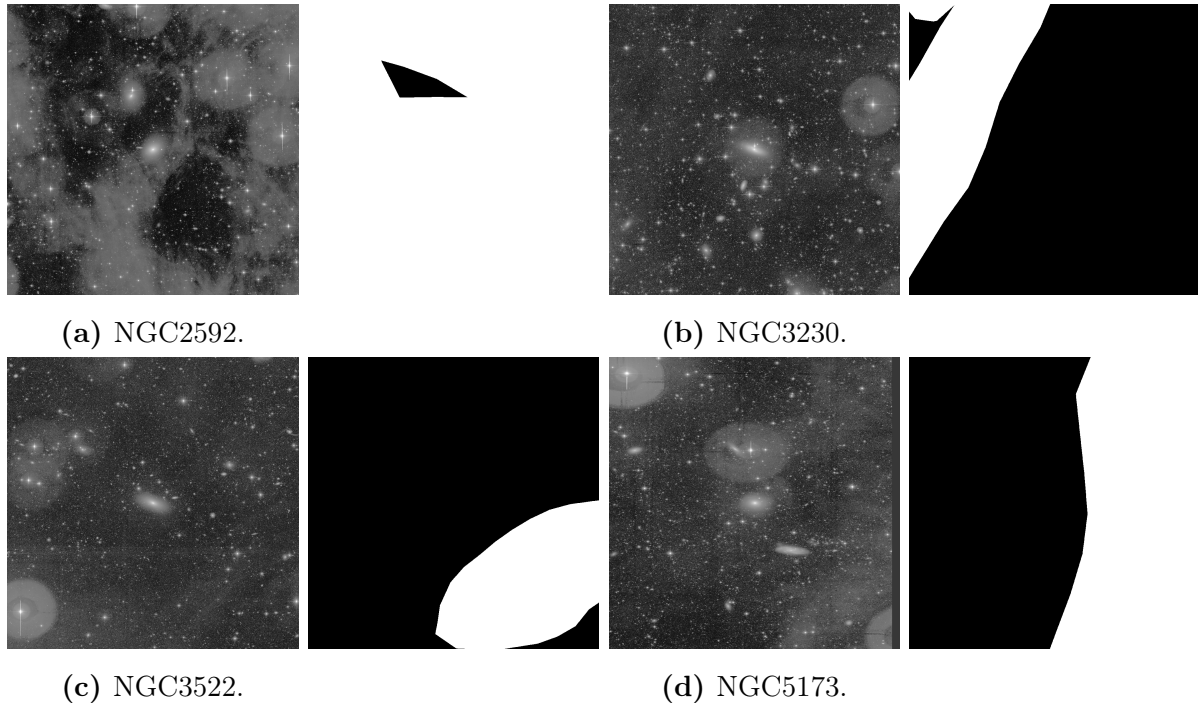


Figure 5.1: Examples of cirrus contaminating different galaxies in various strengths, and their associated annotation. Images are taken from the r-band.

We utilise multiple attention operators at different spatial scales in order to enhance local and global contextual understanding. Methods of generating multi-scale features have been developed in various fashions [36, 37, 130, 210, 215], including methods utilising attention [219]. A proven approach for generating features at multiple spatial scales is to divide features maps into different sizes and apply convolutional layers in parallel [89, 126, 188, 218]. Yu et al. [216] use attention operators in succession to refine feature representation and propose an auxiliary loss to guide attention maps towards better class separation, demonstrating increased performance on datasets of limited size. Sinha and Dolz [182] apply attention to features from different intermediate layers of ResNet [90] in parallel to assess contextual information on multiple spatial scales, and extend the guided supervision of [216] to dual attention modules. Tao et al. [191] create copies of input images and generate segmentation predictions and attention maps from each copy, before combining low-scale predictions with higher scale attention maps in a hierarchical fashion.

5.2.1.1 Background on attention mechanism

We utilise attention to encode contextual dependencies. In this approach, correlations between features are measured along a given dimension, thus modelling inter-feature dependency. For example, two features in separate dimensions that only present either in a pair or not at all would score a high correlation, as a dependency is implied. Following this, correlations are scaled based on the strength of given features. This has the effect of prioritising correlations that have a larger effect on model classification, and suppressing those that do not. This process can be performed along any dimension or axis which represents some internal organisation of a feature map, allowing contextual dependencies to be measured in a variety of fashions.

In this work, we use the attention operator described in [77]. This attention operator, denoted h , can be implemented generally over any tensor dimension as follows. Let X represent a single input feature map tensor with axes $A_1 \times A_2 \times \dots \times A_d$, where d denotes the number of dimensions used in the internal organisation of the tensor. For example, in traditional CNNs a feature tensor may have axes channels $C \times$ height $H \times$ width W . For attention over a desired axis A_k , X is reshaped to a matrix with axes $A_k \times N$ where $N = \prod_{i \neq k}^d A_i$, denoted query Q . A copy is made of Q and transposed, to create key $K = Q^T \in \mathbb{R}^{N \times A_k}$. Correlations $M \in \mathbb{R}^{A_k \times A_k}$ between features are then calculated with some alignment function σ , i.e. $M = \sigma(QK)$. For example, a common choice of σ is the Softmax function,

$$M = \frac{\exp(QK)}{\sum_l^{A_k} \exp((QK)_l)} \quad (5.1)$$

where subscript l denotes matrix column. Scaling of correlations according to feature strength is achieved through a further matrix multiplication between M and an additional reshaped copy of X , termed value V . This quantity is then reshaped to the original axes. The final attention output $E \in \mathbb{R}^{A_1 \times A_2 \times \dots \times A_d}$ is given as,

$$E = h(X) = \gamma \cdot MV + X, \quad (5.2)$$

where γ is a learned parameter initialised to zero which further controls the strength of correlations.

Typically, attention is measured across feature channels, referred to as channel-wise attention, or pixels, referred to as positional attention. For the latter operation, height and width axes are combined into one single axis containing all the present pixels, and convolutional layers are placed before the reshaping operation to create query Q , key K and value V . In combination, these operators are able to capture different interdependen-

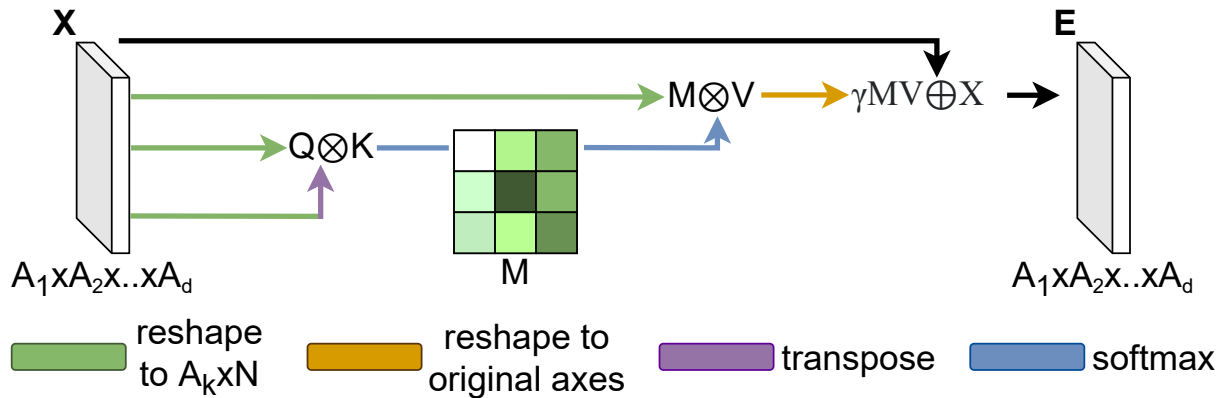


Figure 5.2: Diagram of the general attention module implemented in this work. Note that with positional attention, a convolutional layer is placed before each initial reshaping operation (green arrows). Matrix multiplication is denoted by \otimes , element-wise addition is denoted by \oplus .

cies among features: we refer to this combination as dual attention. After each attention based feature map has been computed, they are combined simply through element-wise summation.

We use an adapted attention framework described in [182] to further enhance semantic representation in feature maps. In this approach, two attention operators applied in succession are used to refine the generated feature maps. While attention inherently prioritises information relevant to class separation, referred to as semantic information, this effect can be encouraged more explicitly through additional regularising losses. For the first module, attention operator output $h_1(X)$ is scaled by the original input feature map passed through a small encoder-decoder style network, $g_1(X)$, i.e. $E_1 = h_1(X) \odot g_1(X)$ where \odot represents element-wise multiplication. The second attention module functions similarly to the first, however input is formed by multiplying previous attention module output E_1 by the original feature map X ,

$$E_2 = h_2(E_1 \odot X) \odot g_2(E_1 \odot X). \quad (5.3)$$

From the two chained attention modules, an auxiliary loss is constructed from two constraints. In the first loss L_V , the encoded representations of g_1, g_2, g_1^V, g_2^V are forced to be close together. For the second loss L_S , the outputs of g_1, g_2 are forced to be close

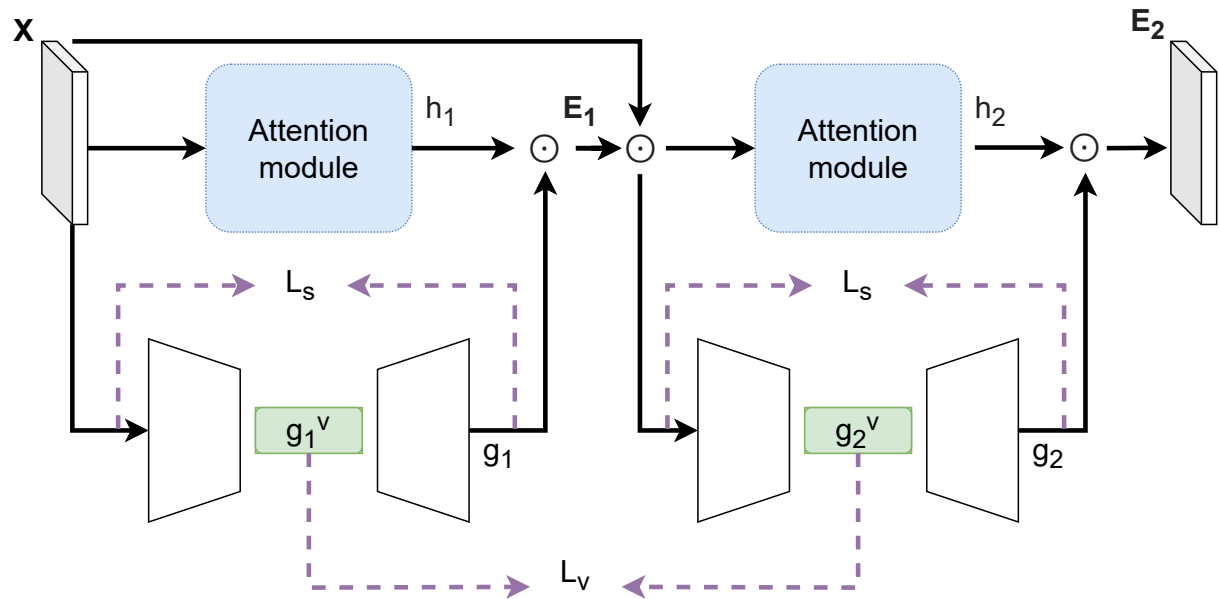


Figure 5.3: Diagram of the guided attention block implemented in this work. Element-wise multiplication is denoted by \odot .

to their inputs. Both of these constraints are enforced by minimising per-pixel error:

$$L_{\text{aux}} = L_V + L_S = \left(\|g_1^V - g_2^V\|_2 \right) + \left(\|g_1(X) - X\|_2 + \|g_2(E_1 \odot X) - E_1 \odot X\|_2 \right). \quad (5.4)$$

Autoencoder networks such as g_1 and g_2 must reconstruct their input from a learned low dimensionality encoding, the success of which is ensured by minimising L_S . Through this process, the encoding must preserve the most descriptive features such that the input can be recovered as accurately as possible. A side effect of this is that features of low relevance or with little discriminative information are discarded, such that the reconstructed outputs of g_1 and g_2 contain proportionally more semantic information. By minimising the difference between each autoencoder’s encoded representation, this semantic information is retained across successive attention modules.

5.2.1.2 Gridded attention

The ability to generate multiscale attention maps from an entire image is beneficial for segmentation of global contaminants. A consequence of using additional spatial scales is that computational resources are increased by some factor for each considered scale. This problem is compounded in models using positional attention as each of these models uses memory proportional to squared image size. These factors are particularly relevant

to segmentation of cirrus in LSB images, as images are very large ($>5000\text{px}^2$), making it difficult to process an entire image with a single pass of the model. While it is possible to manage this effect through downscaling and cropping, the former compromises key local textural information and the latter compromises key global contextual information. It is therefore desirable to use multi-scale attention and minimise the use of both downscaling and cropping.

We propose a computationally inexpensive method for computing attention over multiple spatial scales. We divide each set of multiscale features into groups of tiles so that spatial size is standardised across all scales according to the feature set with the smallest spatial size. By using different grids for separating features of different spatial scales into regions, we obtain tiles of consistent smaller spatial size but with multiple underlying spatial scales. Tiles are then treated as an individual sample similar to a batch of different images. Our architecture then consists of multiple network branches, each handling a different spatial scale and composed of a separate attention module, similarly to [182], though with gridded Gabor attention. In addition, where as [182] upsample smaller scale features independent of other scales, we tie the upsampling process across scales. We illustrate this gridded attention mechanism in Figure 5.4. With this approach, due to the massive saving on computational resources, the model is able to inspect both local texture and global context of an entire LSB image without requiring downscaling images or only using a small region of an image. After attention maps have been generated per scale set, tiles are then reassembled to recover the entire grid and smaller scale feature maps are up-scaled to the original resolution with upsampling blocks consisting of upsampling bilinear interpolation followed by two convolutional layers. In addition, the original feature maps are up-scaled as necessary to the original resolution with similar upsampling convolutional blocks. During initial experiments we found that other upsampling methods introduced artefacts, such as a checkerboard effect produced by transposed convolutions even after carefully setting stride and padding values (see [154]).

To further simplify computation and utilise weight tying across scales, we use spatial scales with a common factor. With this constraint, each underlying grid which is used to disassemble feature maps into tiles shares dividing common boundaries. In this way, each rescaling operation can be composed as a smaller common rescaling operation which is successively applied depending on the desired final scaling. For example, if the desired set of spatial scales are N , $N/2$ and $N/4$ (with N as pixel size), the smallest scale can be achieved by downscaling twice by a factor of 2. An important aspect of the upscaling process is that one upscaling convolutional block is used per scale transition, tying weights across different scale branches of the model. With this choice of spatial scales, all rescaling operations also benefit further parallelised computations.

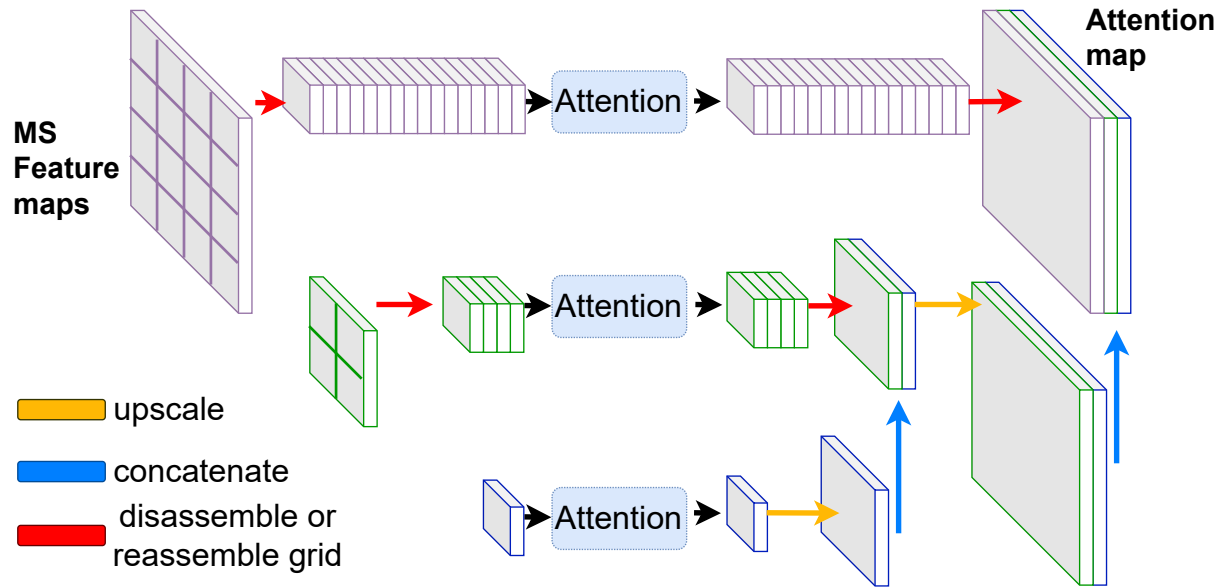


Figure 5.4: Diagram of the proposed gridded attention mechanism. Here number of scales $s = 3$, and common downscaling factor $f = 2$.

It is possible to perform an analysis of the runtime memory usage of gridded attention with spatial scales sharing a common factor. As previously stated, standard attention has memory usage proportional to N^4 , where N is image height or width (for simplicity we assume unit aspect ratio). We perform attention on tiles with resolution equal to that of the smallest scale image: for s different scales with common downscaling factor f , this resolution can be written as $(\frac{N}{f^{s-1}})^2$. Then for each spatial scale considered, with $i \in 0, \dots, s-1$ representing each scale and $i = 0$ as the smallest scale, the number of tiles at a given scale be expressed as $(\frac{N}{f^{(s-1)-i}})^2 / (\frac{N}{f^{s-1}})^2 = f^{2i}$. The total number of tiles is then $\sum_{i=0}^{s-1} f^{2i}$, resulting in a memory usage of,

$$\sum_{i=0}^{s-1} f^{2i} \left(\frac{N}{f^{s-1}}\right)^4 = N^4 \sum_{i=0}^{s-1} \frac{f^{2i}}{f^{4(s-1)}}. \quad (5.5)$$

In our case, with $s = 3$ and $f = 2$, runtime memory consumption becomes $\frac{21}{256}N^4$, equating roughly to a 10 fold reduction in runtime memory usage for attention calculation.

Finally, it is necessary to decide how to generate multiscale features to feed into each network branch. One approach is to retrieve features from intermediate layers of some convolutional network composed of blocks separated by downsampling pooling layers, which we refer to as the backbone network. This approach fits nicely with the choice of spatial scales, as features obtained this way similarly successively downscale with chained pooling operations. While this is efficient, intermediate layers may contain less relevant

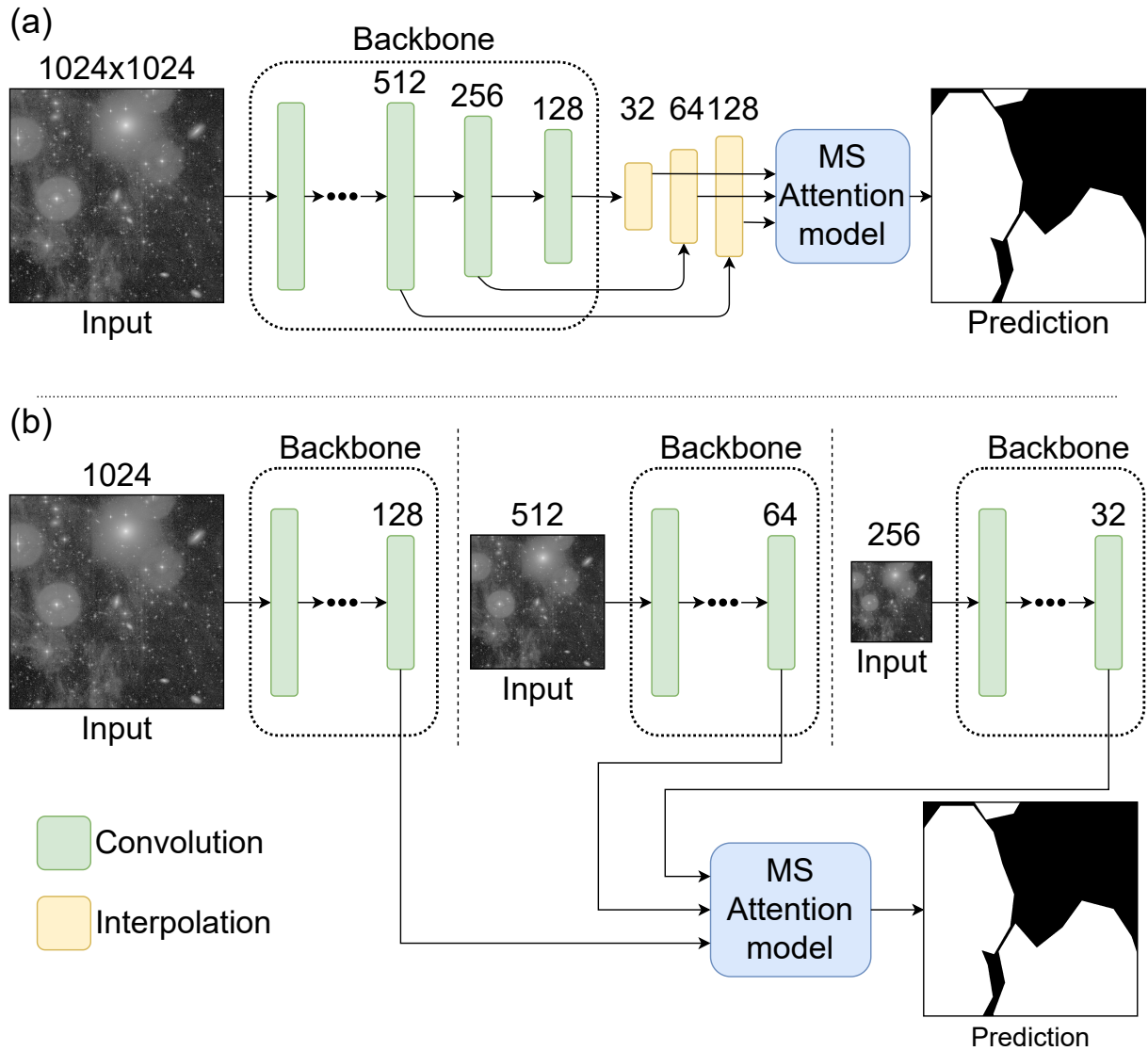


Figure 5.5: Different strategies for generating multiscale features, with number of scales $s = 3$, downscale factor $f = 2$ and image size $N = 1024$. (a) Features are taken from intermediate layers; (b) downsampled copies of the input image are created and then each fed into the backbone separately.

features that would be discarded through the not yet performed pooling layer. Another approach is to first downscale the image into multiple scales through simple spline interpolation and then to pass each rescaled copy through the backbone separately. By tying all weights across all spatial scales, the model may be exposed to a consistent view of objects that appear in different sizes, i.e. scale invariance is encouraged. A downside of this approach however is the computational cost of generating features with the backbone network for each scale. These strategies are illustrated in Figure 5.5. We implement both approaches into our attention framework and later compare results.

5.2.2 Gabor Attention

We attempt to encode some understanding of long range orientational patterns in images, to aid the segmentation model’s ability to spot large contaminants. Local textural information alone may not be sufficient for reliable segmentation of large contaminants, due to variation in textural patterns and possible confusion with other classes that can appear similar locally. To handle these issues, we investigate integrating local and global correlations between orientational features into our segmentation model. In this way, the model can more effectively use the context surrounding uncertain areas to inform the final prediction.

We propose a novel attention operator for studying orientational context. In Chapter 4 we demonstrated that learnable Gabor modulated convolutions significantly increased performance on synthesised cirrus images. We integrate such convolutions into the proposed multiscale attention architecture by calculating attention over the new tensor axis denoting different Gabor orientation parameters. This attention operator is used in combination with the channel and positional operators, creating an attention module with three separate attention operators each measuring different interdependencies among features.

5.2.2.1 Tri-attention module with orientation-wise attention operator

Using the attention mechanism described in Section 5.2.1.1, we compute attention with respect to features dependent on different orientations. In this approach, the Gabor attention operator requires feature maps with a new axis representing the orientations used by modulating Gabor filters. By measuring correlations across this Gabor axis, we are able to study relationships among orientation rich features. To create the required tensor axis, we simply place independent non-cyclic learnable modulated Gabor convolutions before each initial reshaping operation in the attention operator, as shown in Figure 5.7. This choice of a single non-cyclic layer is motivated by several factors. First is the necessity to minimise computational cost, as the new axis multiplies runtime memory usage by the number of orientations considered, or by the square of this number in the cyclic case. Second is that the problem only demands rotation invariance, for which cyclic Gabor convolutions do not offer a significant performance increase in comparison to the non-cyclic version. Finally this choice keeps the architecture flexible, allowing for example the use of any backbone network, or simple comparison between the choice of attention.

The proposed Gabor attention operator can be easily integrated into any attention module. We add Gabor attention to the dual attention framework, and calculate Gabor attention maps in parallel with positional and channel-wise attention maps. These three attention maps are then combined with element-wise summation. We refer to this new

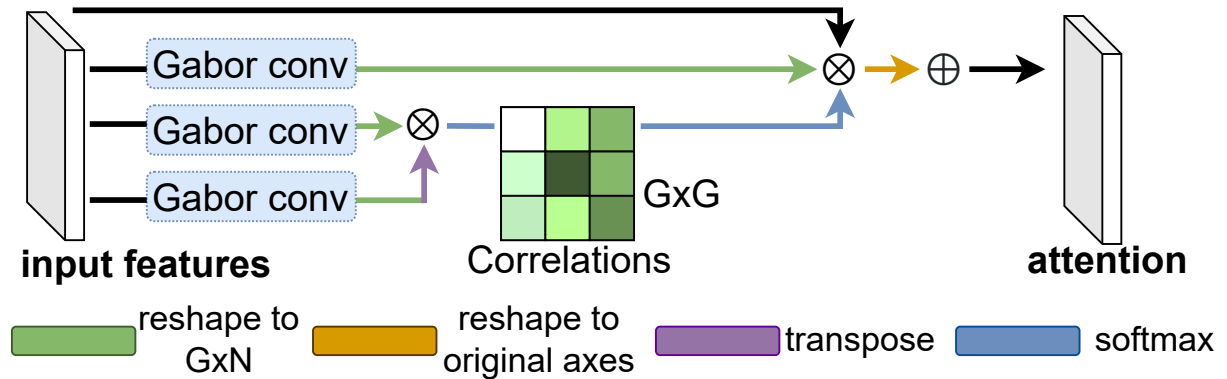


Figure 5.6: Proposed Gabor attention operator. G is number of modulating Gabor filters, N is the product of other axes. Matrix multiplication is denoted by \otimes , and element-wise addition by \oplus . An intermediate correlation result is shown.

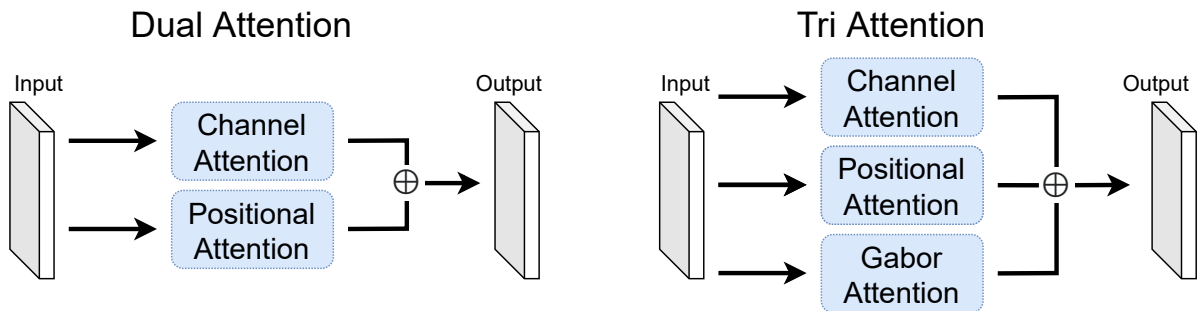


Figure 5.7: Diagram of dual attention versus tri-attention.

attention module as tri-attention.

We also implement versions of the positional and channel-wise attention operators which are compatible with outputs with a Gabor axis. This allows the investigation into whether introducing orientational information into other correlation measurements is beneficial for cirrus segmentation. This is trivially achieved with channel-wise attention through the reshaping described in Section 5.2.1.1. For positional attention we replace the convolutional layers used to generate query, key and value with a non-cyclic Gabor convolution. These variants are later compared to their traditional counterparts in Section 6.4.2.

5.2.3 Constructing a Segmentation Model with Attention

The attention modules presented in this work are versatile and can be substituted in place of any existing attention module, combined with any feature generating backbone, and used to analyse any arbitrary scale or number of scales. These factors make the

proposed methodology applicable to a large variety of computer vision models and potentially valuable to any task involving very large images with local and global feature dependencies.

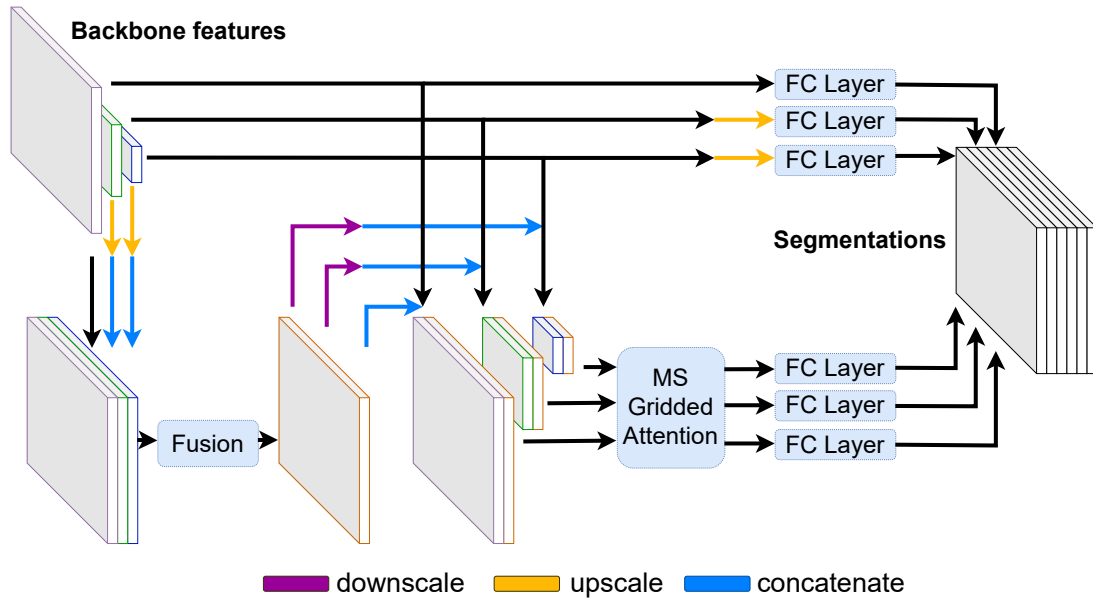


Figure 5.8: Fusion layer and segmentations generated from both attention maps and initial backbone features.

We implement a multiscale attention architecture similar to [77, 182], with multiple considered spatial scales (a comparative analysis is performed later to study the effect of changes the number of scales and their scale factor). Images are first passed through a backbone network in order to generate multiscale features, according to one of the two processes described in Section 5.2.1.2. In Figure 5.8 we illustrate an overview of our model, where multiscale features are generated by passing the input image through the backbone three times: one with original scaling, one downscaled by a factor of 2, and one downscaled by a factor of 4. For ablation studies we use a simple backbone network with four convolutional blocks, with each block consisting of a convolutional, batch normalisation, ReLU and max pooling layer. In final experiments we use ResNet-50 [90] as a backbone. A further fused feature map is created by upscaling features of each scale to the largest size present, concatenating, and passing through a fusion layer. This fusion layer is composed of three convolutional blocks, each consisting of a convolutional, batch normalisation, and ReLU layer. This fused feature map is then concatenated onto each feature map after being rescaled by the feature map’s corresponding scale factor.

The three feature maps created from the backbone network and fusion process are passed through separate guided attention modules after being disassembled into tiles of common size $H/4 \times W/4$, where H, W represent height and width, respectively. The

resulting attention map tiles are then reassembled into three feature maps of size $H \times W$, $H/2 \times W/2$ and $H/4 \times W/4$. Following this, each attention map and original feature map created by the backbone network is upsampled with convolutional blocks as described in Section 5.2.1.2. The original feature maps and attention maps are then passed through a fully connected layer to generate six segmentation predictions. For training, we treat each prediction equally and measure loss in the same way for all predictions. In this way, the backbone is explicitly forced to preserve spatial locations of features, relieving the attention section of the network any realigning effort. Provided the backbone network can produce a reasonable segmentation, the only job of the attention section is to utilise attention to refine the segmentation prediction. For inference, we take the average of the three segmentation predictions computed from the attention maps.

5.3 Results

Modifications detailed throughout this chapter are validated in this section, including those proposed in this work and other related works. We seek to discern the effect of different techniques, and carry out experiments on two tasks performed in Chapter 4. Experiments are first performed on synthetic and real low surface brightness images, where we investigate the benefit of incorporating the ability to study long range interdependencies when segmenting large contaminants. The proposed model is then evaluated on the segmentation of cloud cover in natural images, using the SWIMSEG dataset [53].

5.3.1 Segmentation of Cirrus Dust

We assess the performance characteristics of different modifications on the cirrus segmentation task. Multiple comparative analyses are performed in order to identify the best performing setups of methods detailed in this chapter. All experiments are performed on both real and synthesised images with 5-fold cross validation. The ability to study global context is a major motivation of this work, thus for both datasets we ensure the model is trained on large images, in comparison to the previous chapter. For synthesised images, we generate images similar to Chapter 4 but with size 1024px^2 , representing a 16x increase in pixels. This dataset contains 300 images in total, with 160 for training, 40 for validation, and 100 for testing. For real images, we use the same dataset as is detailed in Chapter 4, involving 48 2-channel LSB images of approximate size 5000px^2 . However, instead of training on small randomised image crops, we downscale the entire image with bilinear interpolation to the maximum size the model can accomm

<https://www.overleaf.com/project/6261bd1d82f354184bb6da04odate>

(with batch size fixed across experiments). Targets are obtained by taking labels from a single expert annotator. Of the 48 images, 32 are used for training, 8 for validation and 8 for testing.

Training setup is fixed across all following experiments to ensure a fair comparison. Networks are trained for 200 epochs over 5 splits using the Adam optimiser [111] and a binary cross entropy loss function. We set learning rate to 10^{-4} which is decayed with an exponential schedule of 0.95 per epoch, and apply an L2 weight regularisation penalty of 10^{-7} . On both datasets, augmentation is employed in the form of random flips and 90° rotations to mitigate against limited sample size.

In this first experiment, the performance of modifications proposed in [182] is validated. While the cited work achieves boosted performance on medical images, we wish to confirm whether this relationship extends to cirrus samples given that segmented structures in [182] are not translucent contaminants. An ablation study, shown in Table 5.1, is performed with three modifications: concatenating a fused feature map onto multi-scale features before calculating attention; generating multi-scale features from intermediate layers of the backbone versus multiple copies of the input image with different scales; and the use of guided attention. Isolated experiments show that fused features and guided attention respectively increase performance over the control dual attention model by 0.15 and 0.23 on synthesised data, and 0.23 and 0.41 on real data. Guided attention appears to offer a larger benefit on real images, possibly because the generalisation effect offered has a larger impact on the smaller sample size of real low surface brightness images. The benefit of these components is also consistent across all combination runs. We observe that using multiscale features from intermediate layers of the backbone appears to decrease performance on real LSB images but increase on synthesised images: this pattern is demonstrated on the isolated experiment and also in combination with the other components. This could be because intermediate layers do not generate features that are discriminative enough for the additional detail in real images, thus the higher expressivity afforded by deeper layers outweighs the benefit of using multiscale features with an inherent hierarchical relationship.

Next, the performance of gridded attention is analysed. Based on the previous experiment, fused features and guided attention are incorporated into the base model of this experiment. A simple grid search experiment, shown in Table 5.2 is performed in order to examine segmentation scores as the number of scales s and downscaling factor f parameters are altered. We see that gridded attention offers an increase in segmentation performance on the large cirrus images in almost all setups. Exceptions to this pattern occur with $s = 4$ and $f = 3$, possibly due to the common tile size becoming too small resulting in excessive erosion of semantic information in larger scale tiles and no added

Fusion	Inter.	Guided	Synth	Real
			0.805 ± 0.0008	0.700 ± 0.024
✓			0.820 ± 0.0006	0.723 ± 0.021
	✓		0.808 ± 0.0009	0.694 ± 0.028
		✓	0.828 ± 0.0011	0.741 ± 0.022
✓	✓		0.823 ± 0.0007	0.718 ± 0.027
	✓	✓	0.830 ± 0.0009	0.734 ± 0.029
✓		✓	0.843 ± 0.0006	0.765 ± 0.020
✓	✓	✓	0.846 ± 0.0010	0.752 ± 0.026

Table 5.1: Ablation study of modifications presented in [182]: fusing scales (see Fig. 5.8); generating multiscale features from intermediate layers; and guided attention. Results presented as mean IoU over 5 splits on real and synthesised cirrus samples.

		$s = 2$	$s = 3$	$s = 4$
Synth	$f = 2$	0.866 ± 0.0012	0.902 ± 0.0011	0.862 ± 0.0014
	$f = 3$	0.875 ± 0.0009	0.886 ± 0.0013	0.831 ± 0.0025
Real	$f = 2$	0.796 ± 0.022	0.842 ± 0.018	0.781 ± 0.021
	$f = 3$	0.819 ± 0.023	0.815 ± 0.016	0.754 ± 0.034

Table 5.2: Segmentation scores of different gridded attention models on synthesised and real data. Here s represents the number of scales and f denotes the downscaling factor. Results presented as mean IoU over 5 splits on real and synthesised cirrus samples.

benefit to global context in smaller scale tiles. This seems likely when considering that tile width and height are $f^s = 3^4 = 81$ times smaller than the largest MS feature map, roughly representing 0.02% of the feature map’s size. There is a similar but less drastic effect with $s = 4$ and $f = 2$ where performance is higher than non-gridded attention but still lower than other gridded setups, indicating forcing a tile size that is too small is not beneficial to performance. It can be seen that setting $s = 3$ and $f = 2$ achieves the highest performance, with $f = 3$ following closely behind, on both datasets. While all setups with $s = 2$ do increase performance over non-gridded models, the minimal variation in spatial scales appears to be suboptimal.

Thus far discussion surrounding computational savings of gridded attention has been theoretical in nature. We seek to empirically validate such theoretical discussions by measuring runtime and memory usage per batch on cirrus data. Results are detailed in Table 5.3. It is clear that gridded attention significantly reduces memory and runtime across all setups. As expected, increasing either the number of scales s or downscaling factor f reduces both the cost of both resources. Gridded attention with two scales decreases runtime by a factor of 0.37 and 0.24 for $f = 2$ and $f = 3$ respectively. As

		$s = 1$	$s = 2$	$s = 3$	$s = 4$
Runtime (s)	$f = 2$	-	0.260 ± 0.009	0.129 ± 0.003	0.112 ± 0.005
	$f = 3$	-	0.167 ± 0.005	0.092 ± 0.006	0.081 ± 0.002
	Non gridded	0.639 ± 0.015	0.702 ± 0.014	0.704 ± 0.011	0.724 ± 0.034
Memory (MiB)	$f = 2$	-	436	212	100
	$f = 3$	-	292	91	33
	Non gridded	2258	2258	2258	2258

Table 5.3: Runtime (seconds) and memory usage (MiB) of multiscale attention calculation for a single batch, for different gridded attention modules and non-gridded attention.

s increases runtime continues to decrease, following an exponential trajectory. Memory usage shares a similar pattern, though the exponential decrease is more significant, with memory usage approximately being divided by f each time s is incremented. This slight difference in trajectories is likely due to the parallelisation achieved between network branches which is reflected in runtime but naturally not in memory usage. The optimal segmentation configuration, $s = 3$ and $f = 2$, offers a sweet spot in terms of computational cost, with a similar runtime to $s = 4$ setups and memory cost that is far more manageable than non-gridded attention modules. We note that this configuration uses 9.3% of the memory used by non-gridded attention, which very closely matches the number calculated during earlier theoretical discussions (see Eq. 5.5).

We now turn to evaluate the performance of Gabor attention and its different implementations. In this experiment, we use the optimal gridded attention configuration with $s = 3$ and $f = 2$ in combination with multiple attention operators: dual attention; dual attention with Gabor convolutions inserted before each reshaping operation (see Fig. 5.2 and Fig. 5.7), denoted Gabor dual attention; Gabor attention in addition to dual attention, denoted tri-attention; and Gabor dual attention in addition to dual attention, denoted Gabor tri-attention. We observe that there is a slight detrimental effect when calculating positional and channel wise attention from features generated with Gabor modulated convolutions, with Gabor dual attention underperforming dual attention and Gabor tri-attention underperforming tri-attention. There is however a clear increase to segmentation scores when using tri-attention, with performance increased by 0.25 on synthesised data and 0.32 on real data. This significant difference likely demonstrates that Gabor-wise attention increases the model’s ability to handle orientational patterns, given that oriented streaks are a major discriminating factor of cirrus regions.

Finally, we perform an ablation study involving the optimal components from the previous comparative experiments. We seek to investigate the effect of different components in combination. We test the application of fused features and guided attention, gridded

	Dual	GaborDual	Tri	GaborTri
Synth	0.902 ± 0.0006	0.891 ± 0.0010	0.927 ± 0.0006	0.923 ± 0.0008
Real	0.842 ± 0.018	0.835 ± 0.020	0.874 ± 0.016	0.857 ± 0.021

Table 5.4: Comparison of attention frameworks: dual attention [77], dual attention with Gabor conv. features, tri attention, tri-attention where channel and positional attention use Gabor conv. features. Results presented as mean IoU over 5 splits on real and synthesised cirrus samples.

Fuse+Guided	Gridded	Tri	Synth	Small Synth	Real
			0.805 ± 0.0008	0.830 ± 0.0009	0.700 ± 0.024
✓			0.843 ± 0.0006	0.859 ± 0.0017	0.765 ± 0.020
	✓		0.868 ± 0.0012	0.822 ± 0.0031	0.796 ± 0.026
		✓	0.841 ± 0.0014	0.867 ± 0.0024	0.782 ± 0.020
✓	✓		0.902 ± 0.0011	0.844 ± 0.0027	0.842 ± 0.018
	✓	✓	0.896 ± 0.0007	0.850 ± 0.0027	0.850 ± 0.023
✓		✓	0.884 ± 0.0008	0.883 ± 0.0021	0.823 ± 0.019
✓	✓	✓	0.927 ± 0.0006	0.861 ± 0.0023	0.874 ± 0.016
U-Net [172]			0.638 ± 0.0678	0.806	0.673
LGCN			0.664 ± 0.0731	0.898	0.715

Table 5.5: Ablation study of the proposed gridded attention and tri-attention, and modifications of previous work [182]. Results presented as mean IoU over 5 splits on three cirrus datasets: large synthesised images (used in previous experiments), small synthesised images (used in Chapter 4), and real LSB images.

attention and tri-attention on real and synthesised cirrus images, shown in Table 5.5. We also test each model on the synthesised dataset used in Chapter 4 containing smaller images, in order to fairly compare performance against non-attention segmentation models. The most significant performance increases on large images are obtained with the addition of gridded attention, likely owing to the minimal downscaling required to process images, thus retaining local textures. This is supported by the fact that these same increases are not observed on the smaller synthesised images across all ablations involving gridded attention, where downscaling is not necessary and gridded attention only decreases global receptive fields. There is a strictly positive effect across all combinations of components on larger images, with the use of all components scoring 0.874 and 0.927 mean IoU on the real and synthesised cirrus datasets, respectively. In comparison to the non-attention based method of the previous chapter, LGCN, these results reflect a large relative increase of 22% on real data and 39% on synthesised data.

5.3.2 Cloud segmentation in natural images

The proposed model is evaluated on the Singapore Whole sky IMaging SEGmentation database [53] (SWIMSEG), and tasked with segmentation of clouds in natural images. This task is highly relevant to the proposed methodology, as the imaged clouds are global structures with high variation. Both local information and global context are key to good performance on this dataset, as difficult positive regions of light cloud can only be correctly identified based on subtle textural patterns and comparison with surrounding regions. The SWIMSEG dataset contains 1013 images of sky patches and corresponding binary cloud segmentation maps, with each image containing 600×600 RGB pixels. Training, validation and testing sets each contain 861, 101 and 51 samples, respectively.

The optimal architecture from the previous section is selected with gridded attention and tri-attention. In addition, we train three further models for comparison, ablating gridded and tri-attention. To make a fairer comparison with previous works we swap our feature generating backbone with a ResNet-50 [90] and increase the number of parameters in following layers. The Adam optimiser [111] is used with learning rate is set to 10^{-4} , exponential learning rate decay to 0.99, and weight decay to 10^{-4} . Networks are trained over 200 epochs for a single fold. Data augmentation is employed in the form of random flips and 90° rotations, and slight adjustment of contrast, brightness, saturation and hue.

Results are detailed in Table 5.6, where we report the segmentation IoU and Dice score on the testing set as in [53, 54, 186]. The base model with no gridded or Gabor attention scores marginally below state of the art [186]. Gridded attention and Gabor attention each improve on this score by absolute differences of +0.3 and +0.2 IoU, respectively. These increases demonstrate that the proposed methodologies each improve the model’s ability to process large homogeneous structures. The combination of these components achieves the best performance with respective IoU and Dice scores of 0.90 and 0.95, representing a significant improvement over state of the art results. Several examples randomly chosen from the testing set are displayed in Figure 5.9. The proposed model produces reliable and accurate segmentations of cloud cover, even in highly translucent areas. There are some areas where the predicted boundaries are more coarse than the ground truth boundaries. Thus, it is likely that incorporating skip connections similar to U-Net [172] style architectures, where fine structural details produced in intermediate backbone layers are provided to the final mask generating layers, could improve performance further. We leave such optimisations to future work, and assert that results presented demonstrate the effectiveness of the proposed method in segmenting global structures.

	Base	+Gridded	+Tri	All	Dev et al. [53]	Dev et al. [54]	Song et al. [186]
IoU	0.85	0.88	0.87	0.90	0.69	0.80	0.86
Dice	0.92	0.94	0.93	0.95	0.82	0.89	0.92

Table 5.6: Segmentation scores on the SWIMSEG sky/cloud segmentation dataset, comparing the proposed gridded attention and tri-attention against previous works.

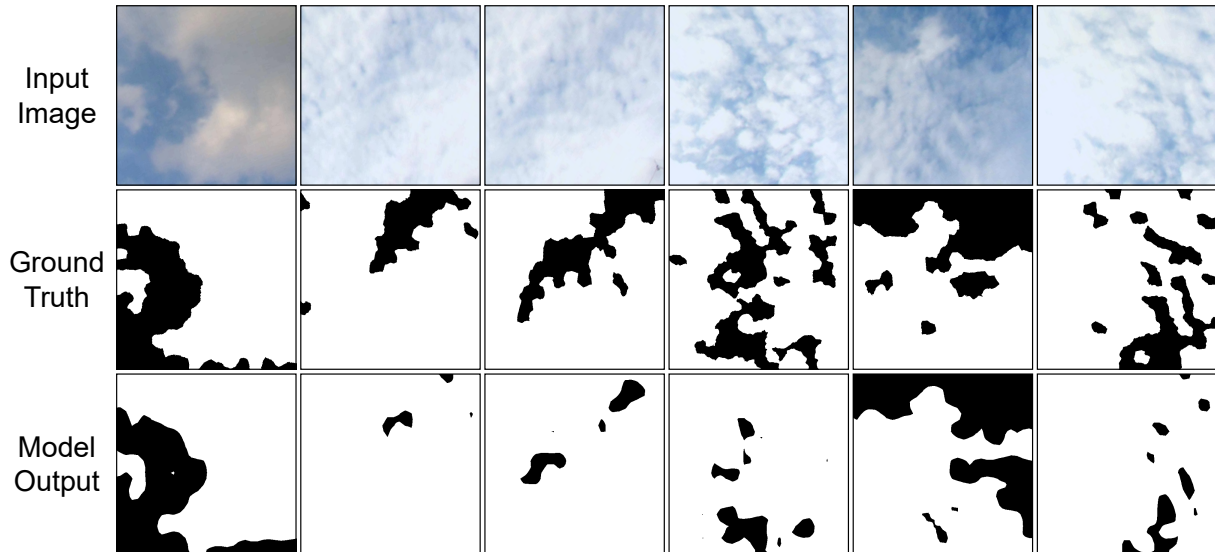


Figure 5.9: Sky/cloud examples from the SWIMSEG testing dataset (top), corresponding ground truth cloud segmentations (middle), and model predictions (bottom).

5.4 Summary

We presented a computationally efficient multi-scale attention architecture that is sensitive to texture orientation for segmentation of global contaminants in large images. Efficiency is achieved through a gridded architecture, allowing global context to be considered while retaining local textural information. Multi-scale features are divided into tiles of size equal to the smallest feature map’s size so that spatial size is constant across all tiles. Attention is computed on each tile and then tiles are stitched back together to retain the original feature map size. As attention has memory usage proportional to the squared image size, computing attention on smaller tiles hugely reduces computational cost. We created an attention operator that measures correlations between orientations by utilising Gabor modulated convolutions to generate orientation-dependent features. Attention is then computed with respect to the different orientations of modulating Gabor filters. These contributions were combined into a new state-of-the-art model for the segmentation of global contaminants in large images. Our model can process multiple entire images of spatial resolution $>1024\text{px}^2$ in a single pass, meaning that the proposed method can

be easily integrated into data processing pipelines for imaging instruments to obtain contamination masks.

The proposed methodology was validated on multiple datasets through a set of comparative analyses. The exact optimal configuration of our methodology was discovered through a grid searches and a following ablation study. The computational characteristics of gridded attention were analysed empirically, with findings concurring with theoretical discussions. We firmly showed that gridded attention and tri-attention improve performance on real and synthesised cirrus samples, indicating the ability of the proposed method to consider both local information and global context simultaneously. State of the art performance was achieved on a sky/cloud segmentation dataset, SWIMSEG, demonstrating the effectiveness of the proposed model in segmenting difficult global structures.

In this chapter we achieved reliable cirrus segmentation performance on a small dataset of real LSB images. An important note is that this dataset of real LSB images, while helpful and informative for validating methodologies, is slightly ill-conditioned as targets are formed from a single annotator’s binary labels, rather than a consensus of all annotators. Furthermore, we trained on only images containing cirrus and thus the potential use of our methodology as an automated catalogue method has not been verified. In the following chapter, we attempt to address these limitations and develop methodologies specific to training supervised ML models on LSB images.

Chapter 6

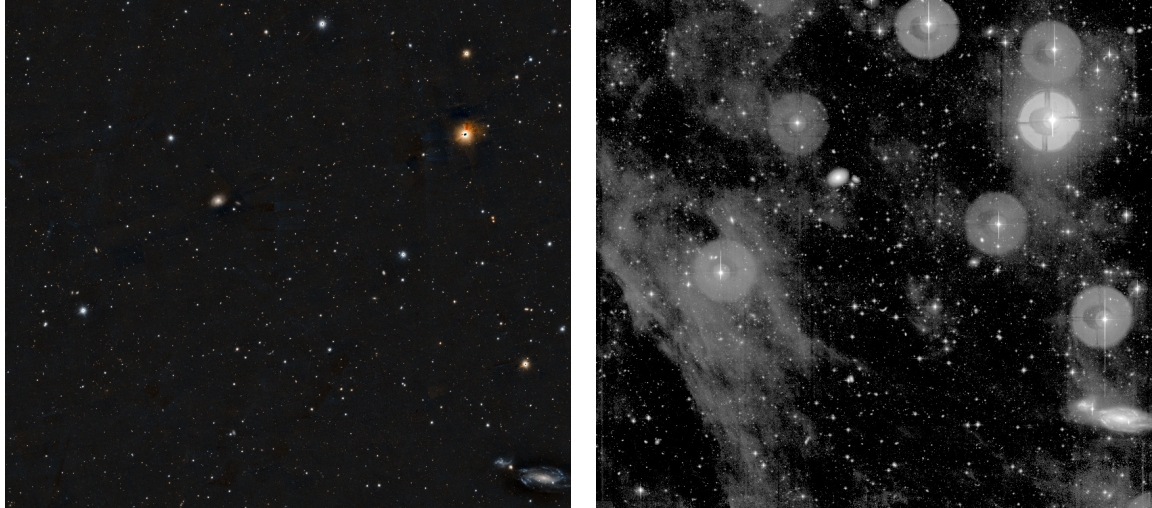
Segmentation of Cirrus Contamination: a Deep Learning Approach

In this chapter, we refocus our attention purely on the annotated LSB data presented in Chapter 3 and further investigate automated segmentation of Galactic cirri. We build upon the attention model proposed in Chapter 5 in an attempt to specialise the machine learning method to handle LSB images. Several novel method components are proposed to address inherent challenges associated with identifying galactic cirrus structures in LSB images. Results of the proposed methodology on the annotated MATLAS images detailed in Section 3.3 are presented in the form of multiple ablation studies. In addition, we thoroughly review predictions generated by the final method on a selection of unseen MATLAS images.

6.1 Introduction

Identifying cirrus contamination in low surface brightness imaging is a major priority for future astronomical surveys. Galaxies captured by traditional instruments in the optical and near-optical bands do not suffer from such contamination; cirrus is typically captured in the infrared band due to thermal emission. The sensitivity of LSB imaging captures scattered light from dust in cirrus clouds, even in the optical and near-optical bands, as illustrated in Figure 6.1. These captured cirrus structures appear in the foreground of some galaxies and contaminate the image in a degree ranging from a slight change in background level to occlusion of interesting structures within the target galaxy. Contamination of LSB images by cirrus clouds impedes the statistical analysis of LSB galaxies and their fine structures [185], and in some cases can even appear visually similar to tidal features [63]. Cirrus regions captured in LSB pictures can be used for high spatial resolu-

tion studies of the interstellar medium, as was conducted by Miville-Deschênes et al. [148] to investigate turbulence cascade. Additionally, the study of correlating factors between cirrus clouds and nearby stellar structures may direct further investigation into possible associated phenomena [21].



(a) PanSTARRS.

(b) MATLAS.

Figure 6.1: Surrounding region of NGC1253 captured in different astronomical imaging surveys.

At the surface brightness ranges captured by deep surveys such as MATLAS [61], Galactic cirrus emission covers a significant portion of the sky. Comparisons to cirrus observed in IR wavelengths, where the dust’s thermal emission can be captured, has shown that the cirrus is highly likely to be Galactic in nature [62, 195]. While existing studies of Galactic cirrus in far-IR imaging provide a map of affected regions, the resolution is far inferior to modern LSB images and thus greatly diminish the quality of any possible cross examination. Cirrus presents in LSB images as a wispy texture often with filamentary structures sharing a common dominant orientation. These dust clouds do not exhibit a constant color, with the $g-r$ colour index, which is defined as the difference between magnitudes of the g and r bands, varying intra-structure. Because of this, subtraction of this foreground component, for example as is done in cosmic microwave background analysis, is made significantly more difficult. Such removal of polluting structures is thus unlikely to be achieved in the near future.

It is vital that cirrus contamination within LSB images can be identified and delineated in an automated fashion. A typical approach for handling contaminants such as image artefacts or foreground dust is to catalogue these contaminated regions so that they can be excluded from or handled specially during any statistical analysis. Traditionally this

process is carried out manually by expert astronomers familiar with the contaminants, requiring a significant time investment. Duc et al. [62] classified the presence of LSB features, including cirrus, in 92 galaxies from the ATLAS^{3D} survey. Bilek et al. [21] performed a similar classification process on a larger sample set of MATLAS images. Annotation effort becomes especially expensive if precise masking of affected regions is desired. For example, Sola et al. [185] precisely delineate and classify LSB features of 352 galaxies captured by MATLAS and CFIS, using the tool presented in Chapter 3: in this work, authors estimate that annotation took approximately 10 minutes per sample. Thus, manually cataloguing contaminated regions is unfeasible for future surveys producing petabytes of LSB image data. There is a clear need for automated detection of cirrus clouds in LSB images.

Segmentation of cirrus in LSB images currently presents multiple challenges. LSB imaging that boasts both high sensitivity and high resolution is a relatively recent technology, with only a handful of deep surveys each containing a small sample set in comparison to what is typically necessary for training ML methods. Images commonly suffer from various instrument artefacts such as internal reflections surrounding bright stars and incomplete CCD columns causing areas of artificially high background levels. Cirrus pollution can be very difficult to spot in many cases, only distorting the background level slightly. The combination of these factors complicate the training process, and significantly increase the difficulty of attaining generalisation. Even in non-polluted regions of images, background intensity levels vary both between and within images, meaning that it is necessary to process the entire image in one pass to ensure the ML model has global contextual information. A complication for this requirement is that images are very large ($>5000\text{px}^2$) and multispectral, thus the method must be carefully designed to accommodate this computational cost. Finally, annotations of cirrus used as training masks in this study were generated by multiple experts, meaning targets are inherently probabilistic.

We propose a machine learning pipeline for the automated identification and delineation of cirrus structures in LSB images. These images present a multitude of challenges that complicate training, such as their large multi-spectral resolution and frequent instrument artefacts. To handle variations in background intensity we implement a preprocessing layer inside the ML mode where multiple intensity scaling transformations are learned in parallel in an end to end fashion. We propose a novel loss for training on probabilistic consensus of multiple human generated annotations, which is suited for scenarios where the number of expert annotators is limited or there is significant disagreement among annotations. Figure 6.2 shows examples of different contaminated regions along with their uncertain consensus annotations. These novel components are combined with the gridded tri-attention segmentation model presented in Chapter 5, to construct a comprehensive

pipeline for automated cataloguing of galactic cirri.

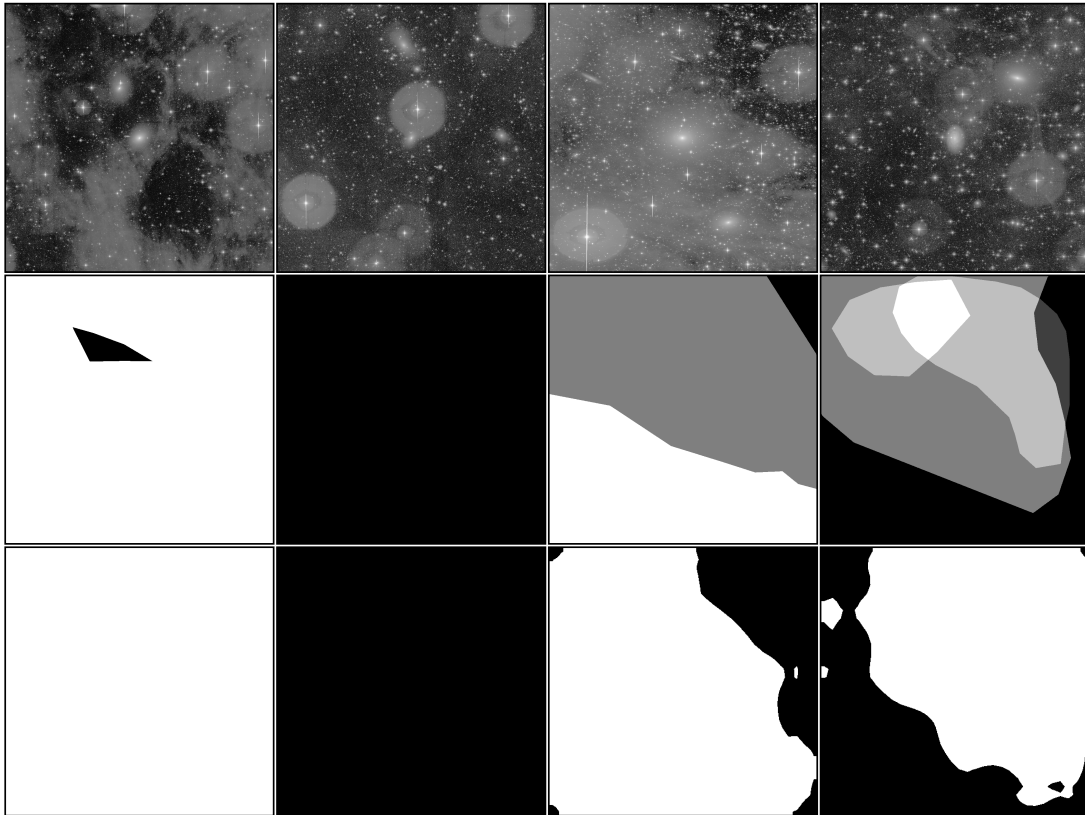


Figure 6.2: Cirrus dust of various strengths (top), with uncertain annotations (middle) and predictions (bottom).

This chapter is organised as follows. We detail our strategy for training on LSB images in Section 6.3, and propose a novel loss for training on probabilistic consensus of multiple human generated annotations. We validate components of our methodology in isolation and in combination in Section 6.4. In Section 6.5 we present results on a dataset of annotated MATLAS images, demonstrating the reliability of our approach. Finally, we provide discussion on the contributions of this work in Section 6.6, and summarise conclusions in Section 6.7.

6.2 Related Work

Modern machine learning techniques have been applied to various astronomy tasks with great success. In particular, numerous authors have exploited deep learning models on image data in recent years. Kim and Brunner [110] perform star-galaxy classification with a combination of images from SDSS [23] and CFHTLenS [67, 93, 94], and pre-computed SExtractor [19] features. Reliable morphological classification with standard CNNs has

been achieved on images from CANDELS [101] and SDSS [59]. Domínguez Sánchez et al. [60] demonstrate the effectiveness of transfer learning for finetuning deep learning models to unseen images from astronomical surveys different to the original survey used for training. Akeret et al. [2] mitigate radio frequency interference using the fully-convolutional U-Net architecture [172] on simulated data.

Works that involve adaptation of the machine learning pipeline to accommodate challenges specific to the data have been particularly successful. Dieleman et al. [55] characterise basic galaxy morphologies using convolutional layers that replicate features with 90° rotations, to aid network understanding of rotational transformations. Pasquet et al. [158] utilise inception blocks [188] in CNNs for regression of photometric redshift, increasing robustness to different spatial scales. Careful design of the ML pipeline is crucial to attaining good results, especially on difficult data.

Attempts to automate cataloguing of cirrus regions has thus far utilised traditional non-ML algorithms. Jackson et al. [104] construct a catalogue of strong filamentary cirrus structures using a differential geometric computer vision algorithm [132]. Akrami et al. [4] manually produce rough masks delineating bright or highly variable regions of cirrus contamination, which is taken into account for statistical classification of contamination surrounding source objects. While such methods have achieved good results on specific types of cirrus contamination, cirrus classification indiscriminate of strength or structural properties is clearly a difficult challenge. Román et al. [171] isolate cirrus contamination in LSB images by exploiting differences in colour signatures between multi-spectral bands, regions closely surrounding source objects cannot be reliably identified due to their strong impact on colour indices.

Object segmentation in astronomical images has also been made possible with deep learning. Bekki [18] utilise a U-Net to segment spiral arms of galaxies in synthesised images. Hausen and Robertson [87] train a U-Net on CANDELS data and masks created with SExtractor to generate a background-object segmentation, which is postprocessed with graph-based techniques [49] to separate objects. Mask R-CNN [91] has been used for segmentation of individual instances of objects by training on SExtractor generated masks [69] and simulated data [31]. Boucaud et al. [27] propose a modified U-Net to simultaneously segment and predict the photometry of high-redshift galaxies. While such studies have achieved high accuracies, segmentation is performed on bright source objects in relatively clean images, while we aim to segment structures with inconsistent geometrical envelopes, textures and brightness. Additionally, models in these discussed works are trained on auto-generated masks and not human generated annotations as is the aim of this study.

6.3 Training on LSB Images

In this section, we detail techniques we use in combination to craft a robust training strategy for training on annotated LSB images, addressing the data-specific challenges related to this work. First, training data is described. Second, we detail the transfer learning and data augmentation used to mitigate against model overfitting and improve generalisation. Third, we present a pixel rescaling operation which we encode as a network layer with parameters that can be learned in an end to end fashion. Finally, we present a novel loss function for consensus of annotations and construct the total loss function used for training our network.

6.3.1 Data

In this study, we use a subset of the dataset presented in Section 3.3.1, containing 186 MATLAS images with two input channels. To construct this subset we first grab all samples with the r -band present, as during the annotation process, annotators reported that the r -band was most useful for identifying cirrus. Of these 186 images, 180 also contain the g -band which we use as the second channel. For the remaining 6 samples without a g -band present, we simply duplicate the r -band to form the second input channel. The average resolution is approximately 6000×6000 px though corners of images often contain null values due to how sections of the sky are captured and pieced together by the instrument. Each image covers a $1^\circ \times 1^\circ$ region in world coordinates, though only $.5^\circ \times .5^\circ$ degrees around the target galaxy is guaranteed to contain no null values. There is a clear motivation to using the entire image for training rather than only the guaranteed non-null region, as there is a significant amount of training data to be gained. These bad pixels are thus ignored for all analysis in this work, including calculation of the training loss so that the model is not trained based on its segmentation of these regions.

During ablation studies, we use 70% of samples as training data and then set aside 15% for validation, which we use to choose the model state that performed best over all training epochs, and 15% for testing the trained model’s performance on unseen samples. This choice allows us to more confidently validate model and training protocol modifications by average over multiple splits, where the validation set is chosen differently in each split. Samples to form the validation and testing set are chosen carefully so that the proportion of images containing cirrus is similar to the training set. Of the total 186 images, 48 contain contain cirrus contamination. The training set is thus composed of 32 cirrus samples and 96 non-cirrus samples, while the validation and test set each contain 8 cirrus samples and 21 non-cirrus samples. We also ensure that the test set contains examples of both strong and weak cirrus contamination.

For final testing, we use 80% of samples as training data and then set aside 20% for the test set. Samples to form the testing set are chosen carefully similarly to the previous setup, so that class balances remain similar between the training and test set, and so the test set contains a good representation of cirrus samples. The training set is composed of 38 cirrus samples and 110 non-cirrus samples, while the test set contains 10 cirrus samples and 28 non-cirrus samples.

In comparative analyses, we also run experiments on another dataset to further validate findings. This is a subset of the aforementioned dataset, composed of only images that contain cirrus contamination, resulting in a sample size of 48 images. Of these 48, 32 are used for training, 8 are used each for validation and testing. While this dataset will not demonstrate the model’s ability to discern cirrus contamination from regular sky as well as the original dataset, there is a benefit to using it for ablation studies. Class imbalance is a major hurdle in this study, as only 25% of images contain any cirrus contamination, and in contaminated images, 60% of pixels contain cirrus. By removing non-contaminated images, we remove the first half of this class imbalance, relieving training efforts of capturing this distribution and hopefully allowing the model to focus further on discriminative features describing cirrus regions. Thus this additional subset may more clearly reveal findings during ablation studies.

6.3.2 Data augmentation and Transfer Learning

The lack of a large and well-balanced dataset is a common challenge encountered in works utilising modern ML algorithms. This is especially the case in this work, where we aim to perform object segmentation on a small dataset of difficult images with only $\sim 15\%$ of all pixels containing cirrus, and $\sim 25\%$ of images containing any cirrus contamination. Typically, to mitigate against dataset limitations, two strategies are implemented in combination. Data augmentation is used, involving applying small transformations to the data such as rotations or translations, exposing the models to more variation in the training set. Transfer learning is also used, where the model is pretrained on some large and balanced dataset before training on the target dataset. Astronomy focused works have extensively made use of data augmentation [59, 110, 158, 201] and transfer learning [31, 60, 69, 101] to improve the ability of trained models to generalise to unseen data. The effective use of both approaches is therefore incredibly important to this study.

Such augmentations must be carefully chosen so that semantic information is not distorted. For example, affine transformations may excessively warp the brightness profile of stars, weakening the model’s ability to process unseen samples. We first extract a random crop of 3000×3000 pixels, representing approximately a $.5^\circ \times .5^\circ$ region. The

following transformations are then applied:

1. Spatial downscale to 1024×1024 px.
2. Random horizontal and/or vertical flip.
3. Rotation by a random multiple of $\pi/2$.

An illustration of example augmentations that can be generated is shown in Figure 6.3. We also experiment with further pixel-level augmentations: adjusting the contrast of either image channel by a random factor between 1 ± 0.02 ; and element-wise addition of Gaussian noise with zero mean and variance of 0.1. While the application of element-wise Gaussian noise may distort semantic information, we note that several astronomy focused works have seen success with the approach provided that variance is small relative to the maximum pixel values of images [31, 69]. This is also the case for contrast adjustment, which was used in [201]. We verify whether the use of these additional augmentations improves generalisation through an ablation study in 6.4.1.

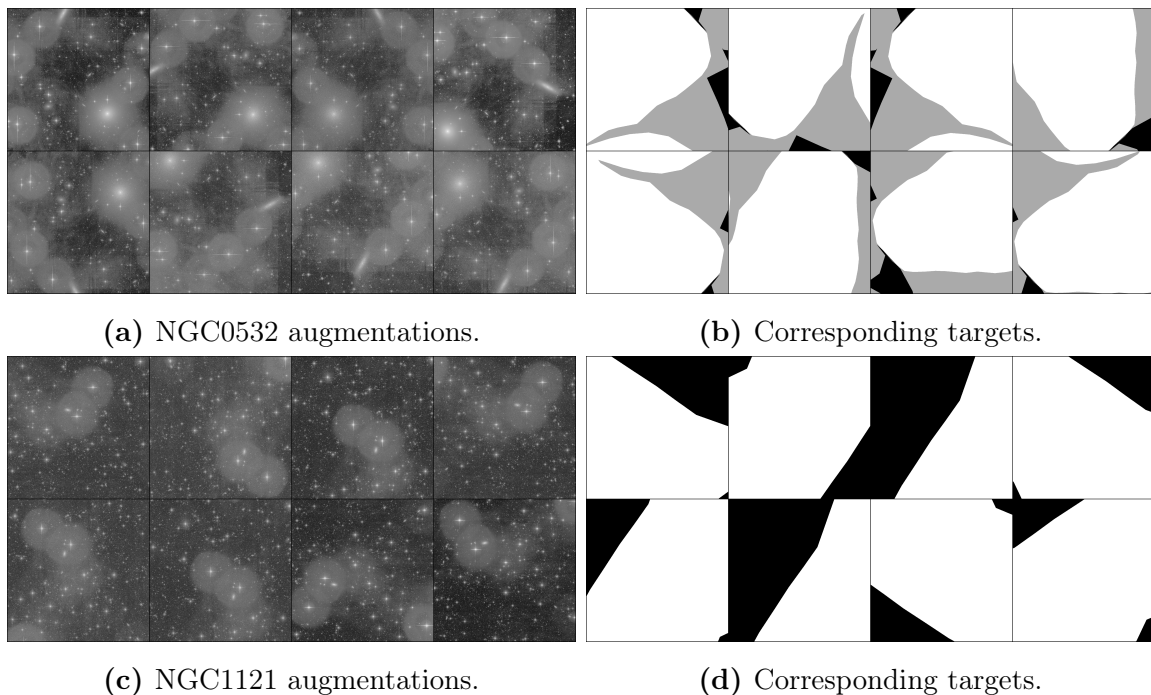


Figure 6.3: Mosaics of augmented images, and their corresponding augmented target masks, generated through our augmentation pipeline.

For transfer learning, we utilise the synthesised cirrus images presented in Chapter 3. We create a dataset of 1200 synthesised images with all difficulty variations applied, described in Section 3.4.2. Before training on LSB images, we train the model to segment

cirrus in these synthesised images. In this way, the model is exposed to a larger sample of cirrus textures and other contaminants before encountering real data, mitigating overfitting.

6.3.3 Adaptive Intensity Scaling

Preprocessing is an important part of any machine learning pipeline, easing the burden of the model by improving the quality of the data. For example, normalisation of input images to have zero mean and unit standard deviation is applied in almost any recent deep learning work based on lessons learned in [123]. Further pixel scaling is commonly used before inspecting astronomy images, as is used in 3.3.1 for annotation. In images' original form, objects other than bright stars are very faint, as the number of photons emitted by the brightest stars is several orders of magnitude larger than other objects such as cirrus clouds. This factor is especially relevant for low surface brightness objects, and Walmsley et al. [200] utilise logarithmic rescaling in combination with pixel clipping in their preprocessing pipeline to train CNNs on LSB images. Such rescaling is particularly necessary in this work to craft a robust training strategy, given that cirrus structures are much fainter than galaxy cores and stars.

To compensate for the large range in pixel values we add an initial layer to the networks implementing arcsinh scaling, popular in astronomical image processing, with learned parameters: $X_s = \text{arcsinh}(aX + b)$, where $a, b \in \mathbb{R}$ are learned. This formulation allows the model to control the strength of the scaling through a while centring values with b . Following this a sigmoid function is applied with a similar weight formulation: $Y = \frac{1}{1 + \exp(-cX_s - d)}$, where $c, d \in \mathbb{R}$ are learned. This formulation acts as a learnable pixel clipping operation, as the sigmoid function is bounded between $[0, 1]$ and inputs > 5 mapping to outputs > 0.99 . Initial scaling parameters were determined from a simple gradient descent algorithm, with the target set as auto-scaled versions of images using Aladin [24, 25] (GPL v3 license).

We design two layers utilising the learnable scaling operation. In the first, multiple scaling operations are learned in parallel to produce multiple scaled versions of each band. Variation in initial weight values is ensured by slightly varying the values generated using the above protocol with a random sample of a Gaussian distribution $\mathcal{N}(0, .25)$. With K different scaling operations, this layer results in $K \times B$ input channels, where B is the number of input wavelength bands. In the second layer design, scaled images are concatenated onto original input to account for oversaturation of very strong cirrus regions. This layer design results in $K \times B + B$ input channels.

6.3.4 Loss function

A necessary component of our training strategy is how the model handles probabilistic annotations. As described in 3.3.2, we combine annotations made by multiple annotators using a weighted average, where annotators with more expertise correspond to a higher weight in the final consensus. As a result, our ‘ground-truth’ samples are not binary labels as is typical in most ML problems. Probabilistic U-Net [116] has been applied to segmentation in medical imaging [95] and astronomy contexts [99]. Walmsley et al. [201] exploit a novel dropout technique to establish Bayesian CNNs with probabilistic output, which in combination with an active learning framework enabled more efficient training on probabilistic annotations. Liu et al. [137] propose a loss function for handling probabilistic annotations, where pixels corresponding to uncertain labels are ignored in the training process.

We propose a loss function to train our attention model on probabilistic annotations. In this study, for target examples we use annotations generated by four annotators as described in 3. These annotations contain pixel values between 0 and 1, where higher values represents a larger consensus between annotators that the pixel contains cirrus contamination. We choose not to use approaches that alter the underlying ML model to become probabilistic, such as [201], due to our comparatively small number of annotators. While approaches such as [95, 116] have seen success with consensus generated from 4 annotators, average annotator expertise was higher than in this work. We instead alter the loss function to be able to handle non-binary target values. Furthermore, as cirrus is only present in roughly a quarter of training images, and occupies a varying portion of each of these images, it is necessary to mitigate against class imbalance issues.

Inspired by works on edge detection with CNNs [137, 213], we separate consensus values into quartiles and conditionally adjust the loss function based on which quartile a pixel falls into. By coarsely dividing probabilities into ranges of confidence, we mitigate against any uncertainty associated with our probabilistic consensus. We also utilise focal loss [131] to encourage the model to focus on hard to classify examples rather than easy to identify negative examples or ‘clean’ pixels which dominate the class balance. With y as the consensus value and p as the prediction, we write focal loss as,

$$\text{FL}(p_t) = \alpha_t(1 - p_t)^\gamma \log(p_t), \quad \text{where } p_t = \begin{cases} p & \text{if } y \geq 0.5 \\ 1 - p & \text{otherwise,} \end{cases} \quad (6.1)$$

With α_t defined similarly to p_t , which scales the loss based on class balance. This can be read as adding a multiplicative factor, $(1 - p_t)^\gamma$, to the standard cross entropy loss function. This multiplicative factor is larger for weak predictions, i.e. low p_t , and smaller

for confident predictions, making the loss larger for examples where the network is unsure of the classification and thus prioritising such cases. Our novel consensus loss is then defined as,

$$L_c = \begin{cases} \beta \cdot \text{FL}(p_t) & \text{if } y \geq 0.75. \\ \text{FL}(p_t) & \text{if } 0.5 \leq y < 0.75. \\ 0 & \text{if } 0.25 < y < 0.5. \\ \text{FL}(p_t) & \text{otherwise.} \end{cases} \quad (6.2)$$

The first and third quartiles, with $0 < y \leq 0.25$ and $0.5 \leq y \leq 0.75$, represent majority consensus for negative and positive examples, respectively. We thus simply use only the focal loss in these scenarios. The second quartile represents uncertain pixels, for which we set the loss to zero. This has the effect of uncertain pixels being ignored by the loss function, and the network is neither encouraged nor penalised based on predictions it makes on these pixels. If this was not the case, there may exist some contradicting patterns where the model is penalised for predicting the existence of cirrus on an annotated region with a weak consensus, even if the region does in fact contain cirrus contamination. Finally, in the fourth quartile we prioritise the loss values on regions with a super-majority consensus by multiplying by a boosting coefficient, which we choose to set as $\beta = 1.25$.

6.4 Comparative Analyses on Proposed Techniques

In this section, we verify the effectiveness of the proposed training strategy and model modifications. This is achieved by carrying out multiple ablation studies in order to assess how different strategies perform. Different training strategies are compared in order to justify our choice of augmentations, the proposed adaptive intensity scaling layer choice, and consensus loss function. We also compare the performance of tri-attention and gridded attention modules similarly to Chapter 5 to verify that findings discovered on the cirrus-only synthesised and real images with non-consensus labels translate to the general LSB image dataset with non-binary labels described in this chapter. All networks are trained for 200 epochs with the Adam optimiser [111] with learning rate 10^{-3} and L2 weight regularisation 10^{-7} . Learning rate is also decreased every epoch by a factor of 0.98. We score different networks using the intersection over union (IoU) metric averaged over five splits, and also report standard error across splits.

6.4.1 Comparing Strategies Specific to Training on LSB images

We validate the proposed training strategies through multiple ablation studies where the underlying model is fixed across experiments. This process allows us to individually assess different components of the training strategy and justify their use. For control model, we use a dual attention framework with positional and channel-wise attention, and a simple feature generating backbone with 4 downsampling convolutional blocks. In these experiments we include results on the two datasets of MATLAS images, the first containing images with and without cirrus contamination, and the second containing only images with cirrus contamination. As ablation studies in this section are concerned with modifications aiming to address challenges specific to training on LSB data, rather than better identifying features associated with galactic structures, we do not include results on the synthesised images.

We also control for training strategy across experiments: models are trained with minimal augmentation, i.e. only steps 1-3 in Section 6.3.2, with no pretraining and with a simple loss function which computes binary cross entropy of rounded consensus probabilities. As choice of intensity scaling layer potentially impacts the optimal set of augmentations, we first find the best scaling layer design with no augmentation. For example, augmentations involving adjusting contrast will have different effects depending on the intensity scaling transformation used. The best scaling layer design is then used in all following experiments.

Comparative results on different scaling layer designs are shown in Table 6.1. The first model uses no intensity scaling; the second uses four learned scaling operations per input channel; and the third uses scaled and non-scaled inputs in parallel, where three learned scaling operations are used per channel and the original input is then concatenated. It can be seen that the parallel technique performs best, and the no scaling model trails closely behind. We observe that there is a significant detrimental effect when using only scaling transformations and not including the original input. This indicates that the learned intensity scaling adds auxiliary information which is helpful in combination with non-scaled images.

	No scaling*	Multiple	Parallel
All images	0.415 ± 0.010	0.370 ± 0.008	0.455 ± 0.012
Only cirrus	0.748 ± 0.021	0.736 ± 0.024	0.803 ± 0.019

Table 6.1: Comparison of different intensity scaling layers. Results reported as mean segmentation IoU over 5 splits. *Control model.

In this experiment, we compare multiple data augmentation strategies. We train five

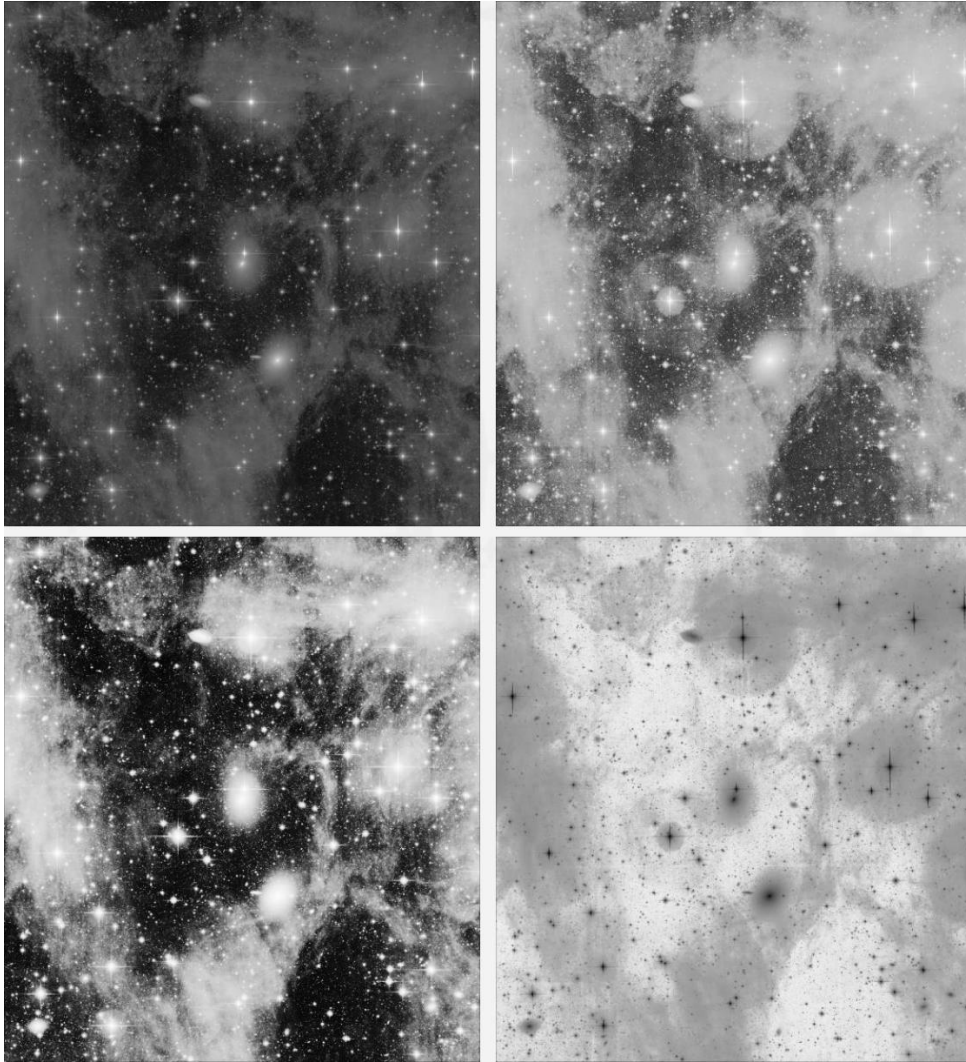


Figure 6.4: Examples of different learned intensity scaling transformations on NGC2592/4.

models, seeking to find the augmentation protocol that leads to the best generalisation on unseen images. In particular, we are interested in whether contrast adjusting and additive noise augmentations aid generalisation. Results are shown in Table 6.2. As expected, we see that augmenting training data through simple geometric transformations is beneficial. Pixel-level augmentations tell a mixed story: element-wise summation of Gaussian noise gives the best test scores on both datasets, while the impact of contrast adjustments is unclear. The combination of both pixel-level augmentations has a detrimental effect on performance. Interestingly, the use of contrast adjustments is correlated with some of the largest margins of error in test scores, indicating that they introduce a significant element of uncertainty to model inference. This is possibly associated with the global nature of contrast adjustments, where a certain contrast range is forced on an entire image which

	No aug.	Min.*	Min.+contr.	Min.+noise	All aug.
All images	0.388 ± 0.018	0.415 ± 0.010	0.418 ± 0.014	0.431 ± 0.011	0.409 ± 0.013
Only cirrus	0.687 ± 0.031	0.748 ± 0.021	0.739 ± 0.026	0.763 ± 0.021	0.751 ± 0.028

Table 6.2: Comparison of different augmentation strategies. Results reported as mean segmentation IoU over 5 splits. Minimal augmentation uses rotations and flips, as described in steps 2 and 3 of Section 6.3.2. Contr. represents contrast augmentations. *Control model.

is not exhibited by any samples in the test set.

We assess the effect of pretraining models on synthesised images, prior to training on MATLAS images. We are interested in whether the synthesised images produced in Section 3.4.2 facilitate knowledge transfer that is valuable for inference on real images. We simply train two versions of the control model on each dataset, one with and one without pretraining. Table 6.3 details the segmentation scores for this experiment. We see that pretraining provides a significant benefit to segmenting unseen samples on both datasets, indicating that training on synthesised images provides features beneficial to processing real MATLAS images. Additionally, there is a large decrease in the margin of error for pretrained models, suggesting that pretraining encourages more robust features to be learned.

	No pretraining*	Pretraining
All images	0.415 ± 0.010	0.442 ± 0.003
Cirrus only	0.748 ± 0.021	0.788 ± 0.009

Table 6.3: Performance of control model with and without the use of pretraining on synthesised images. Results reported as mean segmentation IoU over 5 splits. *Control model.

In this experiment, we compare different variants of loss functions discussed in this work. We are first interested in how our proposed ‘super-majority’ loss framework compares to both a standard loss with no conditions based on confidence, and a similar loss framework [137] which we denote as RCF loss. Briefly, the RCF loss also utilises a conditional loss function, but with only three conditions: the first considers a pixel positive if it is greater than 0.5; the second considers pixel negative if it is zero (i.e. strictly no annotators classified it as positive); and the third considers all other pixels as uncertain and sets their loss to zero so to exclude them from the loss calculation. We are secondly interested in the effect of using focal loss in place of binary cross entropy. In order to

make both of these comparisons, we construct variants of the three loss frameworks with binary cross entropy and focal loss, resulting in six total loss functions.

		Plain	RCF	Super
BCE	All images	0.415 ± 0.006	0.441 ± 0.011	0.449 ± 0.005
	Cirrus only	0.748 ± 0.021	0.782 ± 0.017	0.792 ± 0.015
Focal	All images	0.432 ± 0.008	0.448 ± 0.010	0.456 ± 0.005
	Cirrus only	0.774 ± 0.022	0.790 ± 0.018	0.801 ± 0.014

Table 6.4: Comparison of BCE vs Focal loss functions with different consensus loss frameworks. Results reported as mean segmentation IoU over 5 splits. *Control model.

Comparative results between different loss functions are shown in 6.4. On our dataset of probabilistic annotations of LSB images, there are two clear patterns regarding consensus loss frameworks: ignoring pixels with uncertain annotation is helpful, as shown by RCF and supermajority loss both improving test accuracy; and prioritising very certain pixels is helpful, with the proposed supermajority loss outperforming the other loss frameworks in all scenarios. These patterns are also repeated in the ROC curves of models trained using different loss functions, as shown in 6.6, where the supermajority loss trains models with better precision vs. recall trade-off. Focal loss also appears to boost scores over BCE loss, perhaps owing to its focus on encouraging better handling of the severe class imbalance present in our dataset. This is supported by distributions of predicted positive pixels per image versus the target distribution, as shown in 6.5, where we see that focal loss achieves a further separation between the two modes than BCE loss.

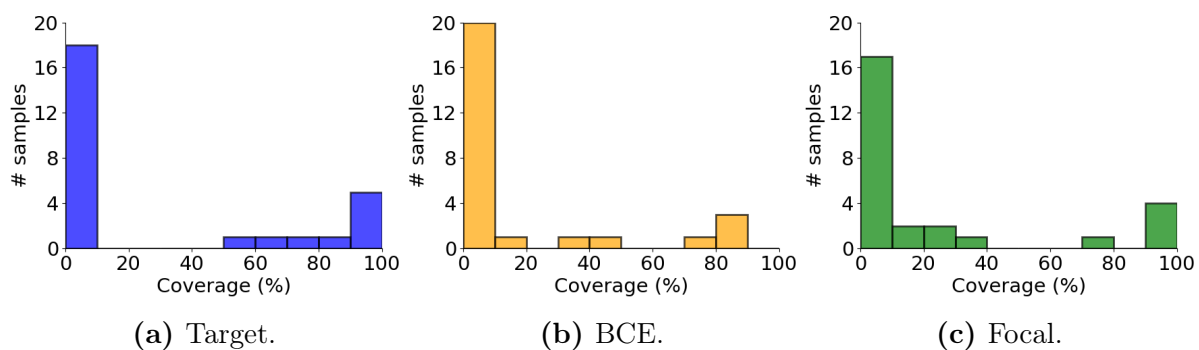


Figure 6.5: Histograms showing the proportion of actual or predicted cirrus across all testing LSB images. Predictions are taken from models trained with BCE and focal loss.

To conclude this section on comparative studies related to training protocol, we perform an ablation study where the best performing components are swapped in and out to create different overall strategies. This ablation study, shown in Table 6.5, allows us to assess the effect of different components in combination. The best performing strategies use

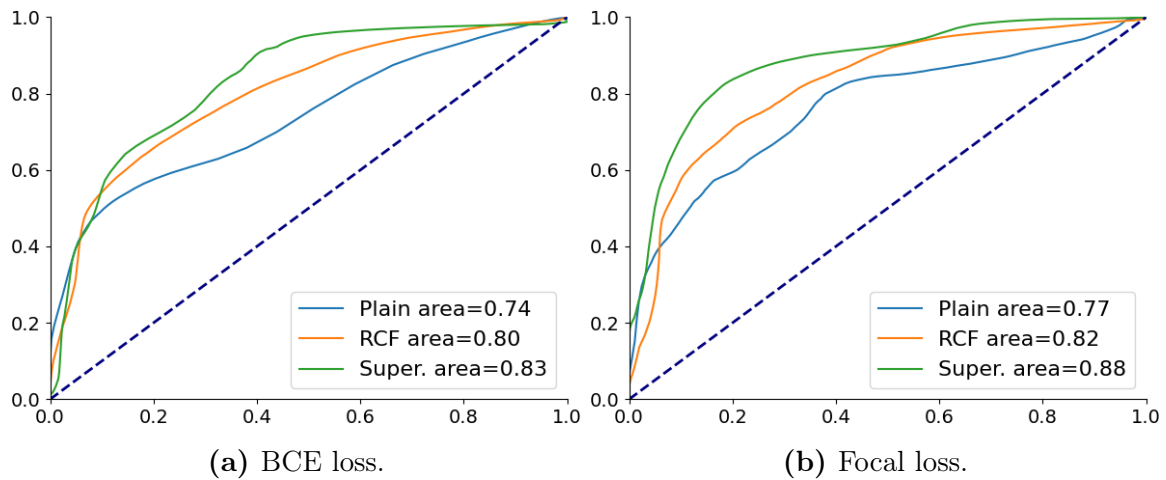


Figure 6.6: ROC curves for different consensus loss frameworks with BCE and focal loss functions, on only LSB images containing cirrus.

the proposed intensity scaling, pretraining on synthesised images and the proposed consensus loss. We see that, in general, there is a positive synergy between training strategies, i.e. almost all combinations outperform experiments that isolated components. Interestingly, an exception to this pattern is the use of augmentation with Gaussian noise, which has an unstable effect on model performance. Particularly, any positive effect of augmentation with Gaussian noise appears to be nullified when used in conjunction with adaptive intensity scaling, suggesting that these two components do not have a positive synergy. This may be due to the adaptive scaling layer providing a form of augmentation that fulfills a similar form of sample variation to Gaussian noise augmentation, but is more representative of LSB images. This nullifying effect is not strictly negative however, and there is a tie for the best performing strategy between using Gaussian noise and not using it (though there is a near-negligible difference in error margins). Given this non-negative pattern, and that this augmentation should in principle expose the model to ‘new’ samples which could exist in an LSB image dataset, we choose to use it in our final training strategy.

6.4.2 Ablation Study on Model Modifications

In this second half of our comparative analyses, we shift our focus to the model components detailed in Chapter 5. An ablation study is performed involving each of the independent components. This process allows us for each modification to firmly verify that the performance gains achieved in the previous chapter carry over to the dataset used in this chapter. We seek to discover whether the modifications are still helpful for cirrus segmentation on a more realistic problem scenario, and how exactly they alter the

Scaling	Pretraining	Min. & noise aug.	Consensus loss	All images	Cirrus only
				0.415 ± 0.006	0.748 ± 0.021
✓				0.455 ± 0.012	0.803 ± 0.019
	✓			0.442 ± 0.006	0.788 ± 0.009
		✓		0.431 ± 0.011	0.763 ± 0.021
			✓	0.456 ± 0.005	0.801 ± 0.014
✓	✓			0.466 ± 0.007	0.816 ± 0.011
✓	✓	✓		0.469 ± 0.005	0.814 ± 0.013
	✓	✓		0.463 ± 0.008	0.811 ± 0.015
✓		✓		0.451 ± 0.014	0.804 ± 0.022
✓	✓		✓	0.483 ± 0.009	0.830 ± 0.014
✓	✓	✓	✓	0.483 ± 0.006	0.832 ± 0.013

Table 6.5: Ablation study of best performing training strategies. Results reported as mean segmentation IoU over 5 splits. *Control model.

model predictions. We perform all ablation studies on both datasets described in Section 6.3.1. Throughout these experiments we control for training strategy and model architecture. We select the best performing components of the previous section to form our control training strategy: parallel intensity scaling, geometric & element-wise Gaussian noise augmentations, pretraining on synthesised images, and the use of our proposed supermajority focal loss. We also use the same base dual attention model as in the previous section (and as in Chapter 5) with a simple feature generating backbone.

We evaluate the use of different components related to the attention network on real cirrus samples. We modify the base control model to test four separate methods, detailed in the previous chapter, individually and in combination: how multiscale features are generated by the backbone; guided attention; gridded attention; and tri-attention. For these experiments, we fix batch size and overall parameter size of each network to be roughly equal, but use the maximum possible image resolution. In this way, we are able to take into account runtime memory efficiency of gridded attention vs other models. While it is possible to perform a ‘sliding window’ or ‘patch based’ segmentation and avoid excessive image downsampling, where a model is fed sections of an image and segmentations are then re-attached to obtain the final prediction, we found this to be massively detrimental in initial experiments (see Fig. 6.7), supporting the argument that global context is highly relevant for cirrus segmentation.

Results for this ablation study are shown in Table 6.6. Firstly, the use of guided attention improves performance on both datasets, and there is a more significant increase on the dataset with only cirrus images. This is likely due to the regularisation encouraging generalisation, as the performance improvement is larger as sample size decreases. We

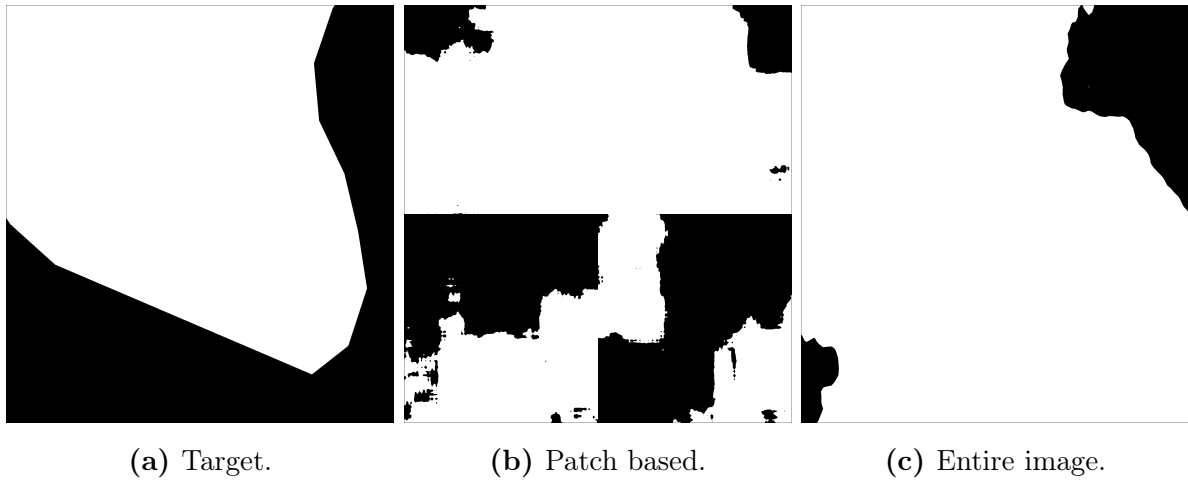


Figure 6.7: Comparison of segmentations generated with a patch based method versus a model that segments the entire image in one pass.

secondly observe that using multiscale features from intermediate layers of the backbone appears to decrease performance. Finally, results show that gridded attention and tri-attention both increase accuracy on all cirrus segmentation scenarios. Each of these findings concur with the results of the previous chapter; we refer the reader to Section 5.3.1 for more indepth experimental discussion of these model modifications. We also note that non-attention methods still suffer on the larger LSB dataset, indicating that limited training data is not necessarily the major bottleneck in these models. To conclude, empirical findings in this section demonstrate that the performance increases achieved by the methodology detailed in the previous chapter do carry over to the more difficult dataset used in these experiments.

6.5 Automated Cirrus Detection on LSB Images

We construct a final model from the best performing components of the previous section and analyse the predictions generated. We significantly increase the parameter size of our segmentation model in this section, and further evaluate the proposed method.

6.5.1 Experiment setup

We use the experiment setup from the previous section with some minor modifications. We first use a larger model with a more sophisticated feature generating backbone. This is achieved by swapping the simple backbone network out for a ResNet-50 [90] network, where we use the final layer before the global pooling operation as features. Given that ResNet vastly outperforms simple CNNs on a variety of classification tasks, we expect that

Inter.	Guided	Gridded	Tri	All images	Cirrus only
				0.483 ± 0.0060	0.822 ± 0.0132
✓				0.476 ± 0.0074	0.815 ± 0.0186
	✓			0.493 ± 0.0057	0.858 ± 0.0114
		✓		0.532 ± 0.0031	0.869 ± 0.0176
			✓	0.497 ± 0.0065	0.861 ± 0.0152
	✓	✓		0.542 ± 0.0029	0.886 ± 0.0162
✓	✓	✓		0.535 ± 0.0023	0.870 ± 0.0158
		✓	✓	0.536 ± 0.0031	0.869 ± 0.0166
	✓	✓	✓	0.548 ± 0.0028	0.892 ± 0.0130
✓	✓	✓	✓	0.543 ± 0.0026	0.885 ± 0.0154
U-Net [172]				0.381 ± 0.1286	0.685 ± 0.0963
LGCN				0.414 ± 0.0492	0.741 ± 0.0327

Table 6.6: Comparison of attention model modifications: generating multiscale features from intermediate layers; guided attention; use of the proposed gridded attention map; and computing Gabor attention in addition to dual attention. Additional networks are included for comparison. Results reported as mean segmentation IoU over 5 splits. First row represents the control model.

this change should provide a more robust and diverse feature set to compute attention across and result in a more accurate cirrus segmentation prediction. To account for this larger network, we increase training epochs to 400. We secondly increase the size of both the training and testing sets by dividing the validation samples between them. As the performance of all components has already been extensively validated, we opt to not to train over multiple splits for cross-validation. We also note that this practice is standard in all works exploring segmentation on astronomical images that have been cited in this chapter.

We experiment with different methods to extract the most accurate segmentation from the methodology. We use test time augmentation, where multiple augmented version of a sample are created, predictions are generated for each augmented version, and the inverse augmentations applied to the predictions to obtain multiple predictions per input sample. These predictions are then averaged, resulting in a more robust segmentation maps. Specifically we apply all permutations of 90° rotations and flips resulting in eight predictions which are averaged over. We also use ensemble predictions, where multiple models are trained, and predictions from each model are averaged over to give the final segmentation. Finally, the combination of both techniques is used.

6.5.2 Results

We report both IoU and Dice scores for model predictions in this section. Dice is more biased to precision rather than recall, so the use of both metrics provides a view of the predictive characteristics of our trained models. Figure 6.8 shows the training curves for a training run of the prediction model. We observe that there is a significant difference between training and testing scores, which is expected given the limited sample size of our dataset. Regardless of this, testing performance does trend upwards along with training performance, indicating that the network is not overfitting. We also see that Dice testing scores seem to increase at a higher rate and for longer than IoU scores, showing that as the model is trained further it predicts less false positives than false negatives relative to the true positive predictions.

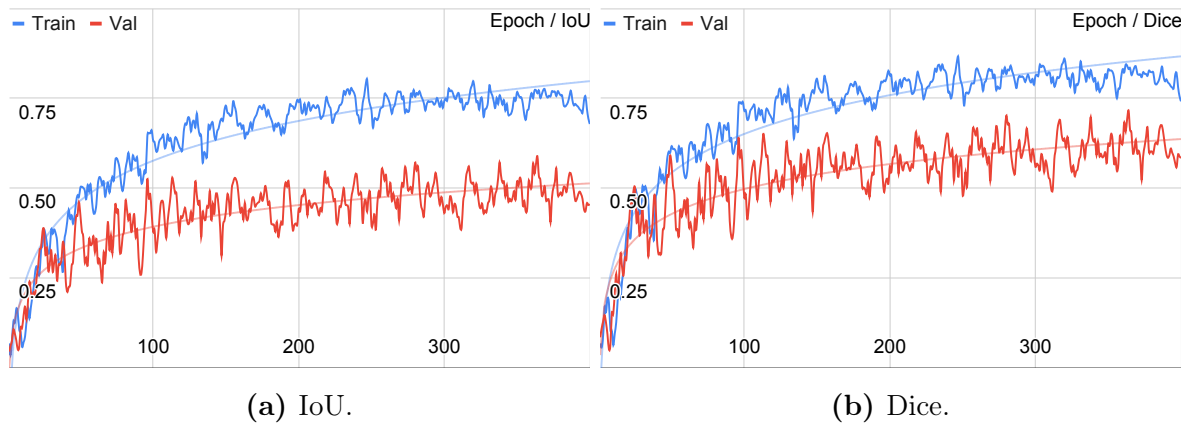


Figure 6.8: Training curves (smoothed) for the proposed model showing how IoU and Dice scores change over training epochs on the training and testing sets. We also fit a logarithmic curve to each plot to help illustrate the convergence trend.

Final segmentation scores on the testing set are shown in Table 6.7. We observe that there is a very large performance increase from models in the comparative analysis, owing to the larger parameter space, additional training samples and longer training period. We see that a single prediction model is outperformed by all other techniques of combining predictions, with the model ensemble achieving the highest scores. While test time augmentation is beneficial on a single model, we find that it decreases performance on the

	Single	Test aug.	Ensemble	Combined
IoU	0.745	0.773	0.790	0.781
Dice	0.766	0.794	0.814	0.806

Table 6.7: Results for the final network with different prediction generation techniques.

ensemble. Methods that aggregate predictions also appear to handle the class imbalance issue better than the single model, when comparing the target class distribution (Fig. 6.9) against predicted class distributions (Fig. 6.10). Kullback-Leibler divergences between the target distribution and each predicted distribution support this, with values of 0.40, 0.19, 0.07 and 0.29 for the single, test time augmentation, ensemble and combined models, respectively. We see that the ensemble model best matches the target distribution in addition to achieving the highest segmentation scores. Interestingly, prediction aggregation appears to also increase the gap between IoU and Dice scores, indicating that precision is increased through averaging over predictions.

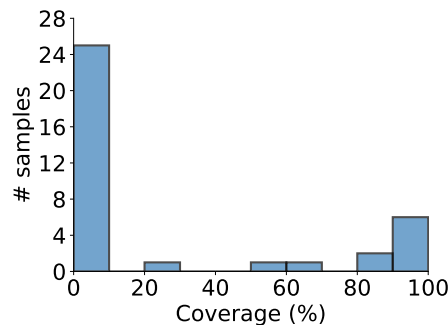


Figure 6.9: Histogram showing the proportion of cirrus coverage across testing images.

To better understand the characteristics of the trained segmentation models, we turn to a brief qualitative assessment of generated predictions. The model handles typical regions in LSB images well, as shown in Figure 6.11, confusing no areas as false positives. This low false positive rate also extends to more difficult samples, as shown in Figure 6.12, where even images with high background levels (NGC6017, UGC03960) or large areas of diffuse light (NGC5846) are predicted correctly. It can also be seen that the ensemble model does appear to increase accuracy in some uncertain scenarios, demonstrating the strength of averaging over multiple predictions. As shown in Figure 6.13, the model deals well with regions entirely contaminated by cirrus dust, with few errors. The inner section of cirrus clouds appears to be reliably predicted, though the model struggles with matching the envelope of contaminated regions. Figure 6.14 shows the 4 samples with the lowest IoU scores: while the model performs well on most images, there are some examples with poor accuracy. In particular the model appears to struggle with more localised areas of contamination, especially when contamination is close to the boundary where there is a small amount of global context.

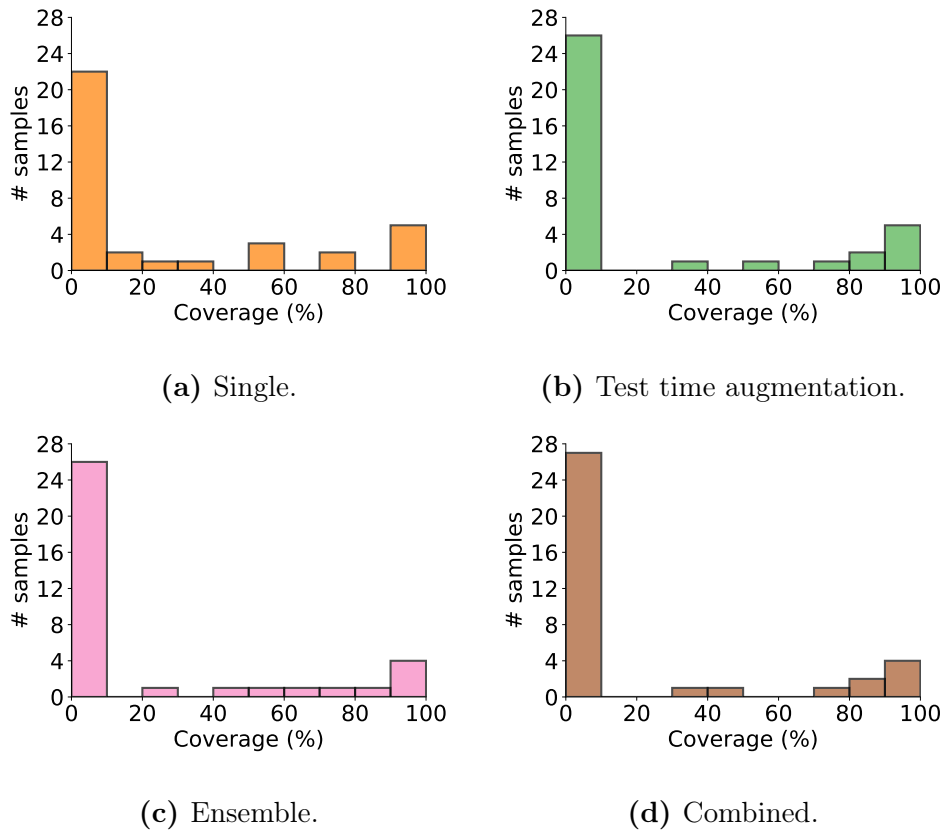


Figure 6.10: Histograms showing the proportion of predicted cirrus across all testing LSB images, for different prediction techniques.

6.6 Discussion

A major strength of the proposed method is that segmentation is performed on images with no preprocessing involving manual tuning of hyperparameters. Given that the scaling transformation is learnable, it should be possible to obtain good performance on LSB images produced with a variety of instruments. Using the presented model to identify cirrus contamination in other astronomical surveys would require a small amount of fine-tuning in order to recalibrate learned weights. In addition, the model can process an entire image in one pass with minimal downscaling, meaning that the proposed method can be easily integrated into the data processing pipeline for LSB instruments to obtain cirrus contamination masks. We were able to perform inference with an average time of approximately 0.4s per sample on a single GTX 1080 Ti. With optimisation efforts such as model compilation, pruning or half precision weights, this time could be significantly reduced making automated cirrus segmentation a relatively computationally inexpensive part of a larger LSB survey’s processing pipeline.

In this work we were able to train a relatively accurate model, however there were

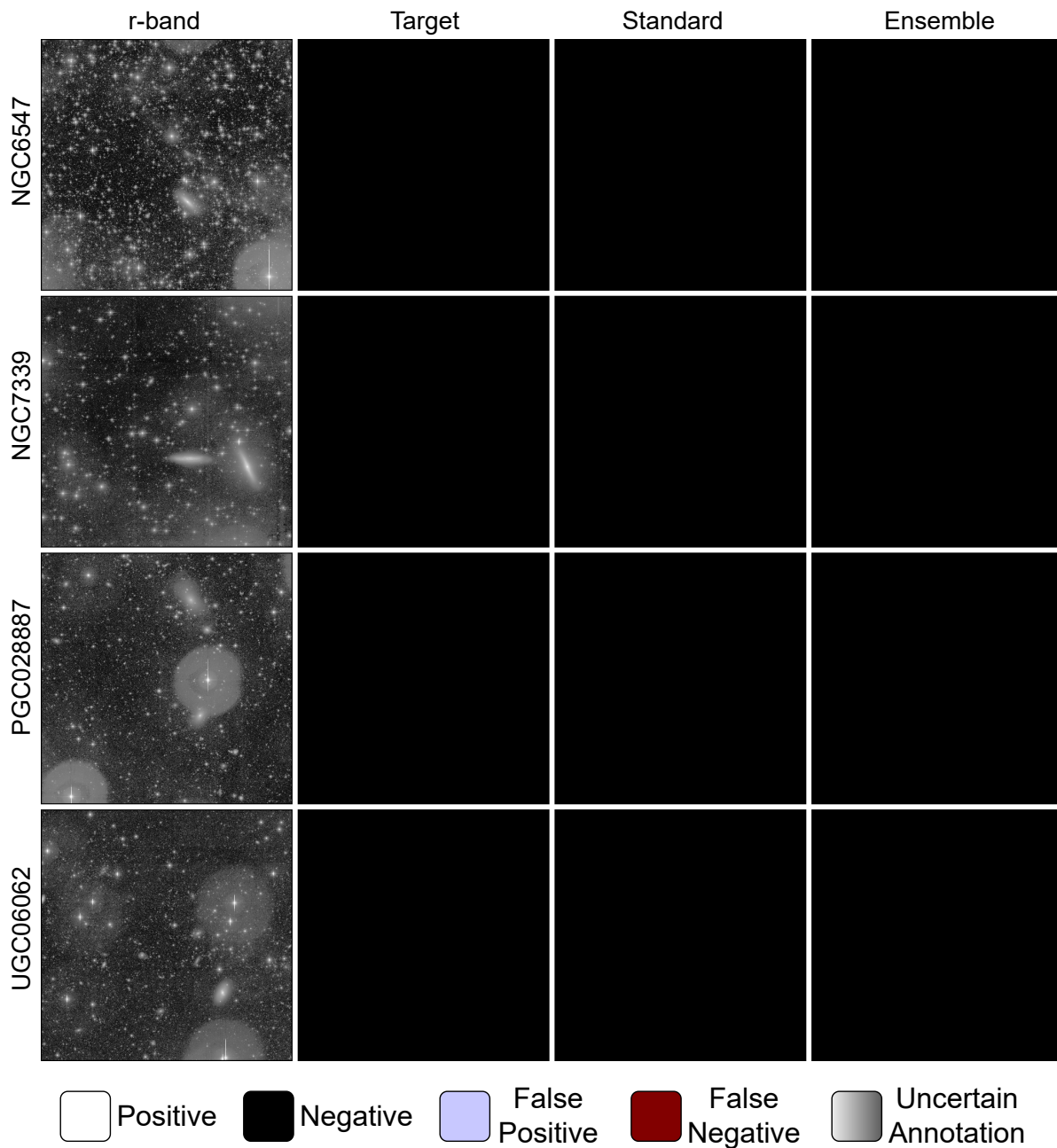


Figure 6.11: Typical LSB images with no cirrus contamination. The proposed model predicts zero false positives.

some limitations of the study. Imperfect dataset quality and small sample size were key challenges, and while the proposed techniques mitigated against these factors, there were some examples where the model failed to identify large regions of cirrus or predicted large amounts of cirrus which did not exist, as in Figure 6.14. It is likely that more training data would reduce these occurrences, though due to the infancy of high resolution and high sensitivity LSB imaging it is not currently possible to measure how well our method

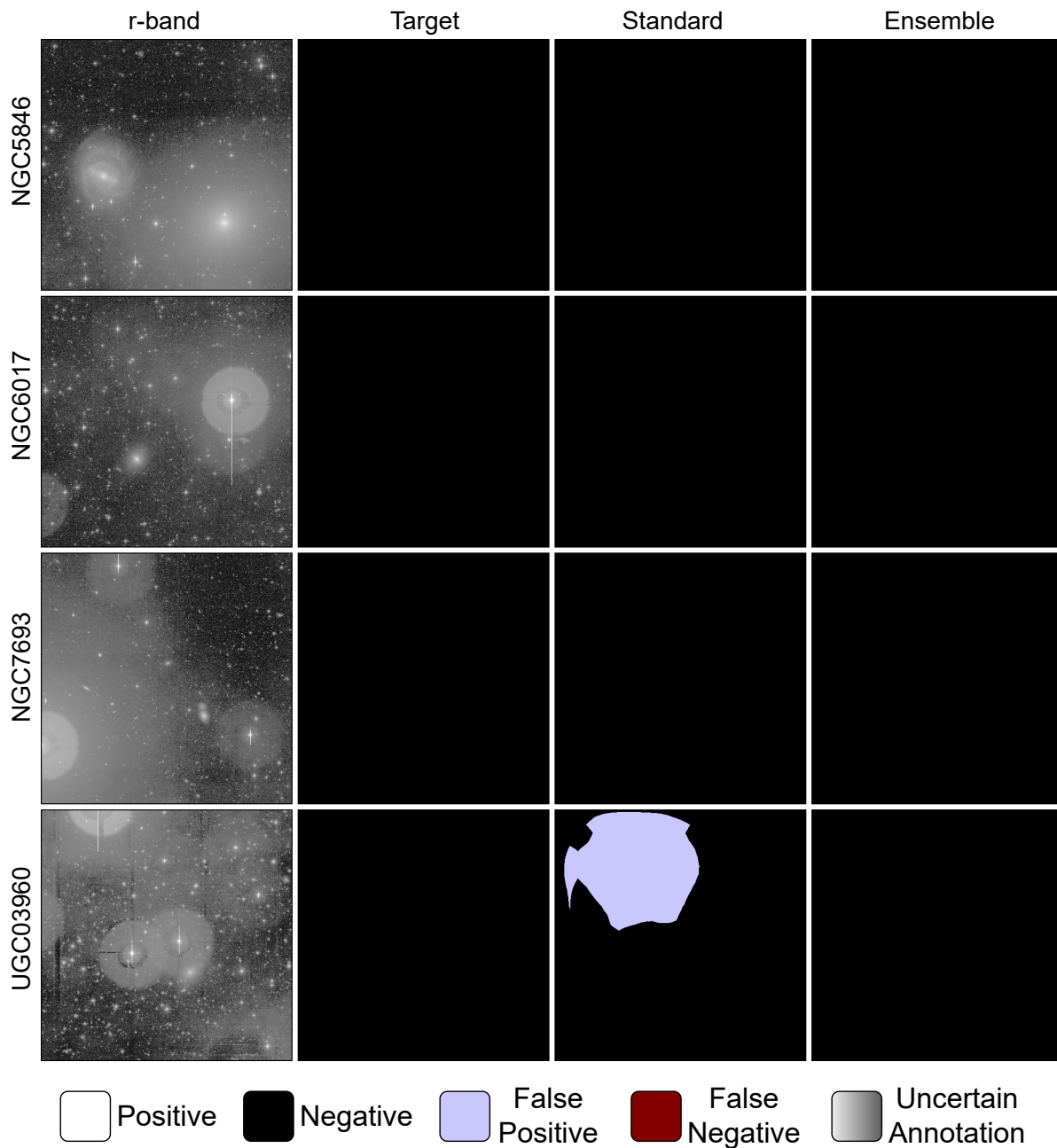


Figure 6.12: Segmentation predictions on difficult examples with no cirrus coverage. In this figure, we specifically chose examples with regions that present similarly to cirrus contamination, such as large regions of diffuse light in NGC5846 (first row), or large areas of high background levels in UGC03960 (fourth row).

would scale to a larger dataset of cirrus contamination. Furthermore, precise estimation of cirrus boundaries was not achieved by our model in most cases. We note, however, that this boundary is naturally ambiguous and there is often disagreement among annotators on the exact envelope (see Fig. 6.3b), so this uncertainty is naturally propagated during

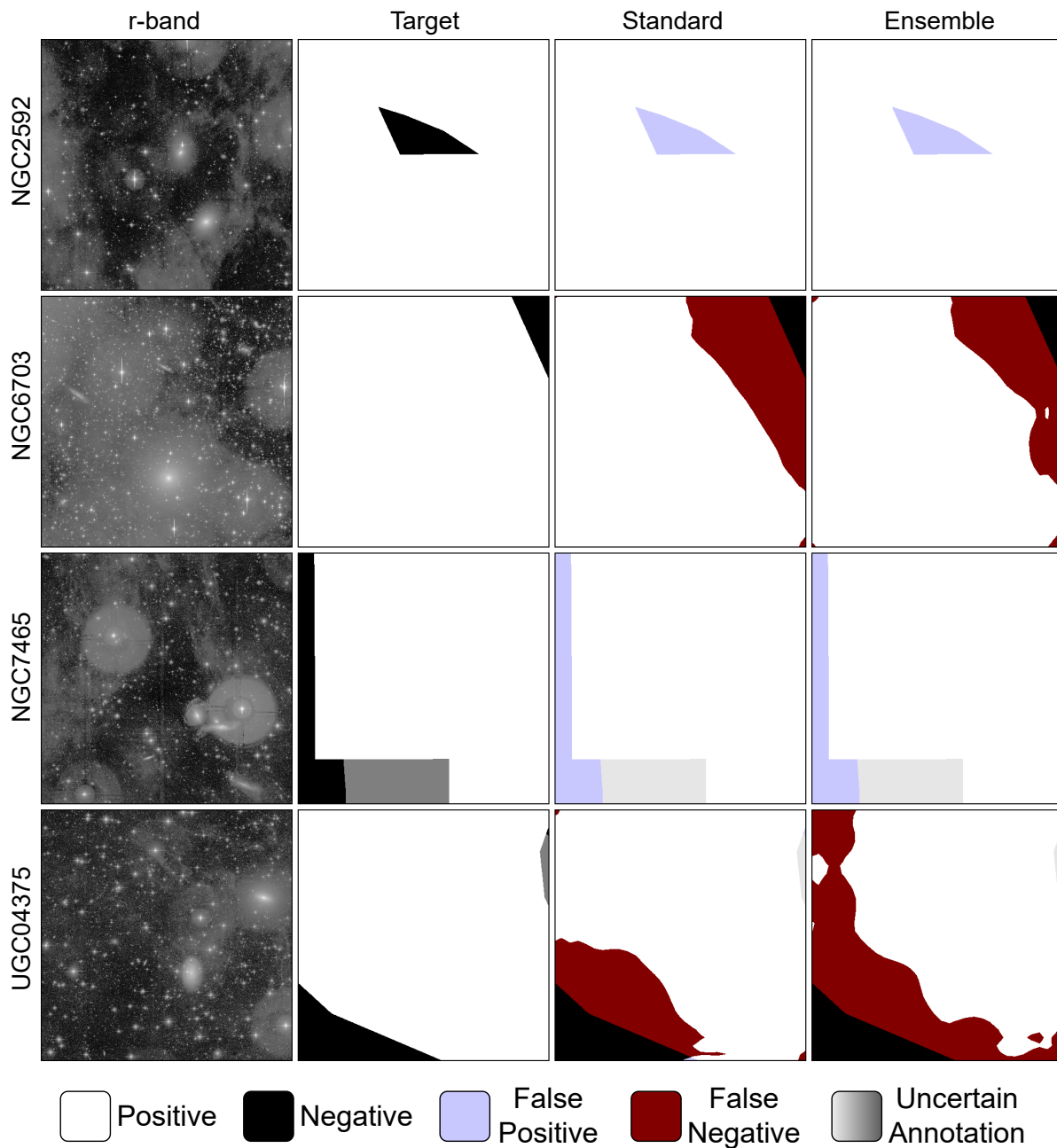


Figure 6.13: Segmentation predictions on examples with high cirrus coverage. Columns three and four show predictions generated by a single model and an ensemble of models, respectively. Light grey in the prediction map, as in the third row, indicates where the model predicted an uncertain pixel as positive.

training even with the consensus loss, which helps mitigate against this issue. Thus it is likely that with more annotators a consistent consensus on contamination boundaries would be reached.

Due to the nature of our annotated data we treated the problem as a binary segmen-

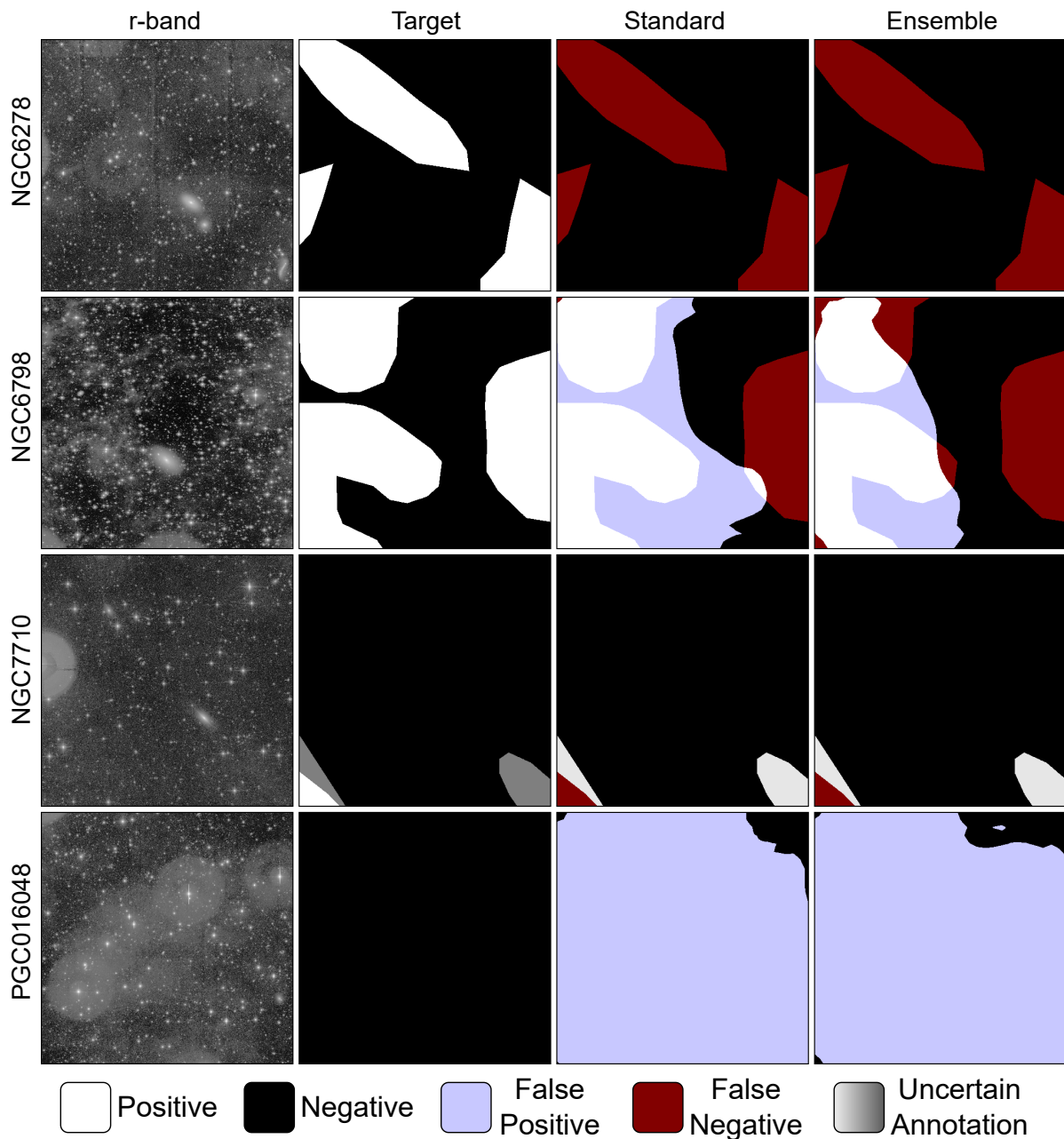


Figure 6.14: The four segmentation predictions with the lowest IoU scores across the testing set. Dark grey in the prediction map, such as in the third row, indicates where the model predicted an uncertain pixel as negative.

tation task. As is evident in various figures, the severity of cirrus contamination ranges from occluding bright objects to a slight change in the background level. While all ranges on this severity spectrum are important to identify, separation of cirrus into sub classes such as strong and weak contamination may be helpful from both a computer vision and astronomy perspective. For the former, this separation may implicitly guide the model

towards learning discriminating features exhibited by each severity range. Comparing performance across classes would also help identify what the model is lacking or where it could be improved. In the case of the latter, the distinction could facilitate studies on contaminated objects, where in the case of minor contamination, analysis could factor in the localised increase in background levels. Prediction masks of specifically stronger cirrus contamination could directly guide studies into the contamination itself.

6.7 Summary

In this chapter, we presented a comprehensive machine learning pipeline for automated segmentation of cirrus clouds in LSB images. We proposed a simple consensus loss to handle probabilistic annotations generated with a limited number of annotators. This loss function coarsely divides probabilities into groups and then prioritises labels where there is a strong consensus. We then designed an adaptive intensity scaling operation for enhancing subtle pixels in LSB images. This operation fits into the deep learning model as a standalone layer, where scaling parameters are learned alongside network weights. Finally, these contributions were combined with the gridded tri-attention architecture presented in Chapter 5.

Cirrus contamination is a significant hurdle for astronomers studying low surface brightness galaxies, and statistical analysis of galactic structures requires masking areas suffering from contaminating cirrus dust. In the near future, new surveys will generate large amounts of LSB data making manual cataloguing of cirrus infeasible. With the pipeline proposed in this work, we were able to segment cirrus contamination with reliable performance with only a small training dataset of LSB images. Future work will involve applying the pipeline to a larger dataset of LSB images sourced from multiple instruments. It would also be interesting, using the methodologies presented in this work, to craft a deep generative model capable of removing cirrus contamination from images.

To our knowledge, automated cataloguing of cirrus in LSB images with deep learning has not yet been attempted. Using our methodology, astronomy researchers will be able to automate the process of masking cirrus in future surveys, saving valuable time and greatly facilitating research into LSB galaxies and contaminated regions.

Chapter 7

Multi-class Segmentation of Galactic Structures

We widen the scope of the automated cataloguing investigated thus far and explore detection and segmentation of galactic structures in LSB images. We construct a Mask R-CNN [91] architecture with the methodology detailed in the previous chapter as a feature generating backbone. This network is then applied to an instance segmentation dataset created using the dataset described in Chapter 3. We find that the proposed methodology is able to reliably detect and segment objects. Furthermore, the trained network is able to generate predictions directly from MATLAS images, i.e. images require no preprocessing.

7.1 Introduction

Classification of galactic structures is a priority for LSB images. Given that high resolution LSB images are a relatively recent advancement, many structures have been uncovered that have not been previously catalogued or processed. For instance, tidal features, which are remnants of interactions between galaxies, can be studied in LSB images. The presence of tidal structures and the morphology they exhibit are clear indicators of galaxy formation history. Stellar halos are also uncovered by LSB imaging, which are an accretion of diffuse stellar material, such as dissolved tidal features. The processing of structures is a necessity for statistical analysis which has the potential to lead to important findings on the nature of LSB galaxies and thus general phenomena relating to galaxy evolution and formation.

It is key that reliable classification of LSB structures can be automated. Currently, classification is possible through manual means, where domain experts visually inspect images and record information on the structures present. This is the case as the quantity of LSB data is limited, with such manual cataloguing efforts [21, 62] being performed

on datasets with sample sizes in the hundreds. Future surveys, such as Euclid, seek to produce datasets of similar images with sample sizes far larger than can be feasibly manually classified. Development of a method to automatically classify LSB structures is crucial for the sphere of LSB galaxy research.

We investigate the use of machine learning to detect and segment galactic structures while simultaneously segmenting cirrus contamination in LSB images. Deep learning models have been applied to astronomical images for instance segmentation in multiple instances. Burke et al. [31] use a Mask-RCNN [91] model to detect and segment galaxies and stars, using a training set of simulated images. Farias et al. [69] also utilise a Mask-RCNN for instance segmentation, but focus on characterising galaxy morphology. The strong performance offered by this architecture in combination with existing successful use cases in astronomy provides a strong justification for its use.

An important distinction in segmentation problems is the idea of ‘stuff’ versus ‘things’ in classes of objects. The former refers to classes that present as uncountable amorphous regions, such as sky, grass or road. The latter, refers to distinct objects that present as countable entities, such as humans, cars or bikes. Semantic segmentation typically deals with ‘stuff’ whereas instance segmentation handles ‘things’. In this chapter, we seek to segment both ‘stuff’, in the form of cirrus, and ‘things’, in the form of galaxies and fine galactic structures. Such a unified problem is referred to as panoptic segmentation.

We first evaluate a standard Mask-RCNN on instance segmentation of galactic structures, and experiment with a simple semi-supervised learning loop where unannotated correct ‘false positives’ are manually added into the training dataset. With lessons learned from previous chapters, we then modify a Mask-RCNN model for panoptic segmentation and attempt to detect and segment galactic structures while simultaneously segmenting cirrus contamination in MATLAS images. We seek to train these networks to make predictions directly from images with no preprocessing, making the method versatile.

7.2 Related Work

Instance segmentation with deep neural networks has become a rich area of research in the past five years. Popular large benchmark datasets such as MS-COCO [129] and Cityscapes [48] have enabled standardised quantitative evaluation and facilitated research progress. The most widely used class of models for instance segmentation use a multi-stage approach, where objects are first detected and then segmented. The object detection component is typically carried out by a Region based CNN style network [82, 83, 168]. Here, regions of interest (RoIs) are identified with a convolutional subnetwork, pooled to reduce the number of incorrect or overlapping RoIs, and then classified. Mask R-CNN

[91] extended the R-CNN family of networks for instance segmentation, adding a mask generator and improving backbone feature generation and object proposals. Mask R-CNN has become a widely used baseline for instance segmentation due to its strong performance and simple design. Numerous modifications to Mask R-CNN have been proposed to improve performance. Liu et al. [136] prioritise the propagation of low-level features in the RPN, based on the principle that earlier layers focus on entire objects whereas later layers focus on local texture [217]. Cai and Vasconcelos [32] improve on object detection by using multiple detection subnetworks each trained on proposals pooled at varying levels of thresholds. Huang et al. [97] improve alignment of predicted masks by regressing the IoU metric given the final proposal features and the generated mask.

There have been several applications of object detection and instance segmentation in astronomy works. González et al. [84] employ a YOLO [166] detection network to detect galaxies and classify morphology. Burke et al. [31] train a Mask R-CNN on simulated images to classify and segment stars versus galaxies and achieve good performance even on overlapping objects. Farias et al. [69] segment galaxies and classify their morphologies in SDSS [23] images using a Mask R-CNN. Levy et al. [125] employ a Mask R-CNN on LSB images and attempt to identify LSB galaxies, though the model suffers from very poor precision due to cirrus contamination being confused with LSB structures. Mask R-CNN has also been used to detect/segment LSB artefacts such as ghosted halos and scattered light [190], though boundary delineation, fine classification, and poor precision on cleaner images are weaknesses of the pipeline. Outside of instance segmentation, reliable classification of galaxies [189] and even tidal structures [200] in LSB images has been achieved, though cirrus contamination appears to be weak in both studies.

Until recently, research into image understanding through segmentation has been split into two categories, semantic segmentation and instance segmentation. Kirillov et al. [113] propose a novel segmentation task that unifies the two categories, named panoptic segmentation. In this task, the model is required to simultaneously segment background classes (referred to as *stuff*) and foreground objects (referred to as *things*). A simple baseline for panoptic segmentation is proposed in [112], where a popular semantic segmentation model, FCN [138] is added into Mask R-CNN so that both instance and semantic networks share a common feature generating backbone. A problem with this approach for most datasets arises when combining the semantic and instance masks: as there is no encoding of consistency between the two masks, overlapping pixels become a problem especially with objects which are partially occluded. While several approaches have been developed [120, 127, 149] to combat this issue, we note that occlusion or overlapping classes are not an issue for data used in this chapter. This is the case as a pixel can be classified as more than one class: for example, a pixel can be contaminated by cirrus and belong to a galaxy.

Unification of segmentation of galactic structures and contaminants has not yet been performed with deep learning. Given that: 1) approaches identifying LSB galaxies suffer from high false positive rates due to cirrus contamination [125]; and 2) approaches identifying contaminants in LSB images suffer from high false positive rates where galactic structures are confused with contaminants in cleaner images [190], there is a clear motivation to combine the two tasks into a unified approach. There is also a plain motivation for the use of Mask R-CNN in this study, as all astronomy works investigating instance segmentation use the architecture and thus our results will be more easily comparable.

7.3 Method

In this section, we detail the methodology of this Chapter. The Mask R-CNN architecture is described, beginning with a brief overview of the model’s design followed by a discussion of more specific implementation details. The process of extending Mask R-CNN is then discussed, where the attention segmentation model of the previous chapter is integrated into Mask R-CNN to create a panoptic segmentation model.

7.3.1 Mask R-CNN Overview

At the core of Mask R-CNN is its feature generating backbone. Similarly to the attention network described in Chapter 5, a large network, such as ResNet [90], is utilised to extract features from the input, which are then used in the following components of the architecture. Numerous early implementations of Faster R-CNN [168] use backbone features from the final convolutional layer of ResNet-50 [50, 96]. Authors of Mask R-CNN propose using features from multiple layers of the backbone through a feature pyramid network [130] design, where intermediate layers produce features of differing scales. A second network is then added which progressively combines smaller scale features with larger scale features, exposing features of each scale to features of all other scales and providing a hierarchy of multi-scale features.

Backbone features are fed into the region proposal network (RPN) [135] to scan for possible RoIs, or areas of the image that have a potential to contain target objects, referred to as anchors. The RPN is a small network consisting of a single convolutional layer, which can be thought of as a sliding window which scans for the existence of an object within the window, followed by two branching parallel convolutional layers. These sibling layers have related tasks, one outputs the probability of the window containing an object, and the other outputs coordinates of the exact region of interest. Candidate regions are finally pruned through non-maximal suppression (NMS) where highly overlapping boxes

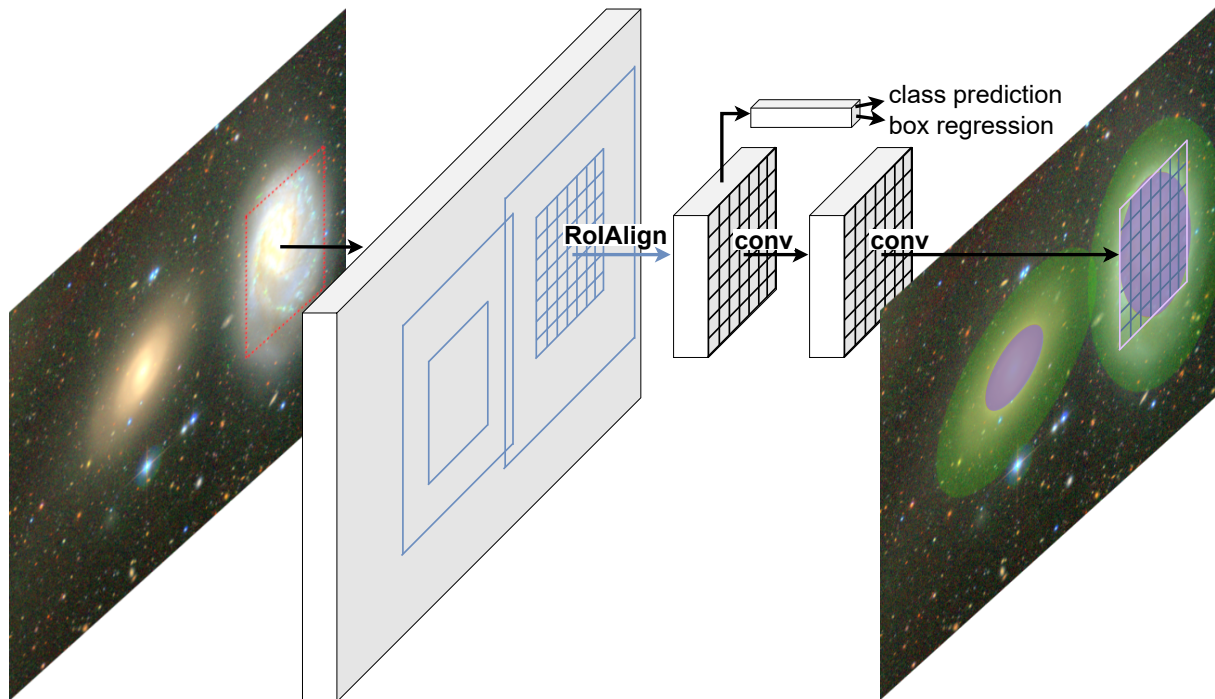


Figure 7.1: Mask R-CNN diagram, based on the first figure of [91].

are removed.

Following the RPN network, features are extracted from candidate regions by an RoI pooling subnetwork. This subnetwork consists of two components: a processing stage where backbone features inside candidate regions are downsized into a smaller feature map of standardised size; and a fully connected network for further feature refining. Mask R-CNN improves on the original RoI pooling network, RoIPool [82], by modifying the first component to downscale features through bilinear interpolation rather than max pooling, referred to as RoIAlign [91]. A subnetwork then predicts both the class of the potential object inside the candidate RoI (or lack of class) and regresses the coordinates of a box bounding the object.

The final step of Mask R-CNN for instance segmentation is mask prediction. An important design principle of the architecture is that mask and class prediction are decoupled. This is in contrast with typical multi-class segmentation networks which predict a class for each pixel through a multinomial cross-entropy loss. In Mask R-CNN, masks are predicted by feeding each output feature (candidate object) of RoIAlign into an FCN [138]. Each mask is a binary segmentation, i.e. whether or not the pixel belongs to the candidate object. The complete loss is then the sum of three tasks: class, bounding box, and mask prediction, i.e. $L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}}$. By combining the three tasks into a single loss function, there is a positive synergistic effect [91]. This is likely because the three tasks are significantly intertwined, for example, the segmentation mask of an object

is linked to the class of the object.

7.3.2 Implementation details

We implement Mask R-CNN with a ResNet-50 backbone as it is a proven choice and it was used in earlier chapters in this work. We use mostly default off-the-shelf parameters: RoIAlign size = 7×7 px, RPN non-maximal suppression threshold = 0.7, object confidence threshold = 0.05. While tuning of hyperparameters could lead to minor performance improvements, we verified that this set up gave consistently good results and thus we leave this computationally intensive optimisation for future work.

An important and relevant parameter to consider is the size of anchors considered by the RPN. The anchors must be a sufficient variety of sizes so that all possible object sizes and shapes are contained within an anchor size. It is also important that anchor sizes/shapes that are unlikely to bound an object well are not used to reduce noise during the training process. To determine reliable anchor sizes we compute histograms of the heights and widths of all target objects, shown in Figure 7.2. It can be seen that the majority of objects have heights and widths between 32px and 512px, and aspect ratios between 0.25 and 2. We therefore set the RPN to consider anchor boxes of widths 32, 64, 128, 256 and 512, and for each box width three aspect ratios are considered: 0.5, 1, and 2. While there are a significant number of object bounding boxes with aspect ratios between 0.25 and 0.5, using unbalanced ratios has the possibility of introducing a bias into the model where tall but narrow objects are favoured by the RPN. This is unwanted as pose variation in localised astronomical objects is naturally balanced (ghosted halos which are artefacts also have unit aspect ratio in most cases). This setup results in 15 total anchor box sizes considered.

Finally, an adaptive intensity scaling layer (see Section 6.3.3) is placed before the backbone network to exaggerate fainter structures within the LSB images.

7.3.3 Cirrus Subnetwork

In this work, we wish to segment cirrus contamination along with localised objects. Mask R-CNN is designed to handle the latter category; segmentation of extended amorphous regions such as cirrus contamination fits poorly into the instance segmentation framework which Mask R-CNN is formed upon. Handling categories of objects that cannot be divided into discrete entities is typically handled by networks of different design, such as FCN [138] or the attention network proposed in Chapter 5. Thus, there is a strong motivation to extend Mask R-CNN to separate the task of cirrus segmentation from instance segmentation.

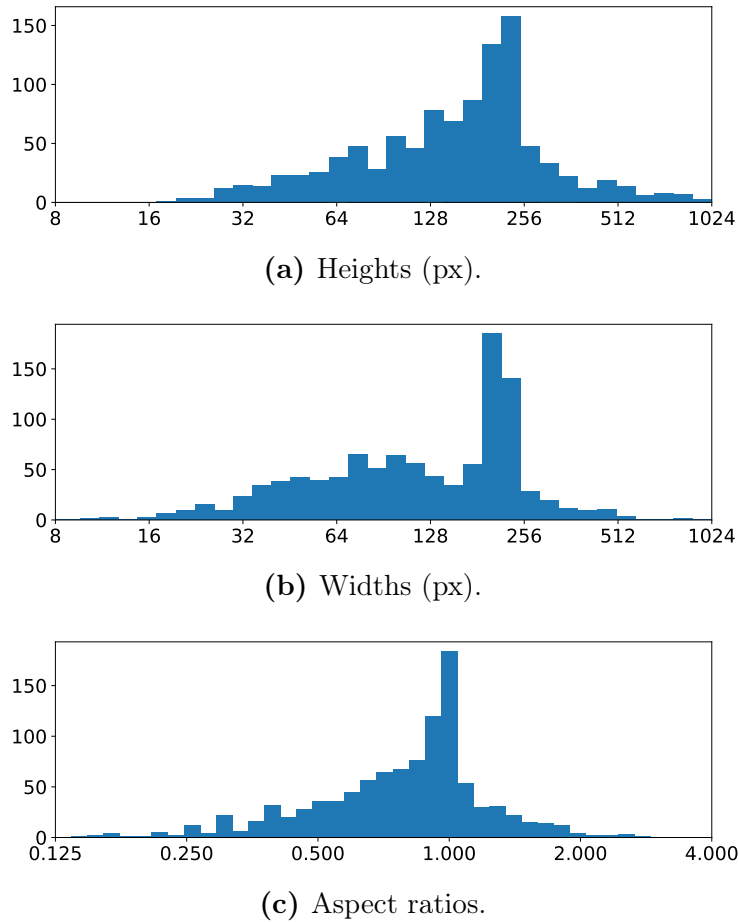


Figure 7.2: Distributions of heights, widths and aspect ratios of all target objects in the annotation dataset.

We combine the gridded Gabor attention network proposed in Chapter 5 with Mask R-CNN, as shown in Figure 7.3. We arrange the two networks so that they share the same backbone features, unifying segmentation of discrete objects and segmentation of galactic cirrus, i.e. the same ResNet-50 features are fed to both RPN/RoIAlign and the gridded attention module. Computation along each branch is then performed in parallel resulting in two segmentations which can be combined to achieve a segmentation of all structures in an input LSB image.

7.4 Data

In this section, we first describe the training set of LSB images and strategies used to improve generalisation of trained models to unseen samples. The preparation of annotations is then detailed to create labels that are compatible with training an instance

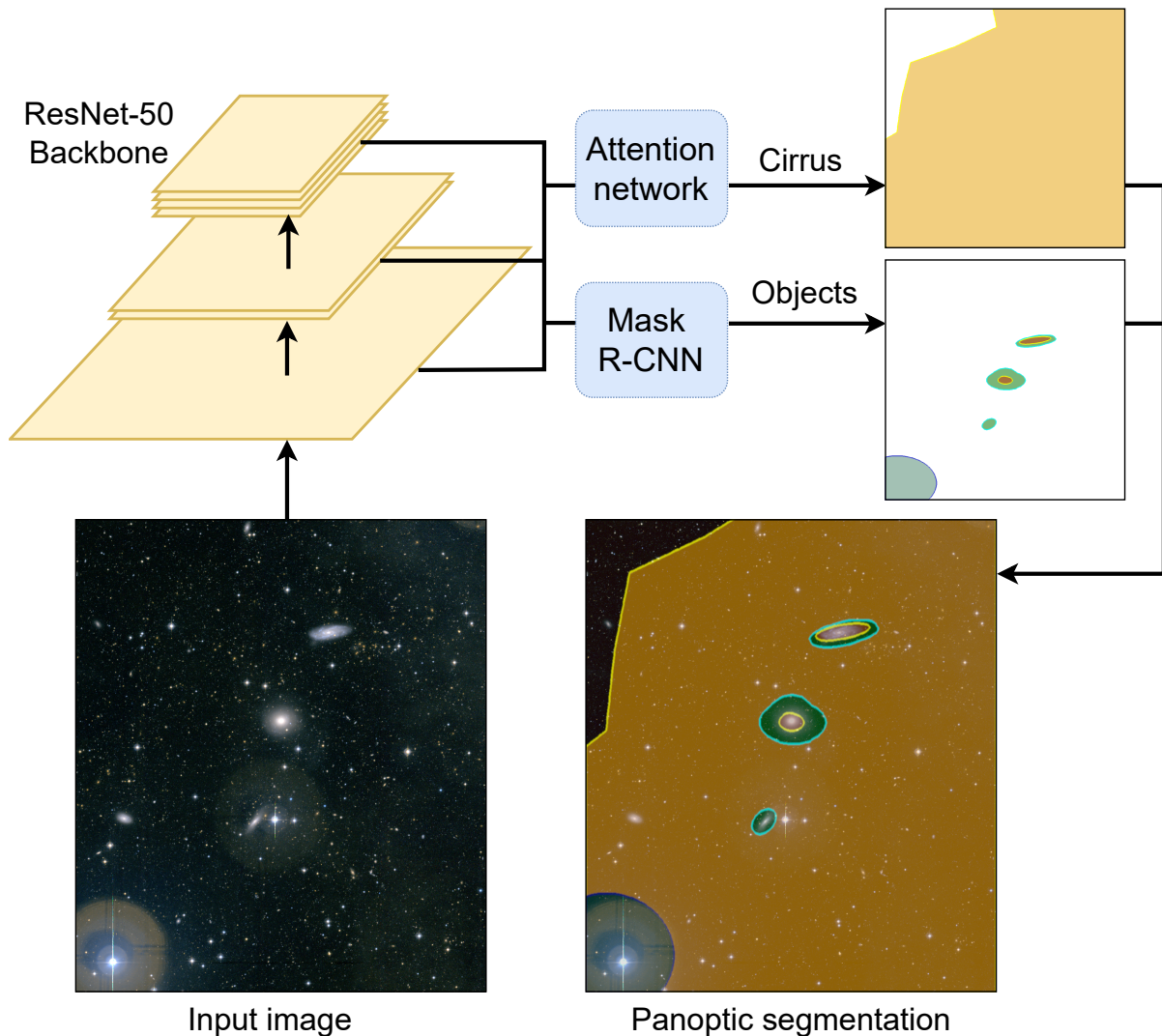


Figure 7.3: Diagram of the proposed segmentation model, combining a gridded Gabor attention model with Mask R-CNN.

segmentation model.

7.4.1 Dataset

In this study, we use the same image set as in Chapter 6. This contains 186 MATLAS LSB images with two spectral channels of average spatial size 6000×6000 px. The same training/testing split as Section 6.5 is used where respectively 80% and 20% of samples are used for training and testing. We take a 3000×3000 px crop around the target galaxy of each image. We employ the same data augmentations of Section 6.5: images are downsized to 1024×1024 px, a combination of random flips and 90° rotations are applied, then element-wise Gaussian noise is applied. Given that these augmentations resulted in

a significant test set improvement on cirrus segmentation models, generalisation should be improved on a modified task with the same images.

To further improve generalisation of the trained model and mitigate against overfitting due to the limited sample size, transfer learning is utilised. Prior to training, Mask R-CNN is loaded with weights trained on the MS-COCO dataset, provided by Torchvision’s [143] default pretrained option. MS-COCO provides a training set of over 300000 natural images with instance segmentation labels of 91 classes of objects. While natural images obviously appear very different from astronomical images, learned features can still be closely applicable on both sets. We also note that this practice is standard in works combining astronomy and instance segmentation [31, 69, 190]. Following this process, the network is further pretrained on the same synthesised cirrus dataset used for pretraining in the previous chapter in order to adjust network weights to features more closely resembled in astronomical images.

From the annotation dataset detailed in Chapter 3, target labels are narrowed down to five classes. First, we combine galaxy classes into one class, i.e. main galaxies and companions, as these definitions rely on the idea of how annotations are performed on suspected galaxies of interest. We task the model with identifying all instances in an image, thus this notion is not relevant to the machine learning model. Secondly, we combine elongated tidal features, tidal tails, plumes and streams, into a single class. These three features appear visually similar, and are all defined as a propulsion of stellar material from a galaxy. Furthermore, given that the distinction between tails and streams is the type of source galaxy, and we group galaxy classes into one class, combining these tidal features naturally follows. Thirdly, we do not predict shells as the shape assigned to their annotation does not enclose a space that can be segmented. Finally, we discard contaminant classes other than ghosted halos and cirrus, as they are either very rare (satellite trails/instrument artefacts) or very difficult to correctly predict without a larger field of view. The final object classes used in this study are: galaxy, diffuse halo, elongated tidal structure, ghosted halo and cirrus.

7.4.2 Obtaining instance masks

Formulating the delineation and detection of individual galactic structures into an instance segmentation framework requires processing of annotation data. Thus far, expert annotations have been combined into consensus masks for each semantic class, allowing the training of semantic segmentation style models which through a softmax operation predict a class label per pixel. Instance segmentation models, on the other hand, require a format that encodes masks of each distinct object as well as their semantic class. There

is thus the task of identifying distinct objects in consensus masks to create such instance labels.

For the majority of object classes: galaxies, companion galaxies, diffuse halos, and ghosted halos, a single shape is used for delineation. In these cases, a simple connected component analysis is sufficient to separate most instances. For this process, consensus masks are binarised by rounding. We consider two positive pixels connected within a mask if there exists a path of positive pixels that starts at one and ends at the other. Parts of the mask that are not connected are then assumed to be separate objects. Tidal features, however, can be described by multiple shapes that are close together. For example, consider a tail that appears separated in its centre. To isolate these shapes/features into instances, we instead use proximity. We found that grouping tidal features that are less than 0.025° (roughly 200 pixels) apart produced visually reasonable results.

There exist some cases where the connected component method does not separate instances. Overlapping objects of the same class fall into this category, a scenario which occurs with diffuse and ghost halos. We choose a compromise for ghost halos and exploit the fact that ghost halos exhibit fixed size. Any ghost halo masks that significantly differ from this size are simply excluded from the dataset. This choice resulted in losing approximately 15% of ghost halo annotations, which is a reasonable trade-off to ensure consistency in class definition.

Diffuse halos require more special attention, as approximately 40% of annotations suffer from overlap. We proceed with a combination of visual inspection and geometry analysis, as shown in Figure 7.4. We first apply a Euclidean distance transform, where the shortest path connecting each positive pixel to a negative pixel is computed. The value of each positive pixel is then set to the length of this path, so that pixels at the centre of a shape have the highest values. Local maximum peaks are identified to find the locations of these centres. We choose two or three of the most separated peaks, based on the number of galaxies involved in the overlap (determined manually through visual inspection). These peaks are then fed into the watershed segmentation algorithm [198] as markers to identify the boundary separating each overlapping shape and separate the combined shape into parts. Finally, in cases where the retrieved part is incomplete, an ellipse is fit optimally to the part and used to interpolate the lost section of the shape.

7.5 Results

Results of the detailed methodology are presented in this section. We begin with a simplified task where only localised objects are segmented with Mask R-CNN to provide a baseline of performance on MATLAS LSB images. We then experiment with a simple

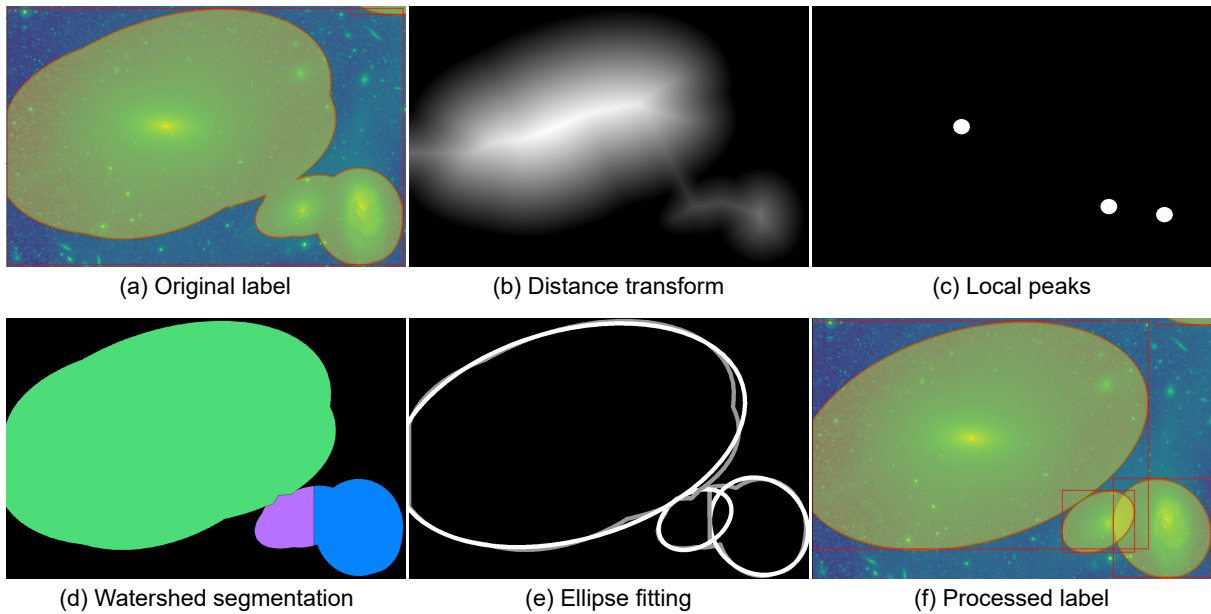


Figure 7.4: Instance labels for diffuse halos surrounding NGC4281, NGC4277 and NGC4273 before and after processing, with intermediate results.

human-in-the-loop training run and investigate how Mask R-CNN can be used on images with partial annotations. Following this, the panoptic segmentation model proposed in Section 7.3.3 is tested on simultaneous segmentation of localised objects and cirrus contamination.

7.5.1 Instance Segmentation with Mask R-CNN

To provide a performance baseline of instance segmentation on LSB images, Mask R-CNN is trained and evaluated on only classes that contain localised objects from the annotated MATLAS dataset. Cirrus contamination in MATLAS images presents a significant challenge when precisely delineating structures. To understand the relationship between the tasks of contamination segmentation and galactic structure segmentation, it is highly relevant to perform a study isolating both parts (see Section 6.5 for cirrus segmentation).

Metrics – For evaluation of network performance we calculate the precision and recall for each image in the test set. A detection is considered positive if its predicted mask has an overlap score (IoU) greater than or equal to a given threshold with a target label of the same class. From these precision and recall values, we compute the average precision (AP) score, which is defined as the area under the precision-recall curve averaged over all images in the test set:

$$\text{AP} = \frac{1}{|R|} \sum_{r \in R} p(r). \quad (7.1)$$

Here, $p(r)$ is the maximum precision at a given recall and $R = \{0, 0.01, \dots, 1.0\}$ contains each recall bin. This metric is computed multiple times with different IoU thresholds $\in \{0.5, 0.55, \dots, 0.95\}$ for deciding whether a detection is positive or negative. As the IoU threshold increases, an increasing number of imperfect detections are considered negative predictions and thus the AP should decrease. The AP at an IoU threshold x is denoted as AP_{100x} , e.g. AP_{50} for an IoU threshold of 0.5, and is calculated for each class individually and combined.

Baseline results – The detailed Mask R-CNN model is trained for 200 epochs using a simple stochastic gradient descent optimiser as in [168] with a learning rate of 0.01 which is halved every 25 epochs. Momentum and L2-regularisation penalty are set to 0.9 and 5×10^{-4} , respectively. Training is performed on a single Nvidia GTX 1080 Ti over approximately 12 hours.

Precision-recall curves are shown for each class in Figure 7.5. Segmentation of elongated tidal structures proves to be a very difficult task, with the network making no positive detections of such structures on the test set at any IoU threshold. The galaxy and diffuse halo classes share similar performance profiles, with strong AP scores that are strong at low IoU thresholds but quickly decrease over higher thresholds. This shared pattern is not surprising as galaxy and diffuse halo annotations are naturally intertwined, as the diffuse halo is the scattered light surrounding a galaxy captured by the LSB imaging instrument. While prediction of ghosted halos exhibits a weaker performance at the lowest overlap threshold, performance is sustained at a much higher rate than other classes. This phenomenon could be due to a combination of two factors: ghost halos have clear boundaries in most cases making localisation easier, and sizes of halos are fixed. Thus AP begins to drop at higher overlap thresholds due to halos that overlap with other bright diffuse light or cirrus contamination, making the halo’s boundary difficult or impossible to observe (e.g. Figure 7.6).

Through manual inspection of network predictions on unseen images, it can be seen that the network is able to predict reasonable boundaries for diffuse halos and galaxies, even in cases where there is significant cirrus contamination (see Figures 7.6 and 7.7). Given that traditional methods typically struggle in such scenarios, this is a significant strength of the method. In Figure 7.7, the network predicts an elongated tidal structure off the edge of the target galaxy, which is most likely cirrus contamination. However, this perhaps demonstrates that the network is somewhat familiar with characteristics of elongated tidal features, despite the null performance on this class, as the prediction is reasonable. Indeed, the predicted region exhibits the correct local texture and is located where a tidal feature could exist. Thus, it is reasonable that tidal feature prediction would be largely improved by more training data.

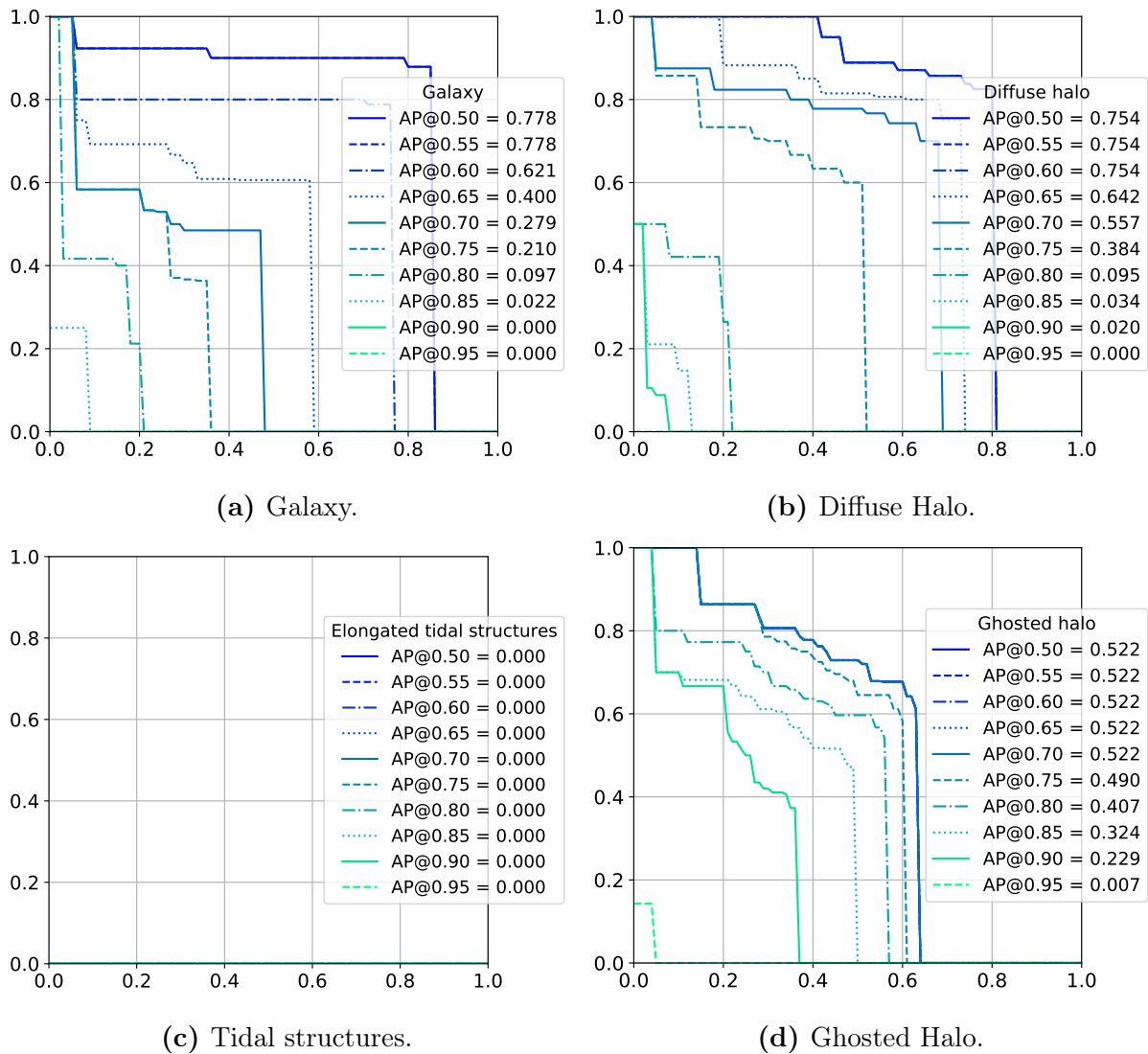


Figure 7.5: Precision-recall curves for each object class and associated AP scores over different IoU thresholds. Horizontal axes show recall, vertical axes show precision.

Human-in-the-loop training – We also observe that false positives are, in almost all cases, correct but unannotated objects. For example, in Figure 7.6, a diffuse halo, a galaxy and a ghost halo have been correctly predicted but are considered false positives as they were not annotated. Based on this fact, we extend the current experiment where we implement a simple human-in-the-loop (HITL) training protocol, illustrated in Figure 7.8. After training for an initial period, we review predictions made on all 184 images. Mask predictions which are of good quality are added into the dataset to be used for training and testing, as shown in Figure 7.11. The network is then trained on the new dataset for a shorter period and then predictions are reviewed. This process is repeated several times to construct a more densely annotated dataset. Specifically, we train for 30 epochs,

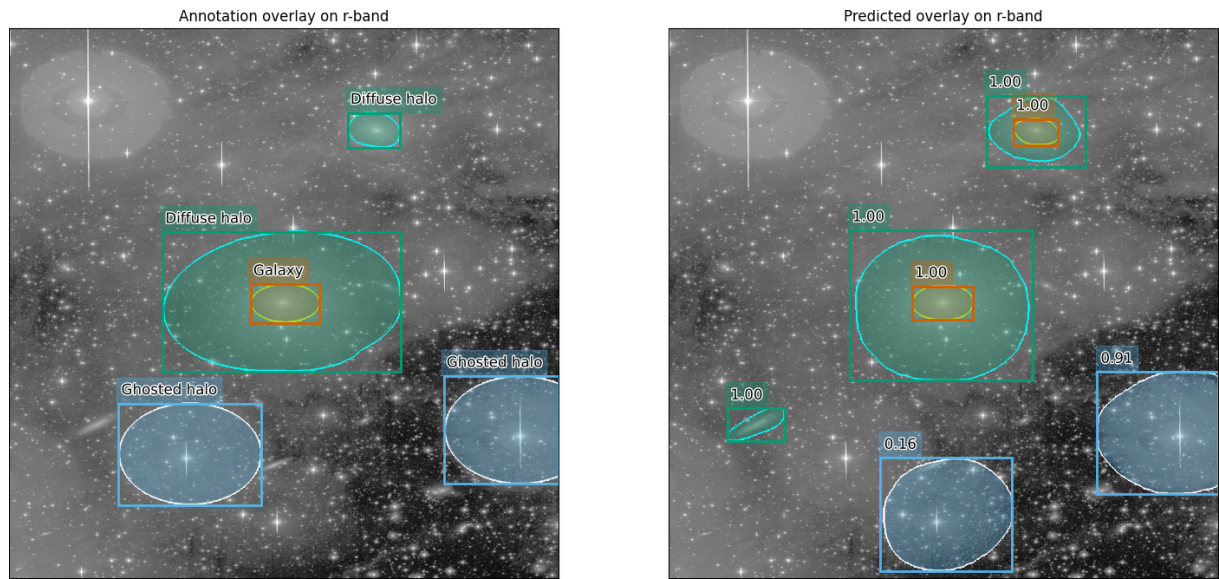


Figure 7.6: Annotated and predicted objects in NGC6703. Numbers above network predictions represent confidence scores.

alternate four times between reviewing and training for 5 epochs, and then alternate four further times between reviewing and training for 10 epochs. The model is then trained until a total of 200 epochs has been reached. Finally, we do not consider tidal structures during this experiment, as they were annotated exhaustively in all MATLAS images.

Figure 7.9 illustrates the number of objects added into the dataset or rejected at each review stage. Over the first four review stages, 67% of the total 472 false positive predictions are added into the dataset. Galaxies, in particular, are reliably predicted, with an acceptance rate of 89% over these review stages. At the fifth review stage there is a large jump in incorrect false positive detections, with 33, 347 and 162 bad detections predicted to be galaxies, diffuse halos and ghosted halos, respectively. This is likely due to the large amount of added objects over previous stages which the model is still adapting to, given that the problem does not repeat over following review stages where the number of epochs between stages is increased. The largest number of added objects in a single review occurs at the sixth review stage, where 250 false positive detections are added into the dataset at an acceptance rate of 67%. The number of accepted objects drops in the final two stages, likely as the annotation fields at this stage are saturated with no clear detections left to add.

Results summarising the performance of instance segmentation experiments are detailed in Table 7.1. On the standard annotation dataset, the model scores a lower AP_{50} when trained with HITL predictions than without. Given that the HITL trained model detects more objects and, assuming that more false positives are predicted, this indicates

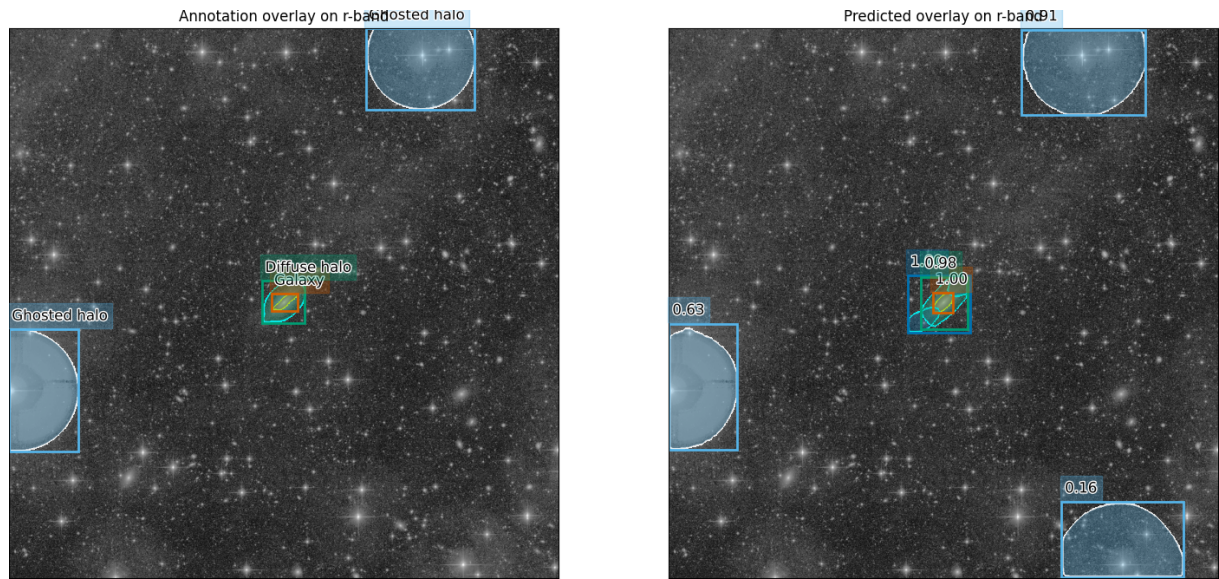


Figure 7.7: Annotated and predicted objects in NGC7710. Numbers above network predictions represent confidence scores. Dark blue belongs to the elongated tidal features class.

HITL trained	HITL eval.	Galaxy	Diffuse halo	Tidal structures	Ghosed halo	All
		0.778	0.754	0.000	0.522	0.514
✓		0.624	0.720	0.000	0.635	0.495
✓	✓	0.797	0.856	0.000	0.814	0.617

Table 7.1: AP_{50} and AP_{75} scores across different classes from models trained with and without HITL data, evaluated with and without HITL data.

that the number of false negatives is not significantly decreased on these classes. It is the contrary for the ghosed halo class, where the decreased number of false negatives outweighs the increase in false positives resulting in a higher AP score. This provides a strong motivation for considering the HITL ghosed halos as part of the ground truth, a scenario in which average precision actually increases by 56% due to HITL training. Across all classes, however, the average pattern appears to fall into the former of the two cases, as shown in Figures 7.10a and 7.10b, where the recall increase is outweighed by the precision decrease. Unsurprisingly, the HITL model performs better on HITL data than non-HITL data (see Figure 7.10c), as a positive bias is introduced where the model is evaluated on objects it is known to be capable of detecting. Nonetheless, this shows that all classes can be well represented by the model and thus much higher AP scores would be possible with a larger dataset.

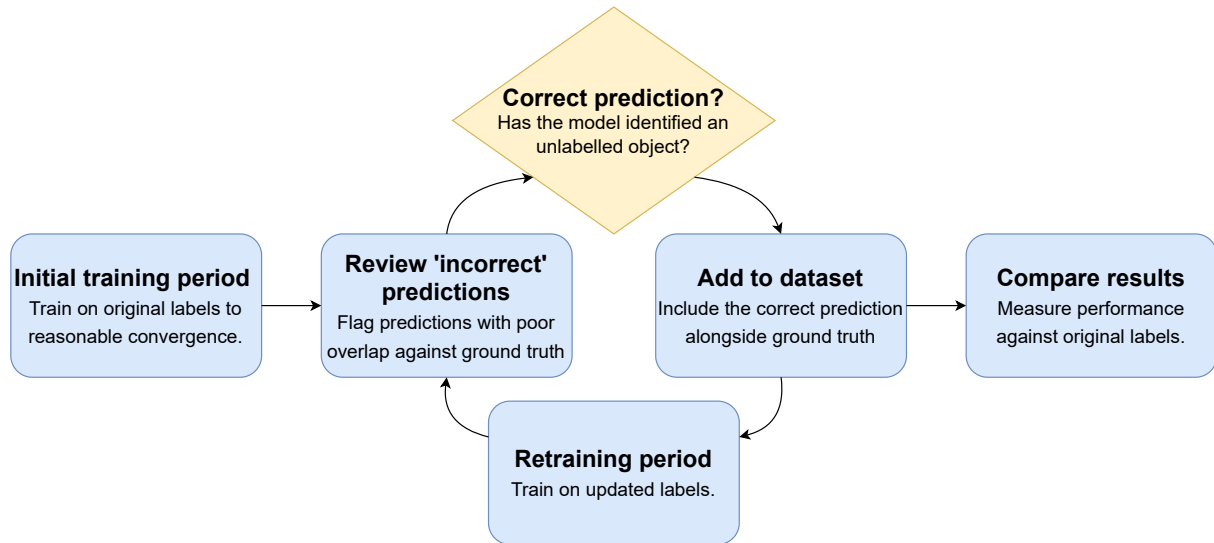


Figure 7.8: Flowchart outlining the implemented human-in-the-loop training protocol.

7.5.2 Panoptic Results

We now turn to the task of panoptic segmentation in MATLAS images. The proposed model is tasked with simultaneously segmenting galactic structures, localised contaminants and cirrus contamination. Specifically, the model must segment the four foreground classes of the previous section: galaxies, diffuse halos, tidal structures and ghosted halos, and the cirrus class. From this experiment, we seek to discover the impact of unifying these segmentation tasks on the model’s predictive characteristics.

The proposed model is trained for 200 epochs with different optimisation strategies for the instance and semantic portions of the network. For each branch, we attempt to match the training setup of the isolated versions. This ensures that a fair comparison can be made between results here and results of isolated tasks. The Mask R-CNN and backbone sections of the model are trained with SGD using a learning rate of 0.01 which is halved every 25 epochs, and L2-regularisation penalty of 5×10^{-4} . The attention network is trained with the Adam optimiser using a learning rate of 10^{-3} which is exponentially decayed by a factor of 0.98 per epoch, and L2-regularisation penalty of 5×10^{-7} . Training is performed on a single Nvidia GTX 1080 Ti over approximately 24 hours.

Precision-recall curves for localised structures predicted by the proposed panoptic segmentation model are shown in Figure 7.12, and summarised in Table 7.2. It can be seen that average precision scores for the all classes except elongated tidal features are increased in the panoptic model, across all IoU thresholds. For the galaxy class, the proposed model offers a minor improvement, with AP_{50} increasing by 0.5%. This difference is likely minimal as the galaxy core is a strong structure and can be delineated

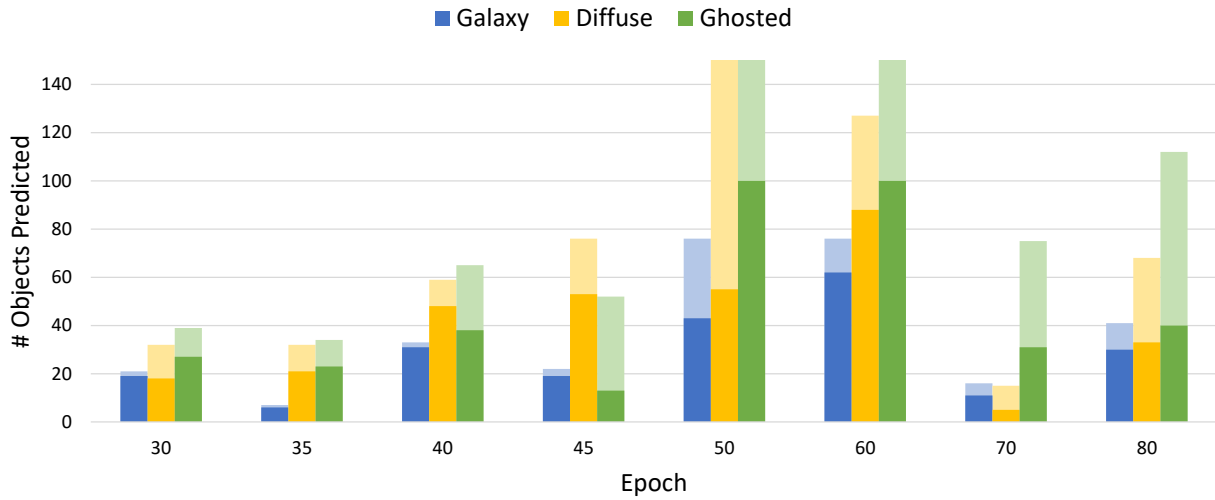


Figure 7.9: False positive objects predicted by Mask R-CNN at different epochs throughout the human-in-the-loop training run. Solid colours represent correct predictions that are added into the dataset, opaque colours are rejected predictions.

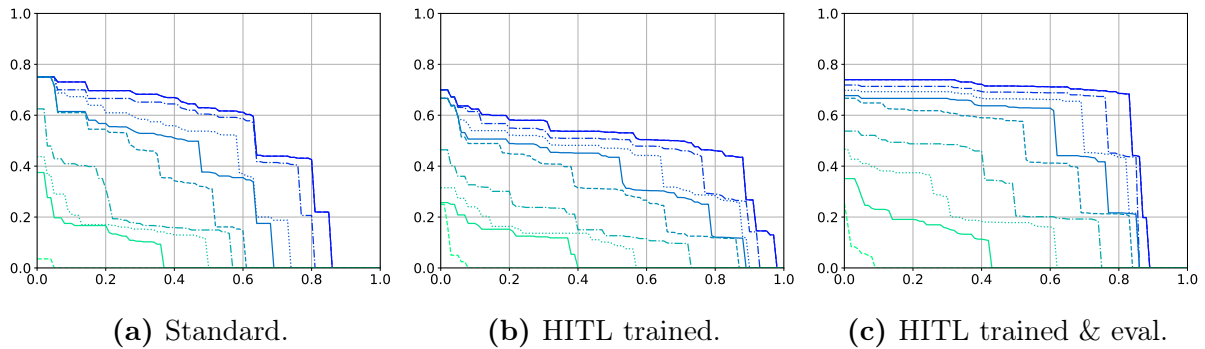


Figure 7.10: Precision-recall curves for all classes over different IoU thresholds. Horizontal axes show recall, vertical axes show precision, brighter colours denote higher IoU thresholds.

relatively easily even in highly contaminated areas. Thus the synergistic benefits of tying the tasks of contamination and object prediction are minimal for galaxy cores. Diffuse halos and ghosted halos obtain a more significant increase, with the proposed model increasing AP_{50} scores by 4.5% and 26.1%, respectively. This concurs with the previous insight, as boundaries of such structures are impacted by cirrus contamination, and thus the panoptic approach offers a larger synergistic benefit for diffuse halos and ghosted halos than for galaxy cores. Scores for these classes are also increased throughout all IoU thresholds on the precision-recall curve; AP_{75} scores are increased respectively by 3.3%, 7.0% and 34.3% for galaxies, diffuse halos and ghosted halos. This indicates that predicted boundaries overlap better in the panoptic approach, i.e. correct detections/classifications

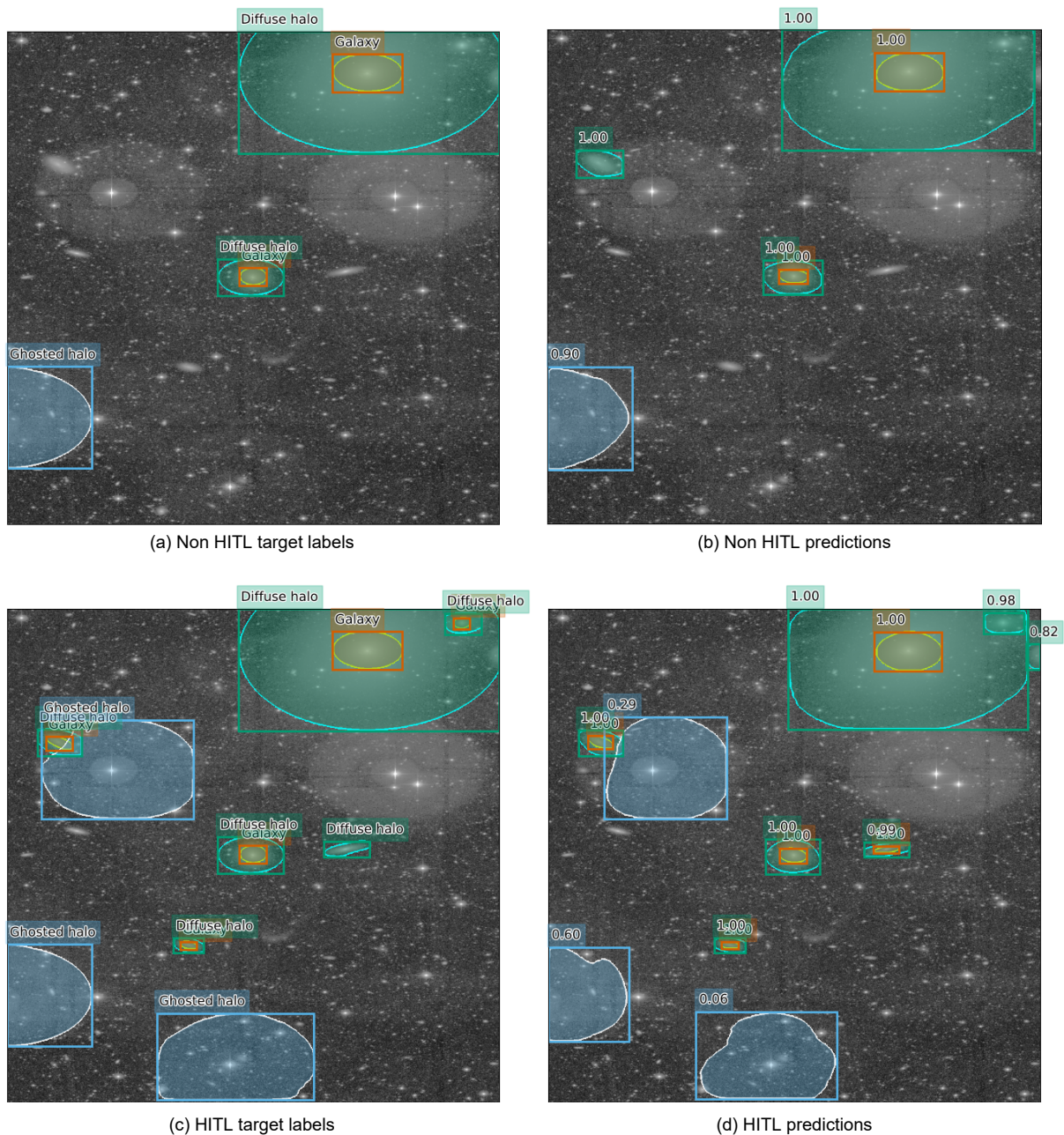


Figure 7.11: Comparison of target and predicted labels (PGC050395) on training runs with and without the human-in-the-loop protocol.

are of better quality.

In addition to improving on instance segmentation performance, the panoptic model also scores higher on the cirrus segmentation task. In Chapter 6, the attention model scored an IoU of 74.5% as a standalone predictor, and 79.0% as an ensemble predictor. The same attention network, as part of a panoptic model, scored an IoU of 85.5%, representing a relative increase of 14.9% and 8.2% over the standalone and ensemble contamination-

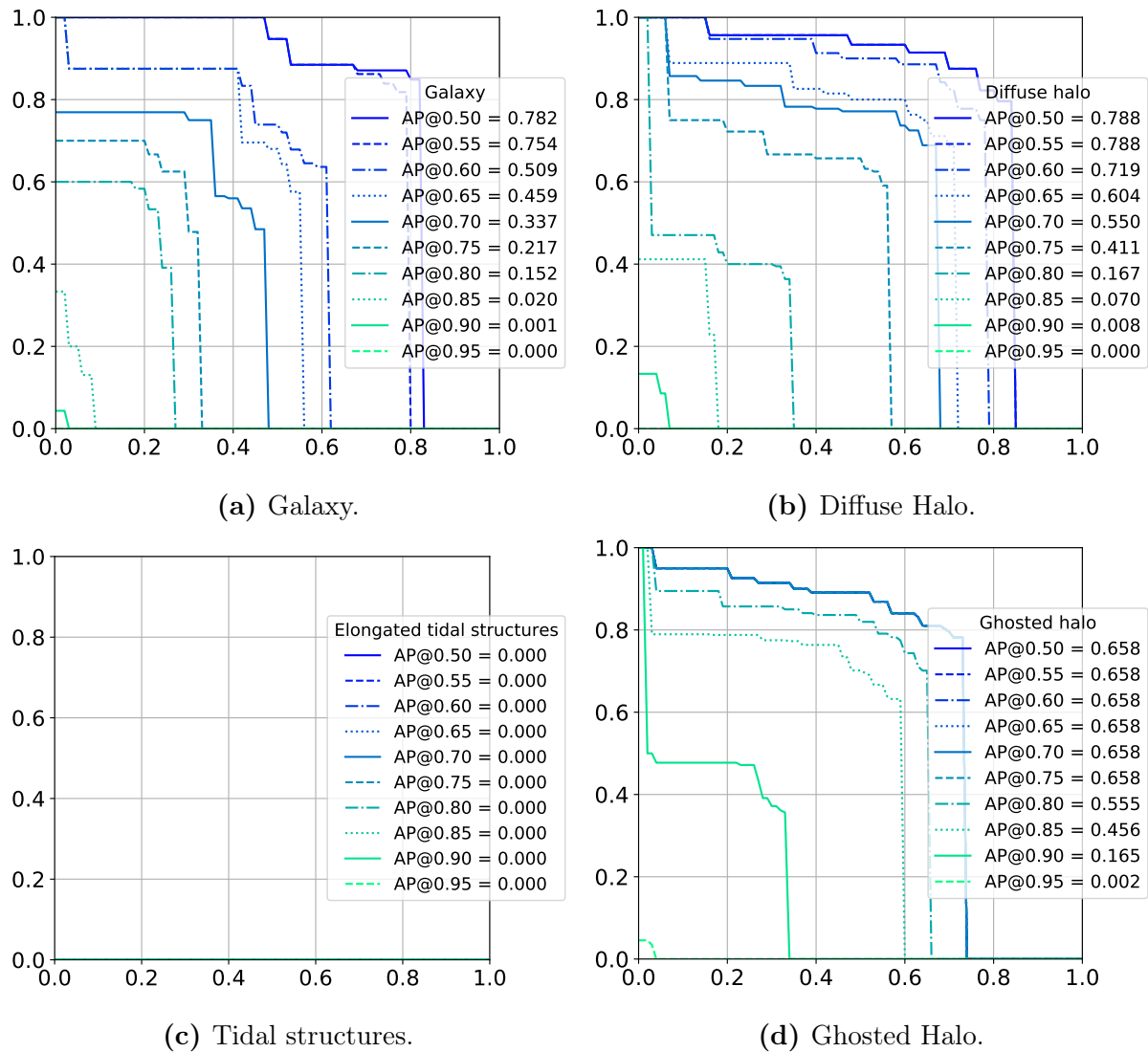


Figure 7.12: Precision-recall curves for each object class and associated AP scores over different IoU thresholds, with the proposed panoptic segmentation model. Horizontal axes show recall, vertical axes show precision.

only models. Based on this significant increase, it would seem that instance segmentation serves as a significantly beneficial auxiliary task for cirrus contamination segmentation. Given that the two tasks are combined through sharing a feature generating backbone, it follows that the addition of more semantic classes allows the model to generate features that better discriminate between cirrus and non-cirrus pixels. Interestingly, the distribution of predicted cirrus coverage appears to differ in the panoptic case versus the non-panoptic case, with the model showing a tendency to binarise segmentation predictions and either predict the image as completely contaminated or containing no cirrus. This may be due to the fact that only a standalone panoptic predictor was used, due

		Galaxy	Diffuse halo	Tidal structures	Ghosted halo	All
Panoptic	AP ₅₀	0.782	0.788	0.000	0.658	0.543
	AP ₇₅	0.217	0.411	0.000	0.658	0.330
Instance	AP ₅₀	0.778	0.754	0.000	0.522	0.514
	AP ₇₅	0.210	0.384	0.000	0.490	0.271

Table 7.2: AP₅₀ scores across different classes from models trained with and without HITL data, evaluated with and without HITL data.

to the complication of combining instance segmentation results. Indeed, the Kullback-Leibler divergence between the target and predicted distributions of cirrus coverage is 0.39, aligning closely with the 0.40 value of the standalone contamination-only model of Chapter 6.

Predictions of elongated tidal features remains a large challenge in this study, with the model failing to identify any of the structures in the test set. The limited number of training samples and difficult class imbalance is likely the culprit here, and a larger study would be required to discern whether or not the panoptic approach improves on the task of segmenting elongated tidal features. Given the benefits observed in performance on other classes, such an investigation is certainly warranted.

7.6 Summary

In this chapter, we presented a method for automated cataloguing of galactic structures in LSB images. Literature surrounding instance segmentation in astronomy was surveyed, and we identified that methods often suffer from a lack of unification of localised objects and homogeneous textures. Mask R-CNN was combined with the gridded Gabor attention network of Chapter 5 to create a panoptic segmentation model. We described the processing steps taken to prepare the MATLAS annotation data of Chapter 3 for training such models requiring separation of consensus labels into distinct entities. Mask R-CNN was first evaluated on an instance segmentation task involving galactic structures. Following this, with a simple human-in-the-loop training protocol, we added 914 unannotated objects into the dataset and significantly improved model accuracy on the ghosted halo class. This success warrants investigation into a more sophisticated active learning or semi supervised framework in future work, especially for the galaxy class which the model consistently predicted correctly. Whereas input images and dataset size remained constant across relabelling iterations, an active learning approach combined with more available LSB images could involve strategically requesting annotations on images con-

taining objects that the network is struggling with. Finally, the proposed method was used to simultaneously segment galactic structures and cirrus contamination, where we showed that unification of the tasks improves performance on both tasks. The proposed panoptic model achieved respective increases of 5.6% and 8.2% over the isolated instance and contaminant-only segmentation tasks.

Segmentation of galactic structures is feasible with deep learning, even in images suffering from heavy contamination. Galaxies and their surrounding diffuse halos, and ghosted halo contaminants were detected reliably, with reasonable delineation boundaries in areas of high cirrus. This success shows great potential for the future of automated cataloguing with deep learning, with inference performed in under a second on images with minimal preprocessing. Despite poor results on elongated tidal features, the nature of false positive detections of elongated tidal features suggests that the network has some understanding of their discriminating features. Thus, we hypothesise that with more training examples, automated detection and segmentation of these subtle structures should be possible.

Chapter 8

Conclusion

In this thesis, we investigated the use of deep neural networks for automated cataloguing of galactic structures and contaminants. The first obstacle of this investigation was that there did not exist training data suitable for segmentation of such structures. A further complicating factor associated with this fact was that a tool for generating 2D labels on large multi-spectral astronomical images did not exist. In Chapter 3, we sought to address both of these challenges and facilitate the training of modern neural network models for segmentation in an astronomy context. A fully featured annotation tool was presented, enabling astronomers to categorise and precisely draw shapes over galactic structures. The tool was designed to accommodate the requirements specific to astronomical images, such as being able to zoom and pan over surrounding regions, and tracking real world coordinates of user interaction. Collaboration was also made possible due to the tool being designed as a web application, where multiple users could contribute to the same annotation. Following this, dataset created using the annotation tool was detailed, comprised of 6573 drawn shapes on 227 MATLAS images. We justified and described how the annotations of the four users are combined into a single consensus labelling which can be as ground truth targets for training neural networks. Finally, a dataset of synthesised cirrus samples was detailed, including multiple variations of increasing difficulty. These synthesised samples are suitable for pretraining CNNs, enabling transfer learning, and for use in ablation studies to reinforce findings on real data.

Sensitivity to orientation is greatly desirable in processing astronomical images, as objects often exhibit orientational patterns, such as the filamentary structures within cirrus clouds. While CNNs are well equipped to handle variation in translation, rotational symmetries are poorly captured by the convolutional operator. In Chapter 4, we addressed this limitation through a novel convolutional layer involving Gabor filters, and created an architecture, LGCN, capable of generating features dependent on exact orientations without interpolation artefacts. We proposed an adaptive modulation method,

where convolutional filters are multiplied by Gabor filters whose analytical parameters are learned during backpropagation, alongside convolutional kernel weights. This learnable modulation was implemented in a fully complex-valued CNN, to enable the use of the full Gabor filter. The modulated convolution was then extended with the proposal of cyclic modulation, where each filter of different modulation orientations are exposed to all orientation dependent features. We demonstrated the effectiveness of LGCNs first on classification of randomly rotated hand-drawn digits, secondly on segmentation of cirrus structures in synthetic and real LSB images, and thirdly on boundary detection in natural images. We found that, in all problem scenarios, vanilla CNNs and static real-valued Gabor modulation were outperformed by LGCNs. This multi-task success displayed the general applicability of learnable complex-valued Gabor modulation.

Generating features dependent on global relations is extremely important for image understanding. Scenes are understood through both short and long range correlations, as the surrounding setting of an object often is linked to its characteristics. For global contaminants in large images which can appear visually similar to interesting objects, context is especially relevant for correct identification. Orientational information can also be a useful distinguishing factor for such identification. A limitation of purely convolutional networks is that global features are typically only learned in the final stages of the network after heavy downsampling. In Chapter 5, we investigated the use of attention to extract global features in contaminated large images. A multi-scale attention architecture was implemented, involving attention computed over different scales in parallel. We utilised the Gabor filter modulation introduced in the previous chapter to extract orientational features, and computed attention with respect to angles of each modulating Gabor filter to capture long range orientational dependencies. To ease the computational burden of attention calculation, a gridded attention method was proposed where features are divided into tiles of different scales before computing attention. The detailed method was validated thoroughly on multiple datasets, both in terms of accuracy performance and computational characteristics, and an optimal network configuration was selected through ablation studies. This optimal network was evaluated on the SWIMSEG dataset of natural clouds, where the model achieved state of the art performance.

While the gridded Gabor attention work described above provided a ML architecture well suited to the cirrus contamination problem, several steps remained for practical application on the cirrus segmentation task in LSB images. In Chapter 6, we presented a machine learning pipeline for segmentation of cirrus contamination. As annotation data of Chapter 3 is naturally probabilistic, where more annotators labelling a structure corresponded with a higher label probability, we proposed a consensus based loss function to enable neural networks to train on coarsely probabilistic labels. An adaptive intensity

scaling operation was designed to enhance subtle structures and adjust scaling parameters based on performance. Scaling parameters were able to be learned alongside network weights through backpropagation. We detailed the data augmentation and transfer learning steps taken to ensure generalisation of the neural network despite limited data. These measures were combined with the attention model of Chapter 5, and applied on cirrus segmentation. A comprehensive study was performed evaluating the performance characteristics of the pipeline, where we demonstrated that the proposed consensus loss significantly tackled class imbalance issues. Methods of combining multiple predictions were evaluated, where we found that an ensemble of multiple models produced the best results. The final pipeline achieved reliable automated cirrus segmentation on limited data with no preprocessing.

A central goal of this thesis was to build knowledge surrounding the application of ML to LSB images and propose a method for automated cataloguing of galactic structures. Armed with findings from thorough studies on applying CNNs to LSB images, detailed in previous chapters, in Chapter 7 we refocused our attention to the wider goal of this thesis. We surveyed literature surrounding instance segmentation and astronomy, and identified that a common weakness in works is confusion between contaminants and interesting structures. Based on this, we hypothesised that unification of contamination and galactic structure segmentation would mitigate against this problem and investigated the use of a panoptic segmentation model. We modified Mask R-CNN, a popular instance segmentation model, combining it with the cirrus segmentation model of Chapter 6, to create a model capable of segmenting both localised galactic structures and extended homogeneous contaminants simultaneously. In the first experiment, Mask R-CNN was applied to only instance segmentation of galactic structures. During this experiment, to investigate the impact of training data quality and quantity, we used a human-in-the-loop training regime where correct predictions of unannotated objects were added into the training dataset. A simple approach resulted in a significant amount of unannotated objects being added into the dataset, and we found that performance was increased on the original dataset for some classes, showing the potential of both a larger training dataset and a more sophisticated semi-supervised learning approach. Finally, we applied the proposed panoptic segmentation model on the task of segmenting galactic structures and contaminants, where we found that unification of tasks indeed led to greater performance. Reliable inference was achieved with the panoptic approach using feasible computational resources and on images with minimal preprocessing, demonstrating the potential of careful deep learning approaches for automated cataloguing in LSB images.

8.1 Contributions

The main contributions of this work can be summarised as follows.

- **Annotation tool for astronomical images and dataset of labelled structures in MATLAS images.** We designed an annotation tool for astronomers to draw shapes overlaying structures in astronomical images. Multiple domain specific considerations were taken to ensure functionality for both machine learning and astronomy researchers using the tool. We detailed a dataset produced using the tool of shapes drawn by four users, delineating structures in MATLAS LSB images.
- **Orientation robustness in convolutions with learnable complex-valued Gabor modulated convolutions.** We presented a convolutional layer where kernels are modified with Gabor filters to render them more sensitive to orientational patterns. This layer was extended to the cyclic Gabor convolution, where further rotational weight-tying is enforced thus increasing orientational robustness.
- **Efficient attention operator for global features rich in orientational information.** We proposed a memory efficient attention layer capable of generating global features on large images. We integrated Gabor filter modulation into the attention operator to facilitate measuring long range orientational patterns.
- **Pipeline for segmentation of cirrus contamination in LSB images.** We applied the gridded Gabor attention network to segmentation of cirrus contamination in LSB images. A loss function for training neural networks on coarsely probabilistic target labels was proposed. We presented a simple adaptive intensity scaling operation which adjusts scaling parameters based on network performance.
- **Automated cataloguing of galactic structures and cirrus with a panoptic segmentation model.** We presented a panoptic segmentation model for simultaneous delineation of galactic structures and cirrus contamination. The processing of MATLAS annotation data into instance segmentation labels was described. Mask R-CNN was used to set a baseline on segmentation of localised galactic structures, and to perform a preliminary study on the potential of a semi-supervised framework.
- **Source code for all work will be freely available online.** All source code relating to methods, studies or tools described in this thesis will be publicly available for other researchers to use.

8.2 Future Work

Throughout each chapter within this thesis we have discussed ideas that would be interesting to explore in future work. In this final section we will reiterate these ideas, elaborate and build on them, and present new research ideas gleaned from viewing the thesis as a holistic work.

Annotation tool – The presented annotation tool enabled efficient and precise drawing of shapes on astronomical images, but after extended use we identified areas where it could be improved. The ability to directly adjust contrast, saturation and hue of displayed images would be greatly helpful for astronomers to be able to inspect subtle features. While this feature can be made partially available by switching between HiPS layers or surveys, it requires images with different colour characteristics to be pre-computed. In addition, the option to annotate association between drawn shapes would be a greatly beneficial feature, as structures such as diffuse halos and tidal features are inherently linked to a ‘source’ galaxy. This feature would first facilitate astronomy motivated statistical analysis investigating correlating factors between the shapes of related objects, and secondly open the possibility of computer vision research attempting to either predict association or use the information to improve segmentation methods, such as those presented in this thesis.

Synthesised cirrus images – The cirrus synthesis approach developed to supplement training data led to benefits through transfer learning, however the algorithm was fairly elementary. It would be interesting to utilise recent advancements in generative modelling to create more realistic examples of cirrus. GANs are one option for this task, having achieved photo-realistic quality on image generation tasks. Another option would be normalising flows which have seen recent use in physics aware contexts. Another avenue extending the synthesis work would be to combine our method with real images as an augmentation method. As cirrus contamination presents in roughly a quarter of MATLAS images but contaminates a substantial region, there is exists a difficult class imbalance problem which we discussed and tackled. Clean images augmented with synthesised cirrus contamination could further help ease this problem and potentially increase prediction accuracy. Data synthesis and these discussed ideas could also be extended to generate tidal features, which we struggled to detect in later chapters due to the even more difficult class imbalance problem and their subtle nature.

Gabor modulation – We found that modulating convolutional kernels with Gabor filters worked well for rotation sensitivity, though it is possible that other analytical filters could change performance characteristics. The use of Gabor filters was motivated by previous works and that they are well understood filters on orientation-dependent tasks.

A comparative analysis of other filters could more concretely justify this choice, or even expose stronger choices on certain tasks. This analysis could also include applications on different datasets to more clearly illuminate how different modulating filter choices affect learned features. It would also be informative to integrate the proposed modulated convolutions into more complex and popular architectures. For example, a ResNet style architecture modified to use modulated convolutions and applied on typical benchmark data would show how the proposed method scales to larger and more difficult tasks.

Gridded Gabor attention – With regards to the gridded Gabor attention methodology, we have identified several avenues for future work. Though the model achieved strong segmentation results, the fixed nature of tile grids could introduce boundary artefacts between tiles. While there exists some overlap tiles of different scales, it could be interesting to investigate how overlapping tiles further would impact results. This could be achieved either by simply increasing each tile size so that boundary regions overlap, or shifting the tiling grid in a cyclic fashion similar to modern transformer models. Further applications of the model on different data would also be interesting to explore. Medical images in particular often exhibit similar characteristics to those used in the study surrounding the attention method, i.e. they are very large, require global features and often suffer from artefacts. Finally, there are certainly other architectural modifications which could be made to improve performance, such as increasing the complexity of the final upsampling stage, and skip connections between the backbone and upsampling network sections.

Cirrus segmentation – We tackled the cirrus contamination problem as a binary segmentation task due to the annotated data available during the study. In reality, cirrus contamination presents in a wide spectrum of strengths and taking this fact into account would be very interesting. Given the large variety in strength, it is possible that there is a detrimental effect on model performance by forcing all cirrus into a single class. This seems likely given that there are instances where strong and obvious cirrus contamination fails to be predicted. Nonetheless, a study that accommodates this variation would inform the community on how ML approaches handle different types of cirrus and importantly would provide guidance on which areas to focus efforts on. This accommodation could be achieved by either adding multiple cirrus classes or using a strength parameter which would be regressed. The annotation tool of this thesis could be adapted to allow astronomers to provide training data for this more granular task.

Galactic structure segmentation – The task of segmenting galactic structures was approached as a panoptic segmentation problem, for which we presented a model combining Mask R-CNN and gridded Gabor attention. Despite the model’s simplicity, we achieved increased performance on the galactic structure segmentation in comparison

to the Mask R-CNN instance segmentation baseline. It would be interesting to explore the impact of more complex panoptic segmentation models on prediction accuracy. This could be achieved by intertwining the semantic and instance branches so that intermediate features of each branch inform the other, as was described in related work. While we determined that this was not a priority as pixel classes overlapped in our problem, connection between branches could alleviate any confusion between similar classes such as tidal features and cirrus contamination. Aside from convolutional networks, transformer models have achieved promising results on panoptic tasks, and thus could also be a good path to pursue.

Semi-supervised segmentation – During experiments on galactic structure segmentation, we performed a preliminary study on using semi-supervised learning to mitigate against weak annotations, i.e. images where not all present objects were annotated. With a trivial human-in-the-loop training protocol, a significant number of objects was discovered and added into the dataset. We also observed that performance was increased on classes with particularly weak annotations, such as ghosted halos. It would be interesting to investigate how a more sophisticated semi-supervised approach could be used to leverage either of these findings. A possible avenue would be to explore using a noisy student framework. This involves first training a ‘teacher’ model as normal, and then training another model, the ‘student’, on available target labels and ‘pseudo-labels’ generated by the first with noisy input data. This process is repeated iteratively with the student being recycled as the teacher. The noisy-student framework could be used to expand the training dataset by adding new unannotated objects, as in our study, or generating labels on new unannotated images.

Future survey data – Throughout this thesis, data quantity and quality were major challenges. While we developed approaches to mitigate against these issues, they undoubtedly limited both model performance and the confidence in any insights drawn from experiments. With future surveys providing higher quality data and more LSB images, it would be incredibly interesting to examine the effects of relieving the data limitations from our studies. Developing a better understanding of how the proposed methodologies scale to future surveys is vital for research progress. A larger and less noisy training dataset would also greatly clarify the strengths and weaknesses of our approaches and guide future work. Increasing the number of annotators would also facilitate interesting research, as target label confidence should increase and thus label noise should decrease. This would also enable exploring the use of inherently probabilistic approaches such as those mentioned during this work.

Cirrus contamination removal – Viewing the thesis as a whole, a project which follows naturally would be to attempt to remove contaminants with deep learning. We

developed techniques for automated cataloguing of structures, despite contaminants occluding objects. It would be interesting to explore how our methodologies could be extended to decontaminate these instances, given that in this thesis we tried to take into account how ML models would process contaminants and interesting objects. Works utilising conditional GANs for image-to-image tasks have seen success, and decontamination naturally falls into this category of tasks. Future work could involve somehow combining the tasks of decontamination and segmentation, so that the tasks become intertwined. We hypothesise that better decontamination should result in better instance segmentation predictions, and thus segmentation could provide a strong auxiliary goal to account for the lack of ground truth.

To conclude, the application of modern deep learning to astronomy and especially LSB image processing is in its infancy. In recent years, basic off-the-shelf approaches have achieved great results in astronomy. With more careful design and domain motivated computer vision research, such as in this thesis, we believe that machine learning will have a large impact on automated cataloguing and the wider astronomy research sphere.

Bibliography

- [1] R. G. Abraham and P. G. van Dokkum. Ultra-low surface brightness imaging with the dragonfly telephoto array. Publications of the Astronomical Society of the Pacific, 126(935):55, 2014.
- [2] J. Akeret, C. Chang, A. Lucchi, and A. Refregier. Radio frequency interference mitigation using deep convolutional neural networks. Astronomy and computing, 18:35–39, 2017.
- [3] M. Akhlaghi and T. Ichikawa. Noise-based detection and segmentation of nebulous objects. The Astrophysical Journal Supplement Series, 220(1):1, 2015.
- [4] Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, K. Benabed, and others. Planck intermediate results-LV. Reliability and thermal properties of high-frequency sources in the Second Planck Catalogue of Compact Sources. Astronomy & Astrophysics, 644:A99, 2020.
- [5] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab. Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. IEEE transactions on medical imaging, 35(5):1313–1321, 2016.
- [6] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour Detection and Hierarchical Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell., 33(5):898–916, 5 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.161. URL <http://dx.doi.org/10.1109/TPAMI.2010.161>.
- [7] S. Arivazhagan, L. Ganesan, and S. P. Priyal. Texture classification using Gabor wavelets based rotation invariant features. Pattern Recognition Letters, 27(16):1976–1982, 12 2006. doi: 10.1016/J.PATREC.2006.05.008.
- [8] M. Arjovsky, A. Shah, and Y. Bengio. Unitary evolution recurrent neural networks. In International Conference on Machine Learning, pages 1120–1128, 2016.

- [9] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, and others. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans. Medical Physics, 38(2):915–931, 1 2011. ISSN 00942405. doi: 10.1118/1.3528204.
- [10] Armato III Samuel G., G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, and others. Data From LIDC-IDRI, 2015. URL <https://wiki.cancerimagingarchive.net/x/rgAe>.
- [11] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano. Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. IEEE Transactions on Medical Imaging, 28(8):1266–1277, 2009. doi: 10.1109/TMI.2009.2014372.
- [12] R. Baena-Gallé, R. Infante-Sainz, M. Akhlaghi, I. Trujillo, and J. H. Knapen. Extended Point-spread Functions for Deep Astronomical Imaging Surveys. Research Notes of the AAS, 4(7):124, 7 2020. doi: 10.3847/2515-5172/abaaa8. URL <https://doi.org/10.3847/2515-5172/abaaa8>.
- [13] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. In International Conference on Learning Representations, 2015.
- [14] N. M. Ball, J. Loveday, M. Fukugita, O. Nakamura, S. Okamura, J. Brinkmann, and R. J. Brunner. Galaxy types in the Sloan Digital Sky survey using supervised artificial neural networks, 2004. ISSN 00358711.
- [15] D. Bazell. Feature relevance in morphological galaxy classification. Monthly Notices of the Royal Astronomical Society, 2000. ISSN 00358711. doi: 10.1046/j.1365-8711.2000.03525.x.
- [16] D. Bazell and D. Aha. Ensembles of Classifiers for Morphological Galaxy Classification. The Astrophysical Journal, 2001. ISSN 0004-637X. doi: 10.1086/318696.
- [17] E. J. Bekkers, M. W. Lafarge, M. Veta, K. A. J. Eppenhof, J. P. W. Pluim, and R. Duits. Roto-translation covariant convolutional networks for medical image analysis. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 440–448, 2018.
- [18] K. Bekki. Quantifying the fine structures of disk galaxies with deep learning: Segmentation of spiral arms in different Hubble types. Astronomy & Astrophysics, 647: A120, 2021.

- [19] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. Astronomy and astrophysics supplement series, 117(2):393–404, 1996.
- [20] R. W. Bickley, C. Bottrell, M. H. Hani, S. L. Ellison, H. Teimoorinia, K. M. Yi, S. Wilkinson, S. Gwyn, and M. J. Hudson. Convolutional neural network identification of galaxy post-mergers in UNIONS using IllustrisTNG. Monthly Notices of the Royal Astronomical Society, 504(1):372–392, 2021.
- [21] M. Bílek, P.-A. Duc, J.-C. Cuillandre, S. Gwyn, M. Cappellari, D. V. Bekaert, P. Bonfini, T. Bitsakis, S. Paudel, D. Krajnović, and others. Census and classification of low-surface-brightness structures in nearby early-type galaxies from the MATLAS survey. Monthly Notices of the Royal Astronomical Society, 498(2):2138–2166, 2020.
- [22] N. Bjorck, C. P. Gomes, B. Selman, and K. Q. Weinberger. Understanding batch normalization. Advances in neural information processing systems, 31, 2018.
- [23] M. R. Blanton, M. A. Bershady, B. Abolfathi, F. D. Albareti, C. Allende Prieto, A. Almeida, J. Alonso-García, F. Anders, S. F. Anderson, B. Andrews, and others. Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. 154(1):28, 7 2017. doi: 10.3847/1538-3881/aa7567.
- [24] T. Boch and P. Fernique. Aladin Lite: Embed your Sky in the browser. Astronomical data analysis software and systems XXIII, 485:277, 2014.
- [25] F. Bonnarel, P. Fernique, O. Bienaymé, D. Egret, F. Genova, M. Louys, F. Ochsenbein, M. Wenger, and J. G. Bartlett. The ALADIN interactive sky atlas-A reference tool for identification of astronomical sources. Astronomy and Astrophysics Supplement Series, 143(1):33–40, 2000.
- [26] L. Bottou. Stochastic gradient descent tricks. In Neural networks: Tricks of the trade, pages 421–436. Springer, 2012.
- [27] A. Boucaud, C. Heneka, E. E. O. Ishida, N. Sedaghat, R. S. de Souza, B. Moews, H. Dole, M. Castellano, E. Merlin, V. Roscani, and others. Photometry of high-redshift blended galaxies using deep learning. Monthly Notices of the Royal Astronomical Society, 491(2):2481–2495, 2020.
- [28] T. Brahe. 1598, Stellarum octavi orbis inerrantium accurata restitutio (manuscript), ed. A. Dreyer, 333, 1916.

- [29] A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. De Bruijne, C. Babusiaux, M. Biermann, O. L. Creevey, D. W. Evans, L. Eyer, A. Hutton, and others. Gaia early data release 3-summary of the contents and survey properties. *Astronomy & Astrophysics*, 649:A1, 2021.
- [30] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.
- [31] C. J. Burke, P. D. Aleo, Y.-C. Chen, X. Liu, J. R. Peterson, G. H. Sembroski, and J. Y.-Y. Lin. Deblending and classifying astronomical sources with Mask R-CNN deep learning. 490(3):3952–3965, 12 2019. doi: 10.1093/mnras/stz2845.
- [32] Z. Cai and N. Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018.
- [33] K. C. Chambers, E. A. Magnier, N. Metcalfe, H. A. Flewelling, M. E. Huber, C. Z. Waters, L. Denneau, P. W. Draper, D. Farrow, D. P. Finkbeiner, and others. The pan-starrs1 surveys. *arXiv preprint arXiv:1612.05560*, 2016.
- [34] H. Chen, X. Qi, L. Yu, and P.-A. Heng. DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation. 2016. ISSN 10636919. doi: 10.1109/CVPR.2016.273.
- [35] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille. Attention to Scale: Scale-aware Semantic Image Segmentation. *CoRR*, abs/1511.0, 2015. URL <http://arxiv.org/abs/1511.03339>.
- [36] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. ISSN 01628828. doi: 10.1109/TPAMI.2017.2699184.
- [37] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.
- [38] X. Cheng, Q. Qiu, R. Calderbank, and G. Sapiro. RotDCF: decomposition of convolutional filters for rotation-equivariant deep networks. In *Proceedings of the 7th International Conference on Learning Representations*, 2018.

- [39] D. Ciregan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In 2012 IEEE conference on computer vision and pattern recognition, pages 3642–3649, 2012.
- [40] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. Advances in neural information processing systems, 25, 2012.
- [41] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In Twenty-second international joint conference on artificial intelligence, 2011.
- [42] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. Journal of Digital Imaging, 26(6):1045–1057, 12 2013. ISSN 0897-1889. doi: 10.1007/s10278-013-9622-7.
- [43] T. S. Cohen and M. Welling. Group Equivariant Convolutional Networks. In International conference on Machine Learning, pages 2990–2999, 2016. ISBN 9781510829008. doi: 10.1017/S1464793101005656.
- [44] T. S. Cohen and M. Welling. Steerable cnns. In Proceedings of the 5th International Conference on Learning Representations, 2017.
- [45] T. S. Cohen, M. Geiger, J. Köhler, and M. Welling. Spherical cnns. In Proceedings of the 6th International Conference on Learning Representations, 2018.
- [46] S. Cole, C. G. Lacey, C. M. Baugh, , and C. S. Frenk. Hierarchical galaxy formation. Monthly Notices of the Royal Astronomical Society, 319(1):168–204, 2000.
- [47] A. P. Cooper, S. Cole, C. S. Frenk, S. D. M. White, J. Helly, A. J. Benson, G. De Lucia, A. Helmi, A. Jenkins, J. F. Navarro, and others. Galactic stellar haloes in the CDM model. Monthly Notices of the Royal Astronomical Society, 406(2):744–766, 2010.
- [48] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3213–3223, 2016.

- [49] M. Couprie and G. Bertrand. Topological gray-scale watershed transformation. In Vision Geometry VI, volume 3168, pages 136–146, 1997.
- [50] J. Dai, K. He, and J. Sun. Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3150–3158, 2016.
- [51] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255, 2009.
- [52] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, and others. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage, 31(3):968–980, 2006.
- [53] S. Dev, Y. H. Lee, and S. Winkler. Color-Based Segmentation of Sky/Cloud Images From Ground-Based Cameras. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(1):231–242, 2017.
- [54] S. Dev, A. Nautiyal, Y. H. Lee, and S. Winkler. CloudSegNet: A deep network for nychthemeron cloud image segmentation. IEEE Geoscience and Remote Sensing Letters, 16(12):1814–1818, 2019.
- [55] S. Dieleman, K. W. Willett, and J. Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. Monthly Notices of the Royal Astronomical Society, 450(2):1441–1459, 2014. doi: 10.1186/1748-5908-4-32.
- [56] S. Dieleman, J. De Fauw, and K. Kavukcuoglu. Exploiting Cyclic Symmetry in Convolutional Neural Networks. International Conference on Machine Learning, 48:1889–1898, 2016. doi: 10.1525/aa.1976.78.4.02a00030.
- [57] M. J. Disney. Visibility of galaxies. Nature, 263(5578):573–575, 1976.
- [58] W. Dobbels, S. Krier, S. Pirson, S. Viaene, G. De Geyter, S. Salim, and M. Baes. Morphology-assisted galaxy mass-to-light predictions using deep learning. Astronomy & Astrophysics, 624:A102, 2019.
- [59] H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer. Improving galaxy morphologies for SDSS with Deep Learning. Monthly Notices of the Royal Astronomical Society, 476(3):3661–3676, 2018.

- [60] H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, S. Kaviraj, J. L. Fischer, T. M. C. Abbott, F. B. Abdalla, J. Annis, S. Avila, D. Brooks, and others. Transfer learning for galaxy morphology from one survey to another. Monthly Notices of the Royal Astronomical Society, 484(1):93–100, 2019.
- [61] P.-A. Duc. MATLAS: a deep exploration of the surroundings of massive early-type galaxies. arXiv preprint arXiv:2007.13874, 2020.
- [62] P.-A. Duc, J.-C. Cuillandre, E. Karabal, M. Cappellari, K. Alatalo, L. Blitz, F. Bournaud, M. Bureau, A. F. Crocker, R. L. Davies, and others. The ATLAS3D project–XXIX. The new look of early-type galaxies and surrounding fields disclosed by extremely deep optical images. Monthly Notices of the Royal Astronomical Society, 446(1):120–143, 2015.
- [63] P.-A. Duc, J.-C. Cuillandre, and F. Renaud. Revisiting Stephan’s Quintet with deep optical images. Monthly Notices of the Royal Astronomical Society: Letters, 475(1):L40–L44, 2018.
- [64] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. Journal of machine learning research, 12(7), 2011.
- [65] V. Dumoulin and F. Visin. A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285, 2016.
- [66] R. Ejbali and M. Zaied. A dyadic multi-resolution deep convolutional neural wavelet network for image classification. Multimedia Tools and Applications, 77(5):6149–6163, 2018.
- [67] T. Erben, H. Hildebrandt, L. Miller, L. van Waerbeke, C. Heymans, H. Hoekstra, T. D. Kitching, Y. Mellier, J. Benjamin, C. Blake, and others. CFHTLenS: the Canada–France–Hawaii telescope lensing survey–imaging data and catalogue products. Monthly Notices of the Royal Astronomical Society, 433(3):2545–2563, 2013.
- [68] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, and others. A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd, volume 96, pages 226–231, 1996.
- [69] H. Farias, D. Ortiz, G. Damke, M. J. Arancibia, and M. Solar. Mask galaxy: Morphological segmentation of galaxies. Astronomy and Computing, page 100420, 2020.

- [70] C. Feng and J. Zhang. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. Solar Energy, 204:71–78, 2020.
- [71] P. Fernique, M. G. Allen, T. Boch, A. Oberto, F. X. Pineau, D. Durand, C. Bot, L. Cambresy, S. Derriere, F. Genova, and others. Hierarchical progressive surveys-Multi-resolution HEALPix data structures for astronomical images, catalogues, and 3-dimensional data cubes. Astronomy & Astrophysics, 578:A114, 2015.
- [72] M. Finzi, S. Stanton, P. Izmailov, and A. G. Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In International Conference on Machine Learning, pages 3165–3176, 2020.
- [73] B. Fischl. FreeSurfer. Neuroimage, 62(2):774–781, 2012.
- [74] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, and others. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron, 33(3):341–355, 2002.
- [75] S. R. Folkes, O. Lahav, and S. J. Maddox. An artificial neural network approach to the classification of galaxy spectra. Monthly Notices of the Royal Astronomical Society, 283(2):651–665, 1996.
- [76] W. T. Freeman, E. H. Adelson, and others. The design and use of steerable filters. IEEE Transactions on Pattern analysis and machine intelligence, 13(9):891–906, 1991.
- [77] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu. Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3146–3154, 2019.
- [78] S. Fujieda, K. Takayama, and T. Hachisuka. Wavelet convolutional neural networks. arXiv preprint arXiv:1805.08620, 2018.
- [79] K. Fukushima and S. Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In Competition and cooperation in neural nets, pages 267–285. Springer, 1982.
- [80] H. Gabbard, M. Williams, F. Hayes, and C. Messenger. Matching matched filtering with deep networks for gravitational-wave astronomy. Physical review letters, 120(14):141103, 2018.

- [81] R. Gens and P. M. Domingos. Deep symmetry networks. In Advances in Neural Information Processing Systems, pages 2537–2545, 2014.
- [82] R. Girshick. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, 2015. ISBN 9781467383912. doi: 10.1109/ICCV.2015.169.
- [83] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v5). Technical report. URL <http://www.cs.berkeley.edu/~rbg/rcnn>.
- [84] R. E. González, R. P. Munoz, and C. A. Hernández. Galaxy detection and identification using deep learning and data augmentation. Astronomy and computing, 25: 103–109, 2018.
- [85] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li. CDnetV2: CNN-based cloud detection for remote sensing imagery with cloud-snow coexistence. IEEE Transactions on Geoscience and Remote Sensing, 59(1):700–713, 2020.
- [86] G. M. Haley and B. S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. IEEE transactions on Image Processing, 8(2): 255–269, 1999.
- [87] R. Hausen and B. E. Robertson. Morpheus: A deep learning framework for the pixel-level analysis of astronomical image data. The Astrophysical Journal Supplement Series, 248(1):20, 2020.
- [88] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, pages 1026–1034, 2015.
- [89] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9):1904–1916, 2015.
- [90] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.
- [91] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961–2969, 2017.

- [92] J. F. Henriques and A. Vedaldi. Warped convolutions: Efficient invariance to spatial transformations. In Proceedings of the 34th International Conference on Machine Learning–Volume 70, pages 1461–1469, 2017.
- [93] C. Heymans, L. Van Waerbeke, L. Miller, T. Erben, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, P. Simon, C. Bonnett, and others. CFHTLenS: the Canada–France–Hawaii telescope lensing survey. Monthly Notices of the Royal Astronomical Society, 427(1):146–166, 2012.
- [94] H. Hildebrandt, T. Erben, K. Kuijken, L. van Waerbeke, C. Heymans, J. Coupon, J. Benjamin, C. Bonnett, L. Fu, H. Hoekstra, and others. CFHTLenS: improving the quality of photometric redshifts with precision photometry. Monthly Notices of the Royal Astronomical Society, 421(3):2355–2367, 2012.
- [95] S. Hu, D. Worrall, S. Knegt, B. Veeling, H. Huisman, and M. Welling. Supervised uncertainty quantification for segmentation with multiple annotations. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 137–145, 2019.
- [96] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and others. Speed/accuracy trade-offs for modern convolutional object detectors. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7310–7311, 2017.
- [97] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang. Mask scoring r-cnn. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6409–6418, 2019.
- [98] E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. Proceedings of the National Academy of Sciences, 15(3):168–173, 1929. doi: 10.1073/pnas.15.3.168. URL <https://www.pnas.org/doi/abs/10.1073/pnas.15.3.168>.
- [99] B. Hubert, B. Alexandre, and H.-C. Marc. Probabilistic segmentation of overlapping galaxies for large cosmological surveys. arXiv preprint arXiv:2111.15455, 2021.
- [100] M. Huertas-Company, J. A. L. Aguerri, M. Bernardi, S. Mei, and J. S. Almeida. Revisiting the Hubble sequence in the SDSS DR7 spectroscopic sample: a publicly available bayesian automated classification. 2010. ISSN 0004-6361. doi: 10.1051/0004-6361/201015735.

- [101] M. Huertas-Company, R. Gravet, G. Cabrera-Vives, P. G. Pérez-González, J. S. Kartaltepe, G. Barro, M. Bernardi, S. Mei, F. Shankar, P. Dimauro, E. F. Bell, and others. A catalog of visual-like morphologies in the 5 candels fields using deep learning. *The Astrophysical Journal Supplement Series*, 221(1):8, 2015.
- [102] J. E. Iglesias and M. R. Sabuncu. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219, 2015.
- [103] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456, 2015.
- [104] T. Jackson, M. Werner, and T. N. Gautier III. A catalog of bright filamentary structures in the interstellar medium. *The Astrophysical Journal Supplement Series*, 149(2):365, 2003.
- [105] J.-H. Jacobsen, B. De Brabandere, and A. W. M. Smeulders. Dynamic steerable blocks in deep residual networks. In *Proceedings of the British Machine Vision Conference*, 2017.
- [106] M. Jaderberg, K. Simonyan, A. Zisserman, and others. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015.
- [107] X. Jia, B. De Brabandere, T. Tuytelaars, and L. V. Gool. Dynamic filter networks. In *Advances in Neural Information Processing Systems*, pages 667–675, 2016.
- [108] A. Kanazawa, A. Sharma, and D. W. Jacobs. Locally Scale-Invariant Convolutional Neural Networks. *Deep Learning and Representation Learning Workshop: Neural Information Processing System*, 2014.
- [109] H. Khan and B. Yener. Learning filter widths of spectral decompositions with wavelets. In *Advances in Neural Information Processing Systems*, pages 4601–4612, 2018.
- [110] E. J. Kim and R. J. Brunner. Star-galaxy classification using deep convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, page stw2672, 2016.
- [111] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, 2014.

- [112] A. Kirillov, R. Girshick, K. He, and P. Dollár. Panoptic feature pyramid networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6399–6408, 2019.
- [113] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár. Panoptic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9404–9413, 2019.
- [114] J. Kivinen, C. Williams, and N. Heess. Visual Boundary Prediction: A Deep Neural Prediction Network and Quality Dissection. In S. Kaski and J. Corander, editors, Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, volume 33 of Proceedings of Machine Learning Research, pages 512–521, Reykjavik, Iceland, 5 2014. PMLR. URL <http://proceedings.mlr.press/v33/kivinen14.html>.
- [115] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Medical Physics, 35(4):1407–1417, 3 2008. ISSN 00942405. doi: 10.1118/1.2842076.
- [116] S. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. Maier-Hein, S. M. A. Eslami, D. Jimenez Rezende, and O. Ronneberger. A Probabilistic U-Net for Segmentation of Ambiguous Images. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 31. Curran Associates, Inc., 2018.
- [117] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, pages 1097–1105, 2012.
- [118] D. Laptev, N. Savinov, J. M. Buhmann, and M. Pollefeys. TI-POOLING: transformation-invariant pooling for feature learning in convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 289–297, 2016.
- [119] R. Laureijs, J. Amiaux, S. Arduini, J.-L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, and others. Euclid definition study report. arXiv preprint arXiv:1110.3193, 2011.

- [120] J. Lazarow, K. Lee, K. Shi, and Z. Tu. Learning instance occlusion for panoptic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10720–10729, 2020.
- [121] H. S. Leavitt and E. C. Pickering. Periods of 25 Variable Stars in the Small Magellanic Cloud. Harvard College Observatory Circular, 173:1–3, 3 1912.
- [122] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.
- [123] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller. Efficient BackProp. In K.-R. Orr Genevieve B. }and Müller, editor, Neural Networks: Tricks of the Trade, pages 9–50. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998. ISBN 978-3-540-49430-0. doi: 10.1007/3-540-49430-8{_}2. URL https://doi.org/10.1007/3-540-49430-8_2.
- [124] C. Leinert, S. Bowyer, L. K. Haikala, M. S. Hanner, M. G. Hauser, A.-C. Levasseur-Regourd, I. Mann, K. Mattila, W. T. Reach, W. Schlosser, and others. The 1997 reference of diffuse night sky brightness. Astronomy and Astrophysics Supplement Series, 127(1):1–99, 1998.
- [125] C. Levy, A. Ciprijanovic, A. Drlica-Wagner, B. Mutlu-Pakdil, B. Nord, and D. Tanoglidis. Detecting Low Surface Brightness Galaxies with Mask R-CNN. Technical report, Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 2021.
- [126] H. Li, P. Xiong, J. An, and L. Wang. Pyramid attention network for semantic segmentation. British Machine Vision Conference, 2018.
- [127] Y. Li, X. Chen, Z. Zhu, L. Xie, G. Huang, D. Du, and X. Wang. Attention-guided unified network for panoptic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7026–7035, 2019.
- [128] C.-H. Lin and S. Lucey. Inverse compositional spatial transformer networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2568–2576, 2017.
- [129] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In European conference on computer vision, pages 740–755, 2014.

- [130] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2117–2125, 2017.
- [131] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal Loss for Dense Object Detection. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 2999–3007, 2017. doi: 10.1109/ICCV.2017.324.
- [132] T. Lindeberg. Feature detection with automatic scale selection. International journal of computer vision, 30(2):79–116, 1998.
- [133] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, and others. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society, 389(3):1179–1189, 2008.
- [134] G. Litjens, O. Debats, W. van de Ven, N. Karssemeijer, and H. Huisman. A pattern recognition approach to zonal segmentation of the prostate on MRI. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 413–420, 2012.
- [135] F. Liu, G. Lin, and C. Shen. CRF Learning with CNN Features for Image Segmentation. CoRR, abs/1503.0, 2015. URL <http://arxiv.org/abs/1503.08263>.
- [136] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia. Path aggregation network for instance segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8759–8768, 2018.
- [137] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai. Richer convolutional features for edge detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3000–3009, 2017.
- [138] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015.
- [139] D. G. Lowe. Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision, volume 2, pages 1150–1157, 1999.
- [140] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu. Gabor convolutional networks. IEEE Transactions on Image Processing, 27(9):4357–4366, 2018.

- [141] Z. Ma, H. Xu, J. Zhu, D. Hu, W. Li, C. Shan, Z. Zhu, L. Gu, J. Li, C. Liu, and others. A Machine Learning Based Morphological Classification of 14,245 Radio AGNs Selected from the Best–Heckman Sample. The Astrophysical Journal Supplement Series, 240(2):34, 2019.
- [142] A. L. Maas, A. Y. Hannun, A. Y. Ng, and others. Rectifier nonlinearities improve neural network acoustic models. In Proc. icml, volume 30, page 3, 2013.
- [143] S. Marcel and Y. Rodriguez. Torchvision the machine-vision package of torch. In Proceedings of the 18th ACM international conference on Multimedia, pages 1485–1488, 2010.
- [144] D. Marcos, M. Volpi, and D. Tuia. Learning rotation invariant convolutional filters for texture classification. In 2016 23rd International Conference on Pattern Recognition (ICPR), pages 2012–2017, 2016.
- [145] D. Marcos, M. Volpi, N. Komodakis, and D. Tuia. Rotation equivariant vector field networks. In Proceedings of the IEEE International Conference on Computer Vision, pages 5048–5057, 2017.
- [146] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, volume 2, pages 416–423, 2001. doi: 10.1109/ICCV.2001.937655.
- [147] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, and others. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging, 34(10):1993–2024, 2015. doi: 10.1109/TMI.2014.2377694.
- [148] M.-A. Miville-Deschênes, P.-A. Duc, F. Marleau, J.-C. Cuillandre, P. Didelon, S. Gwyn, and E. Karabal. Probing interstellar turbulence in cirrus with deep optical imaging: no sign of energy dissipation at 0.01 pc scale. Astronomy & Astrophysics, 593:A4, 2016.
- [149] R. Mohan and A. Valada. Efficientps: Efficient panoptic segmentation. International Journal of Computer Vision, 129(5):1551–1579, 2021.
- [150] A. Naim, O. Lahav, R. J. Buta, H. G. Corwin, G. De Vaucouleurs, A. Dressler, J. P. Huchra, S. Van Den Bergh, S. Raychaudhury, L. Sodrr, and M. C. Storrie-Lombardi.

- A COMPARATIVE STUDY OF MORPHOLOGICAL CLASSIFICATIONS OF APM GALAXIES. Technical Report 4, 1995.
- [151] A. Naim, O. Lahav, L. Sodrr, and M. C. Storrie-Lombardi. AUTOMATED MORPHOLOGICAL CLASSIFICATION OF APM GALAXIES BY SUPERVISED ARTIFICIAL NEURAL NETWORKS. Monthly Notices of the Royal Astronomical Society, 275(3):567–590, 1995. doi: <https://doi.org/10.1093/mnras/275.3.567>.
- [152] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In Icml, 2010.
- [153] P. Nilson. Uppsala general catalogue of galaxies. Acta Universitatis Upsaliensis. Nova Acta Regiae Societatis Scientiarum Upsaliensis-Uppsala Astronomiska Observatoriums Annaler, 1973.
- [154] A. Odena, V. Dumoulin, and C. Olah. Deconvolution and Checkerboard Artifacts. Distill, 2016. doi: 10.23915/distill.00003. URL <http://distill.pub/2016/deconv-checkerboard>.
- [155] S. C. Odewahn, E. B. Stockwell, R. L. Pennington, R. M. Humphreys, and W. A. Zumach. AUTOMATED STAR/GALAXY DISCRIMINATION WITH NEURAL NETWORKS. The Astronomical Journal, 103:318–331, 1992.
- [156] S. C. Odewahn, R. M. Humphreys, G. Aldering, and P. Thurmes. November Star-Galaxy Separation with a Neural Network. II. Multiple Schmidt Plate Fields. Technical report, 1993.
- [157] E. A. Owens, R. E. Griffiths, and K. U. Ratnatunga. Using oblique decision trees for the morphological classification of galaxies. Monthly Notices of the Royal Astronomical Society, 1996. ISSN 00358711. doi: 10.1093/mnras/281.1.153.
- [158] J. Pasquet, E. Bertin, M. Treyer, S. Arnouts, and D. Fouchez. Photometric redshifts from SDSS images using a convolutional neural network. Astronomy & Astrophysics, 621:A26, 2019.
- [159] G. Paturel, P. Fouque, L. Bottinelli, and L. Gouguenheim. An extragalactic database. I-The Catalogue of Principal Galaxies. Astronomy and Astrophysics Supplement Series, 80:299–315, 1989.
- [160] M. Pérez-Carrasco, G. Cabrera-Vives, M. Martínez-Marin, P. Cerulo, R. Demarco, P. Protopapas, J. Godoy, and M. Huertas-Company. Multiband Galaxy Morpholo-

- gies for CLASH: A Convolutional Neural Network Transferred from CANDELS. 131 (1004):108002, 10 2019. doi: 10.1088/1538-3873/aaeeb4.
- [161] K. L. Polsterer, F. Gieseke, and O. Kramer. Galaxy Classification without Feature Extraction The Task: Classifying Galaxies. Technical report, 2011. URL <http://www.astro.rub.de/polsterer/ADASS2011b.pdf>.
- [162] X. S. Poma, E. Riba, and A. Sappa. Dense extreme inception network: Towards a robust cnn model for edge detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 1923–1932, 2020.
- [163] D. J. Prole, J. I. Davies, O. C. Keenan, and L. J. M. Davies. Automated detection of very low surface brightness galaxies in the Virgo cluster. Monthly Notices of the Royal Astronomical Society, 478(1):667–681, 2018.
- [164] T. Prusti, J. H. J. De Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, M. Biermann, D. W. Evans, L. Eyer, and others. The gaia mission. Astronomy & astrophysics, 595:A1, 2016.
- [165] T. M. Quan, D. G. C. Hildebrand, and W.-K. Jeong. FusionNet: A Deep Fully Residual Convolutional Neural Network for Image Segmentation in Connectomics. Frontiers in Computer Science, 3:34, 2021. ISSN 2624-9898. doi: 10.3389/fcomp.2021.613981. URL <https://www.frontiersin.org/article/10.3389/fcomp.2021.613981>.
- [166] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788, 2016.
- [167] R. Reed and R. J. MarksII. Neural smithing: supervised learning in feedforward artificial neural networks. Mit Press, 1999.
- [168] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Technical report. URL <http://image-net.org/challenges/LSVRC/2015/results>.
- [169] A. S. G. Robotham, L. J. M. Davies, S. P. Driver, S. Koushan, D. S. Taranu, S. Casura, and J. Liske. Profound: source extraction and application to modern survey data. Monthly Notices of the Royal Astronomical Society, 476(3):3137–3159, 2018.

- [170] F. Rodrigues and F. Pereira. Deep learning from crowds. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.
- [171] J. Román, I. Trujillo, and M. Montes. Galactic cirri in deep optical imaging. Astronomy & Astrophysics, 644:A42, 2020.
- [172] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241, 2015.
- [173] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. International journal of computer vision, 77(1-3):157–173, 2008.
- [174] M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland. A Generative Model for Image Segmentation Based on Label Fusion. IEEE Transactions on Medical Imaging, 29(10):1714–1729, 2010. doi: 10.1109/TMI.2010.2050897.
- [175] C. Sandin. The influence of diffuse scattered light-I. The PSF and its role in observations of the edge-on galaxy NGC 5907. Astronomy & Astrophysics, 567:A97, 2014.
- [176] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry. How does batch normalization help optimization? Advances in neural information processing systems, 31, 2018.
- [177] S. Scardapane, S. Van Vaerenbergh, A. Hussain, and A. Uncini. Complex-valued neural networks with nonparametric activation functions. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018.
- [178] B. Sekachev, N. Manovich, M. Zhiltsov, A. Zhavoronkov, D. Kalinin, B. Hoff, T. Osmanov, D. Kruchinin, A. Zankevich, DmitriySidnev, M. Markelov, Johannes222, M. Chenuet, a-andre, telenachos, A. Melnikov, J. Kim, L. Ilouz, N. Glazov, Priya4607, R. Tehrani, S. Jeong, V. Skubriev, S. Yonekura, v. truong, zliang7, lizhming, and T. Truong. opencv/cvat: v1.1.0, 8 2020. URL <https://doi.org/10.5281/zenodo.4009388>.
- [179] L. Sifre and S. Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1233–1240, 2013.
- [180] P. Y. Simard, D. Steinkraus, J. C. Platt, and others. Best practices for convolutional neural networks applied to visual document analysis. In Icdar, volume 3, 2003.

- [181] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In International Conference on Learning Representations, 2015.
- [182] A. Sinha and J. Dolz. Multi-scale self-guided attention for medical image segmentation. IEEE journal of biomedical and health informatics, 25(1):121–130, 2020.
- [183] C. T. Slater, P. Harding, and J. C. Mihos. Removing internal reflections from deep imaging data sets. Publications of the Astronomical Society of the Pacific, 121(885):1267, 2009.
- [184] K. Sohn and H. Lee. Learning Invariant Representations with Local Transformations. In Proceedings of the 29th International Conference on International Conference on Machine Learning, ICML’12, pages 1339–1346, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.
- [185] E. Sola, P.-A. Duc, F. Richards, A. Paiement, M. Urbano, J. Klehammer, M. Bílek, J.-C. Cuillandre, S. Gwyn, and A. McConnachie. Characterization of low surface brightness structures in annotated deep images. Astronomy & Astrophysics, 662:A124, 2022.
- [186] Q. Song, Z. Cui, and P. Liu. An Efficient Solution for Semantic Segmentation of Three Ground-based Cloud Datasets. Earth and Space Science, 7(4), 2020.
- [187] M. C. Stome-Lombardi, O. Lahav, L. Sodré, and L. J. Stome-Lombardi. Morphological classification of galaxies by Artificial Neural Networks. Technical report, 1992.
- [188] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9, 2015.
- [189] D. Tanoglidis, A. Čiprijanović, and A. Drlica-Wagner. DeepShadows: Separating low surface brightness galaxies from artifacts using deep learning. Astronomy and Computing, 35:100469, 2021.
- [190] D. Tanoglidis, A. Čiprijanović, A. Drlica-Wagner, B. Nord, M. H. L. S. Wang, A. J. Amsellem, K. Downey, S. Jenkins, D. Kafkes, and Z. Zhang. DeepGhostBusters: Using Mask R-CNN to Detect and Mask Ghosting and Scattered-Light Artifacts from Optical Survey Images. arXiv preprint arXiv:2109.08246, 2021.
- [191] A. Tao, K. Sapra, and B. Catanzaro. Hierarchical multi-scale attention for semantic segmentation. arXiv preprint arXiv:2005.10821, 2020.

- [192] P. Teeninga, U. Moschini, S. C. Trager, and M. H. F. Wilkinson. Improved detection of faint extended astronomical objects through statistical attribute filtering. In International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, pages 157–168, 2015.
- [193] The MSE Observatory. Maunakea Spectroscopic Explorer, 10 2014. URL <https://www.cfht.hawaii.edu/en/news/MSE-new/observatory/>.
- [194] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal. Deep Complex Networks. In Proceedings of the 6th International Conference on Learning Representations, 2018.
- [195] I. Trujillo and J. Fliri. Beyond 31 mag arcsec⁻²: the frontier of low surface brightness imaging with the largest optical telescopes. The Astrophysical Journal, 823(2):123, 2016.
- [196] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 9446–9454, 2018.
- [197] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, pages 5998–6008, 2017.
- [198] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis & Machine Intelligence, 13(06):583–598, 1991.
- [199] T. Von Hippel, L. J. Storrie-Lombardi, M. C. Storrie-Lombardi, and M. J. Irwin. Automated classification of stellar spectra–I. Initial results with artificial neural networks. Monthly Notices of the Royal Astronomical Society, 269(1):97–104, 1994.
- [200] M. Walmsley, A. M. N. Ferguson, R. G. Mann, and C. J. Lintott. Identification of low surface brightness tidal features in galaxies using convolutional neural networks. Monthly Notices of the Royal Astronomical Society, 483(3):2968–2982, 2019.
- [201] M. Walmsley, L. Smith, C. Lintott, Y. Gal, S. Bamford, H. Dickinson, L. Fortson, S. Kruk, K. Masters, C. Scarlata, and others. Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning. Monthly Notices of the Royal Astronomical Society, 491(2):1554–1574, 2020.

- [202] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich. Multi-atlas segmentation with joint label fusion. IEEE transactions on pattern analysis and machine intelligence, 35(3):611–623, 2012.
- [203] X. Wang, B. Zhang, C. Li, R. Ji, J. Han, X. Cao, and J. Liu. Modulated convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 840–848, 2018.
- [204] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging, 23(7):903–921, 2004. doi: 10.1109/TMI.2004.828354.
- [205] M. Weiler and G. Cesa. General E (2)-Equivariant Steerable CNNs. In Advances in Neural Information Processing Systems, pages 14334–14345, 2019.
- [206] M. Weiler, F. A. Hamprecht, and M. Storath. Learning Steerable Filters for Rotation Equivariant CNNs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 849–858, 2018. ISBN 1711.07289v3.
- [207] N. Weir, U. M. Fayyad, and S. Djorgovski. AUTOMATED STAR/GALAXY CLASSIFICATION FOR DIGITIZED POSS-II. Technical Report 9, 1995.
- [208] S. D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. Monthly Notices of the Royal Astronomical Society, 183:341–358, 5 1978. doi: 10.1093/mnras/183.3.341.
- [209] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pages 3–19, 2018.
- [210] D. Worrall and M. Welling. Deep scale-spaces: Equivariance over scale. In Advances in Neural Information Processing Systems, pages 7364–7376, 2019.
- [211] D. E. Worrall, S. J. Garbin, D. Turmukhambetov, and G. J. Brostow. Harmonic Networks: Deep Translation and Rotation Equivariance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5028–5037, 2017.
- [212] J. Wu, Y. Yu, C. Huang, and K. Yu. Deep multiple instance learning for image classification and auto-annotation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3460–3469, 2015.

- [213] S. Xie and Z. Tu. Holistically-nested edge detection. In Proceedings of the IEEE international conference on computer vision, pages 1395–1403, 2015.
- [214] Y. Yan, R. Rosales, G. Fung, R. Subramanian, and J. Dy. Learning from multiple annotators with varying expertise. Machine learning, 95(3):291–327, 2014.
- [215] F. Yu and V. Koltun. Multi-Scale Context Aggregation by Dilated Convolutions. In International Conference on Learning Representations, 2016. ISBN 0894-0282. doi: 10.16373/j.cnki.ahr.150049.
- [216] Y. Yu, Z. Ji, Y. Fu, J. Guo, Y. Pang, and Z. Zhang. Stacked Semantics-Guided Attention Model for Fine-Grained Zero-Shot Learning. In Advances in Neural Information Processing Systems, 2018.
- [217] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In European conference on computer vision, pages 818–833, 2014.
- [218] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2881–2890, 2017.
- [219] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. C. Loy, D. Lin, and J. Jia. Pscanet: Point-wise spatial attention network for scene parsing. In Proceedings of the European Conference on Computer Vision (ECCV), pages 267–283, 2018.
- [220] Y. Zhou, Q. Ye, Q. Qiu, and J. Jiao. Oriented response networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 519–528, 2017.
- [221] F. Zwicky. The redshift of extragalactic nebulae. Helvetica Physica Acta, 6:110–127, 1933.

Appendix A

Appendices

A.1 Annotation Database Design

The users relation stores records of all users registered on the annotation tool, detailed in Table A.1. While there exists four attributes containing non-functional additional information for registered users, other attributes serve some purpose for managing users within the annotation tool. The exact functionalities of these attributes will be discussed in the following section.

Attribute	Description
u_id	The user's primary key identification.
username	The user's chosen login username.
email	The user's email address.
firstname	The user's first name.
lastname	The user's last name.
institution	The user's associated institution.
password_hash	A hashing of the user's chosen login password.
advanced	The privilege level granted to the user, ranging from 0 to 3.

Table A.1: Descriptions of attributes belonging to the shapes relation. Primary key is emboldened.

The galaxies relation stores galaxy information relevant to the annotation tool's functionality, detailed in Table A.2. Viewing specific information is held in the survey, bands, fov and active attributes. Galaxy specific metadata is held in the name, ra and dec attributes.

The annotations relation stores records of submitted annotations, detailed in Table A.3. As each annotation is of a galaxy and created by a user, foreign key dependencies are used between the users and galaxies relations and the annotations relation. This

Attribute	Description
g_id	The primary key identification the galaxy record.
name	The galaxy's object identification, e.g. NGC1121.
survey	The default survey to be displayed on loading the viewing tool.
bands	The default available bands available in the viewing tool.
ra	The galaxy's right ascension.
dec	The galaxy's declination.
fov	The initial field of view to be used in the viewing tool.
active	Whether the galaxy is to be selected for annotation by any users.

Table A.2: Descriptions of attributes belonging to the galaxies relation. Primary key is emboldened.

dependency is one to many as a user can submit multiple annotations, and a galaxy can be annotated multiple times. The exact time the annotation is submitted to the central server is recorded in the timestamp attribute.

Attribute	Description
a_id	The annotation's primary key identification.
<i>g_id</i>	The galaxy ID annotated.
<i>u_id</i>	The user ID which submitted the annotation.
timestamp	The exact time the annotation was submitted to the central server.

Table A.3: Descriptions of attributes belonging to the annotations relation. Primary key is emboldened, foreign keys are italicised.

Information concerning each drawn shape is stored in the shapes relation, detailed in Table A.4. Details related to the purpose of the shape, such as the classification, are stored in the number, feature and note attributes. The exact spatial properties of the shape are stored in the shape, x_points, y_points, ra_points and dec_points attributes.

A.2 Annotation Tool User Management

The annotation tool uses a login system to verify user credentials, integrated with the database schema described in the previous section. Users are divided into two broad categories: expert and non-expert. Non-expert users do not require a password to login (see Figure A.1a), while experts do (see Figure A.1b). In the login interface there exists a hyperlink directing to a registration interface shown in A.1c, where visitors can register an account, thus creating a user record in the central database. New user accounts are set as non-experts.

Attribute	Description
s_id	The shape's primary key identification.
<i>a_id</i>	The annotation id which the shape belongs to.
shape	The type of shape.
number	The shape's identifying number during the drawing process.
feature	The feature or object classified by the shape.
note	The shape's note left by the user.
x_points	The x-axis pixel location of each of the shape's vertices.
y_points	The y-axis pixel location of each of the shape's vertices.
ra_points	The right ascension location of each of the shape's vertices.
dec_points	The declination location of each of the shape's vertices.

Table A.4: Descriptions of attributes belonging to the shapes relation. Primary key is emboldened, foreign keys are italicised.

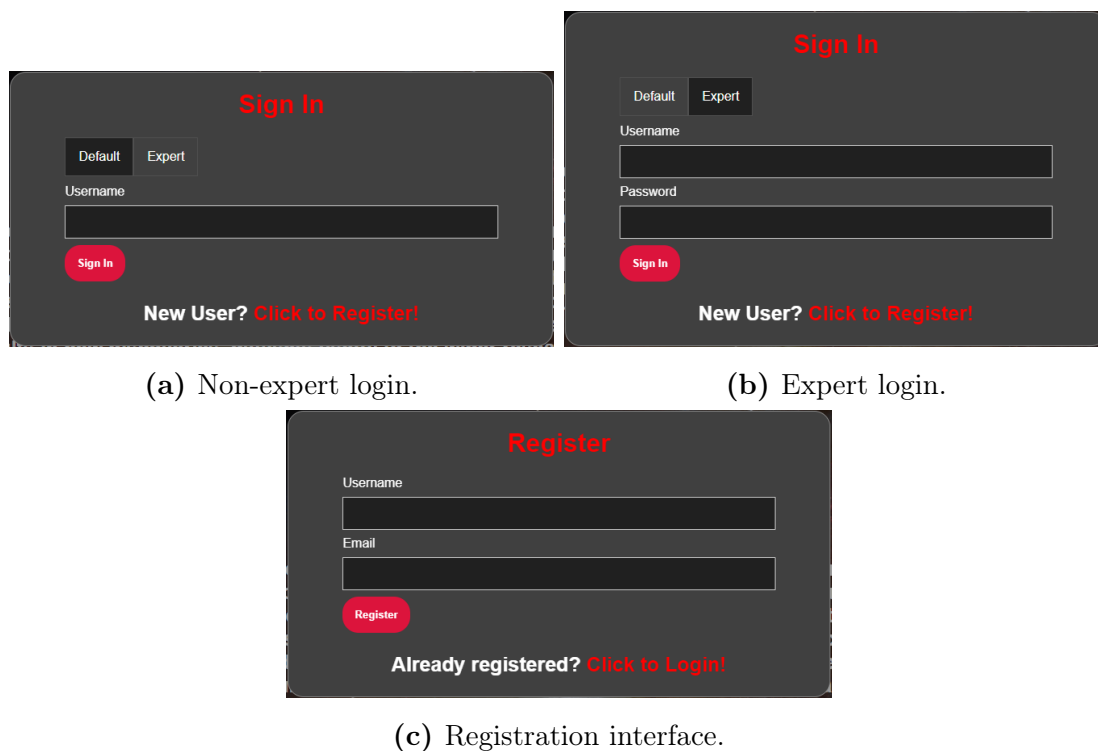


Figure A.1: Different interfaces for user credential verification and registration.

Users are further divided into sub categories based on privilege levels, granting them access to certain features. Privilege levels range from 0 to 3, with 0 as non-experts, and experts between 1 and 3. As previously mentioned, basic non-expert users are not required to set a password, but can submit annotations. Any user with a privilege level greater than 0 and who does not have a password is redirected to set one. This is effectively the only difference between users of privilege 0 and 1: users of privilege 1 are not granted

Link	Galaxy Name	RA	Dec	Timestamp
/verify/id/80	NGC2844	140.45	40.151	2020-07-06 14:22:46.789047
/verify/id/81	NGC4274	184.961	29.6145	2020-07-06 14:25:55.528713
/verify/id/82	NGC5389	209.026	59.7419	2020-07-06 14:33:31.233974
/verify/id/83	NGC5480	211.59	50.7248	2020-07-06 14:39:50.962559
/verify/id/85	IC3102	184.358	6.69	2020-07-06 15:00:54.117404
/verify/id/86	NGC0489	20.4746	9.2065	2020-07-06 15:07:37.606795
/verify/id/87	NGC0518	21.073	9.3308	2020-07-06 15:15:52.111676
/verify/id/88	NGC0522	21.1911	9.9945	2020-07-06 15:20:22.969205
/verify/id/89	NGC0532	21.3223	9.26	2020-07-06 15:26:47.868326

Showing 1 to 9 of 9 entries (filtered from 448 total entries)

Figure A.2: An interactive table showing the user’s annotations.

any tool related functionality aside a password protected account. All users are able to view and search through a list of their own annotations, as shown in Figure A.2, where they can revisit their annotation and submit edits. Finally, users of privilege 2 and 3 are able to view and edit annotations made by themselves and other users, enabling the verification of other user’s annotations.

The annotation tool provides functionality to manage users, so that changes to user records do not need to be made directly to the database. The ability to manage users through the website is provided to users of privilege 2 and 3. Through this interface, the logged in user or current user can lookup other users by searching for their username, displayed in Figure A.3a. To manage another user, the current user must have a higher privilege level than the searched for user, otherwise the search will not return any result. Once a user has been found, the current user can reset the user’s password and increase their privilege level to at most their own privilege level (see Figure A.3b).

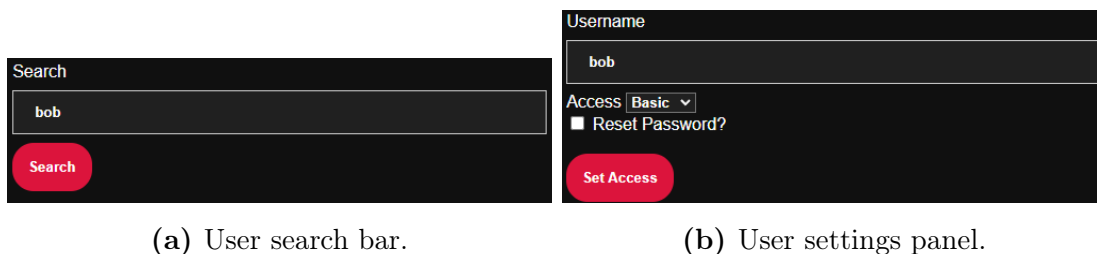


Figure A.3: Elements that allow the management of other users by a user of high enough privilege level.