Multi-Scale WSI Analysis: A Cascade Framework for Efficient Breast Cancer Metastasis Detection

Connor Atkins[1]*[0009−0004−8814−8241], Gary K.L. Tam[1][0000−0001−7387−5180],
Michael Edwards[2][0000−0003−3367−969X], Muhammad Aslam[3], and Jiaxiang
Zhang[1][0000−0002−4758−0394]

[1] Department of Computer Science, Swansea University, Wales {connor.atkins, k.l.tam, jiaxiang.zhang}@swansea.ac.uk
[2] Department of Medicine Health and Science, Swansea University, Wales michael.edwards@swansea.ac.uk
[3] Department of Cellular Pathology, Glan Clwyd Hospital, Betsi Cadwaladr University Health Board, Wales muhammad.aslam3@wales.nhs.uk

Abstract. Analysing whole slide images in digital pathology for disease detection and diagnosis is a challenge, as it requires balancing fine-grained details with broader tissue context. High-resolution images offer detailed information but often result in slow processing times, while lower-resolution images capture larger contextual areas at the cost of missing critical details. This study explores the research question of how to effectively balance these needs by proposing a cascade framework that integrates multiple resolution levels to optimize both accuracy and computational efficiency in detecting breast cancer metastasis using the CAMELYON16 dataset. Surprisingly, intermediate-resolution levels (10× magnification) outperformed the highest resolution (40×), challenging conventional assumptions. Expanding the field-of-view during inference improved performance universally across all resolution levels without retraining. Our cascade pipeline selectively applies high-resolution analysis to regions flagged at lower resolutions. The optimal configuration, combining 5× screening with targeted 20× analysis, achieved a 0.661 FROC score, surpassing single-resolution models by 4.4% and reducing inference time by 12.4%. These findings suggest that strategic multi-resolution approaches can enhance both accuracy and efficiency in computational pathology, potentially accelerating clinical diagnoses without compromising detection reliability.

Keywords: Histopathology · Digital Pathology · Whole Slide Image (WSI) · Deep Learning · Segmentation.

1 Introduction

Whole Slide Imaging (WSI) is transforming pathology by enabling the digital analysis of tissue samples for cancer diagnosis. Traditionally, pathologists perform manual slide examination, a time-consuming and error-prone process that

[*] Corresponding author: connor.atkins@swansea.ac.uk

can impact patient outcomes. As accurate cancer detection plays a crucial role in treatment, automating and improving diagnostic workflows is essential.

Deep learning has demonstrated significant promise in WSI analysis. For example, the CAMELYON16 challenge [12] highlighted the potential of patch-based convolutional neural networks, using GoogLeNet [17] and heatmap-based post-processing [19], achieving high performance and reducing human error by 85%. Several methods have since built on this success, with innovations such as Deep Multi-Magnification Networks [8], HookNet [14], and MAMC-Net [22], all leveraging multi-resolution and attention mechanisms. Techniques like ensemble methods [9] and hierarchical frameworks [20] further aim to balance computational efficiency with accuracy. More recent advances in WSI analysis have explored various architectural innovations beyond traditional CNNs. TransUNet [3] and SwinUNet [2] leverage transformer architectures to capture long-range dependencies, while the top CAMELYON16 challenge submissions achieved FROC scores up to 0.807 using ensemble methods and sophisticated post-processing [19]. Guo et al. [7] investigated the impact of patch size on segmentation performance, finding that larger contexts improve detection accuracy. End-to-end cascade frameworks, such as CSC-Net [16], jointly optimize usage of multiple resolution levels but require complex training procedures.

However, challenges remain in current methods for WSI analysis. First, there is a trade-off between resolution and field-of-view: high-resolution patches capture fine details but lack broader spatial context, while low-resolution patches provide spatial context but can miss critical diagnostic features. Second, methods like MAMC-Net [22] and ensemble approaches [9] introduce computational overhead, limiting their practical use in clinical settings. Third, multi-resolution methods process all regions at all resolutions, leading to redundant computation, especially for areas where low-resolution analysis suffices, such as background tissue. Finally, many methods lack the flexibility to adapt computational efforts based on region complexity or uncertainty, applying uniform processing across the entire WSI. These lead us to reconsider how to optimize the use of resolution levels while maintaining accuracy.

The above observations prompt us to ask the following questions: **RQ1:** *Can we selectively use multiple resolution levels for tissue classification without the computational burden of full multi-resolution fusion?* We hypothesize that leveraging different resolution at certain levels can provide complementary information, allowing for more efficient use of resources. **RQ2:** *Can expanding the field-of-view during inference improve contextual awareness without retraining the model?* We hypothesize that increasing the receptive field during inference can improve contextual awareness while preserving fine details, without needing to retrain the model. **RQ3:** *Can a cascade architecture, mimicking the diagnostic workflow of pathologists, achieve computational efficiency without sacrificing accuracy?* While existing methods have selectively applied high-resolution analysis to regions of interest [4, 20], we hypothesize that a more targeted cascade approach, focusing analysis only on regions identified at lower resolutions, can optimize computational efficiency even further and align more closely with pathologists' diagnostic workflows.

Our cascade approach is inspired by the established diagnostic workflow of pathologists, who initially examine whole slide images at low magnification to identify regions of interest, followed by closer inspection of suspicious areas at higher magnification [6, 18]. This hierarchical strategy has evolved to balance the trade-off between comprehensive slide coverage and time efficiency—a challenge that is equally pertinent to computational pathology. By emulating this clinically validated methodology, our approach seeks to achieve comparable efficiency gains while preserving diagnostic accuracy.

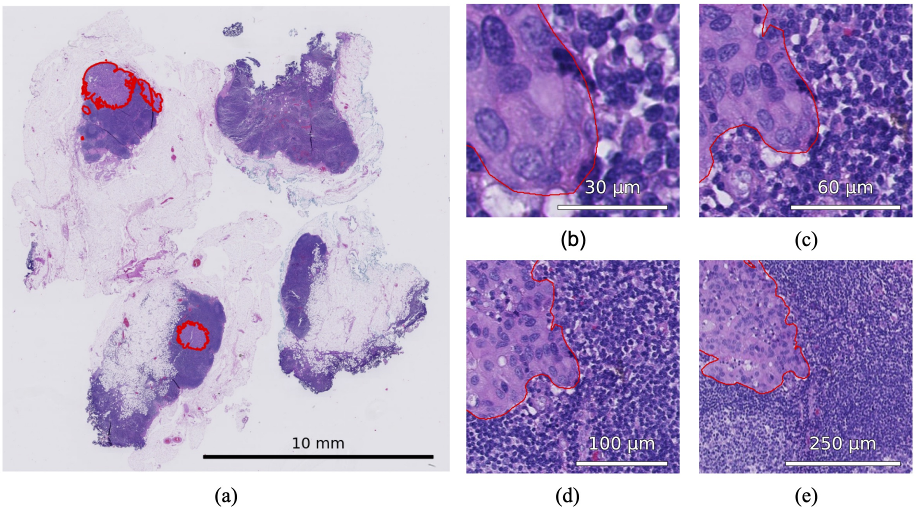
The key contributions of our work include:

– A systematic evaluation of model performance across multiple resolution levels, challenging the assumption that the highest resolution is always optimal for histopathological assessment.
– An investigation into expanding the field-of-view during inference, demonstrating a computationally efficient method for incorporating broader context.
– A computationally efficient cascade architecture that mimics pathologists' diagnostic workflow by selectively applying high-resolution analysis to regions of interest.
– Empirical results showing potential for comparable or superior performance when applied to existing methods while reducing computational requirements, enabling more practical deployment in clinical settings.

2 Data and Preprocessing

This study utilizes the CAMELYON16 dataset [12], which includes a comprehensive collection of 399 whole slide images (WSIs) of lymph node sections from breast cancer patients. The dataset is divided into a training set (n=270) and a test set (n=129). The training set consists of 160 normal slides and 110 slides containing metastatic regions, whilst the test set mirrors this distribution, with 80 normal slides and 49 slides with metastatic regions. Each WSI is stored in a pyramidal .tif format, incorporating a hierarchical resolution structure, with the highest resolution corresponding to 40× magnification (0.25 μm per pixel, Figure 1 (b)). At this magnification, individual WSIs typically span around 100,000 × 100,000 pixels, which presents significant computational challenges for direct processing and requires careful consideration of efficient data handling strategies.

The preprocessing pipeline begins with tissue mask generation to focus computation for downstream tasks on diagnostically relevant regions while excluding background slide area and artifacts (Figure 2). A low-resolution representation (Level 6) from the multi-resolution pyramid [12] is extracted to optimize efficiency while preserving structural integrity [9]. To prevent segmentation errors from scanning artifacts, black pixel regions along slide boundaries are converted to white. Median filtering with a 7×7 kernel reduces noise while preserving tissue



(a) (b) (c) (d) (e)

Fig. 1. Example slide with ground truth overlay (a) and 256x256 patches at 40x (b), 20x (c), 10x (d), and 5x (e) magnification. The horizontal bar indicates the actual length, measured in either mm or μm.

edges, and the RGB color space is transformed to HSV, leveraging the saturation channel for robust tissue-background separation. Otsu's adaptive thresholding [13] is applied to segment tissue regions, followed by morphological dilation with a 3×3 structuring element to enhance continuity. The resulting tissue masks efficiently guide patch extraction, ensuring computation is focused on diagnostically relevant areas while maintaining scalability for high-resolution WSIs.

3 Methods

This study evaluates the efficacy of multi-resolution approaches for WSI analysis in computational pathology through three key experiments. First, we assess the impact of image resolution on model performance in cancer metastasis detection (**RQ1**) by independently training models at multiple resolution levels (Section 3.1). Second, we investigate whether expanding the field-of-view during inference enhances performance without retraining (**RQ2**) by evaluating enlarged inference patches (Section 3.2). Finally, we explore the combination of multiple resolution levels in a cascade framework to optimize accuracy and computational efficiency (**RQ3**) through the development of a cascade inference pipeline (Section 3.3).

Our approach employs the U-Net architecture [15] as an effective baseline due to its demonstrated success in medical image segmentation tasks. While more complex architectures exist, U-Net provides a balanced trade-off between computational efficiency and segmentation accuracy—a critical consideration for the resource-intensive nature of WSI analysis. The architecture's skip connections

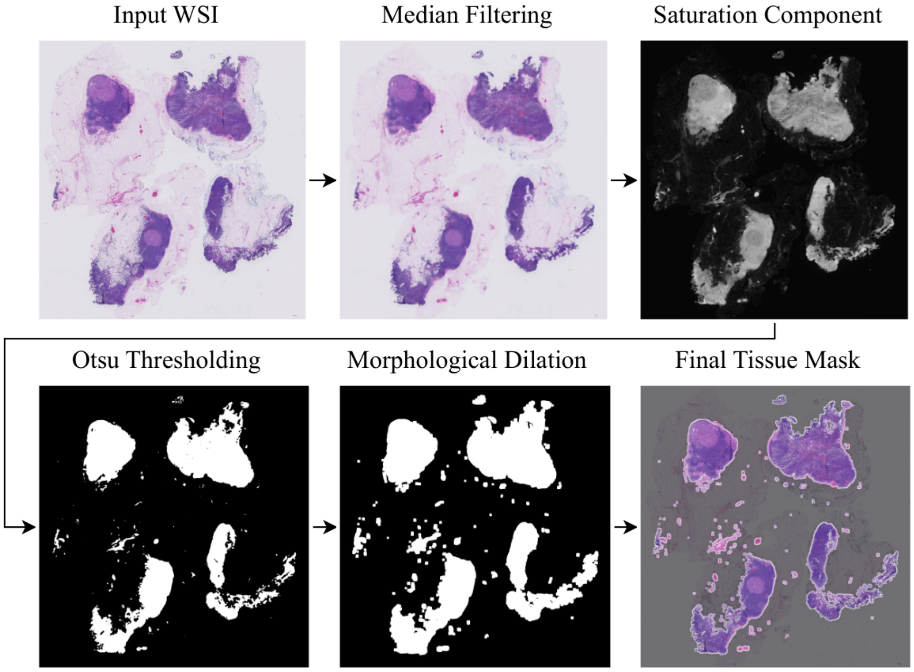


Fig. 2. The preprocessing steps of generating a tissue mask from an WSI image.

are particularly valuable for histopathological analysis, as they preserve fine spatial details while maintaining broader contextual awareness, both of which are essential for accurate metastasis detection.

3.1 Multi-Resolution Model Training

The hierarchical structure of WSIs allows analysis at multiple magnification levels, balancing computational efficiency with visual context (**RQ1**). Our approach systematically evaluates these trade-offs by training four independent U-Net models, each specialized for a specific magnification level. Importantly, we maintain identical U-Net architecture configurations across all models—using the standard architecture as proposed by Ronneberger et al. [15] with four downsampling and upsampling levels—and vary only the resolution of the input training data. Specifically, we train four separate U-Net models for patches extracted from each of four magnification levels available in the WSI pyramid: 40×, 20×, 10×, and 5× magnification. Each model is initialized with the same random weights and trained independently on 256×256 pixel patches extracted exclusively from its corresponding magnification level. This controlled experimental design enables direct comparison of model performance across resolution levels while isolating magnification as the sole variable.

For all models, patch sampling is stratified across slides to ensure representative coverage of the tissue distribution. The sampling strategy implements careful balancing mechanisms to address the inherent class imbalance in histopathological data, maintaining consistent positive-to-negative patch ratios across all resolution levels.

3.2 Increased Inference FOV

To examine the impact of contextual information during inference (**RQ2**), we utilized a key feature of fully convolutional network architectures like U-Net: their capacity to process input images of arbitrary dimensions without modifying the network architecture. This enables us to assess the effect of an expanded field-of-view without the computational cost of retraining models [9]. Our approach involves testing models (originally trained on 256×256 pixel patches) with larger 512×512 pixel patches during inference. Since U-Net applies convolution operations with the same weights regardless of input size, this technique effectively doubles the contextual field of view without requiring architectural modifications or retraining. The convolutional nature of the model handles the increased spatial dimensions while maintaining feature detection capabilities.

This strategy enables direct comparison of the contextual effects on prediction quality, isolating the impact of patch size from other variables. While computationally more expensive, it provides valuable insights into how expanded spatial context influences detection performance.

3.3 Cascade Pipeline

Here, we investigate whether integrating multiple resolution levels within a single inference pipeline can enhance both accuracy and computational efficiency (**RQ3**). We hypothesize that an adaptive multi-resolution approach—leveraging lower resolution for initial screening and higher resolution for regions of interest—can achieve comparable performance to exhaustive high-resolution analysis while significantly reducing computational costs.

This approach mirrors the diagnostic workflow of pathologists, who first scan tissue at low magnification before closely examining suspicious regions at higher magnification. To evaluate this, we designed a hierarchical analysis pipeline (Figure 3) that sequentially applies models trained at different resolutions, with lower-resolution predictions guiding high-resolution analysis. The pipeline consists of several integrated components working in sequence:

Initial Low-Resolution Analysis The first phase employs a low-resolution model to analyse tissue regions identified by the preprocessing tissue mask. Non-overlapping patches are extracted across the entire slide to generate an initial probability map, which guides high-resolution analysis. The chosen resolution balances efficiency with reliable identification of regions of interest.

Region Selection Mechanism A two-parameter thresholding approach, defined by percentile value p and threshold t, determines which patches require further analysis. For each patch, the pth percentile of the pixel-level tumour

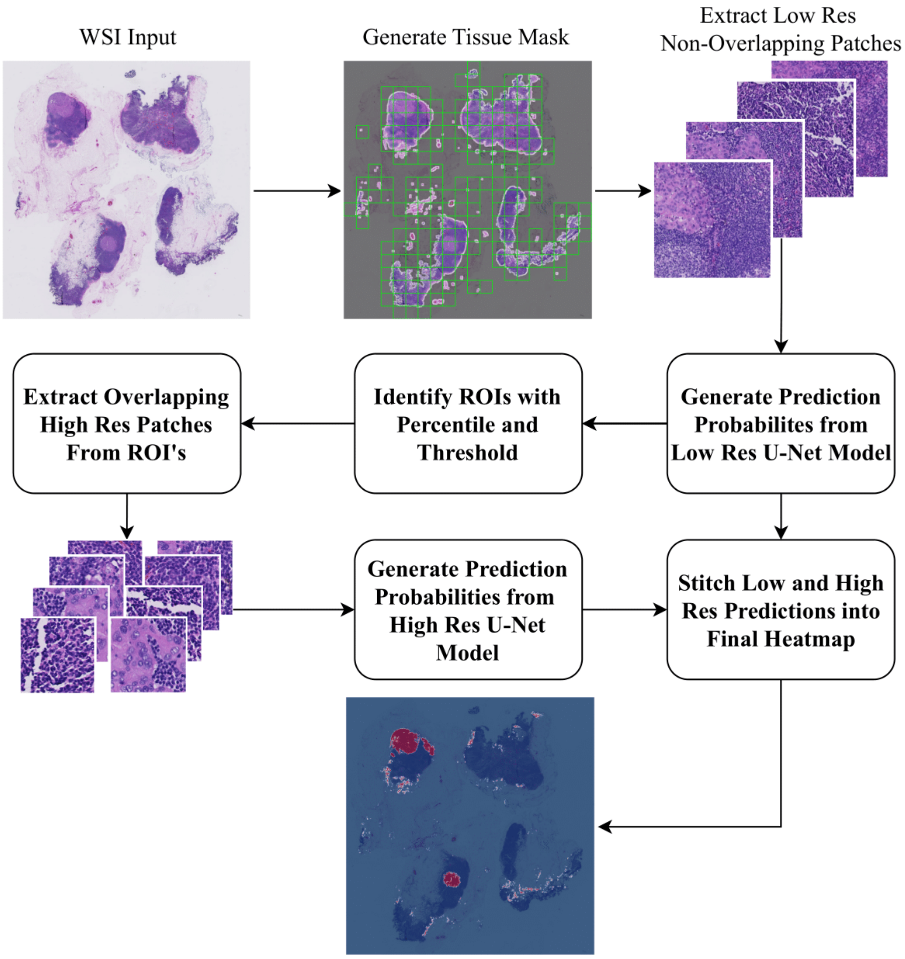

WSI Input
Generate Tissue Mask
Extract Low Res Non-Overlapping Patches
Generate Prediction Probabilites from Low Res U-Net Model
Identify ROIs with Percentile and Threshold
Extract Overlapping High Res Patches From ROI's
Generate Prediction Probabilities from High Res U-Net Model
Stitch Low and High Res Predictions into Final Heatmap

Fig. 3. Overview of the steps taken in the cascade inference pipeline.

probability (from the U-Net model's output) is computed. If it exceeds t, the corresponding region is selected for high-resolution analysis. This mechanism efficiently isolates regions with high tumour probability while minimizing unnecessary computations. Comprehensive ablation studies were conducted to optimize these parameters for accuracy and efficiency.

High-Resolution Analysis Selected regions undergo high-resolution analysis with overlapping patch extraction (50% overlap) to ensure smooth probability maps and reduce boundary artifacts. The examination area extends beyond the initially identified regions by incorporating a border, ensuring comprehensive coverage of potential metastatic areas.

Final Probability Map Generation Spatial averaging is applied at overlapping regions to create smoother probability transitions and minimize chequerboard artifacts. Regions not selected for high-resolution analysis retain their low-resolution predictions, ensuring a seamless integration of all resolution levels into the final whole-slide probability map.

The pipeline supports experimentation with various resolution level combinations, enabling systematic evaluation of high-low resolution pairs. This flexibility allows empirical optimization of the cascade structure through ablation studies, assessing the impact of different resolution settings, patch sizes, and selection criteria on detection performance and computational efficiency.

4 Experiments

4.1 Implementation and Model Training

The CAMELYON16 dataset was initially divided into predefined training and testing sets as described in the Data section. To facilitate model development and hyperparameter tuning, the training set was further split into training (80%) and validation (20%) subsets at the slide level, with stratification applied to maintain consistent class distribution across partitions and prevent data leakage.

Patches were extracted using a balanced sampling strategy (see Section 3), resulting in 1,024,000 patches with equal representation of positive and negative examples. During training, dynamic augmentation was applied on-the-fly, including random rotations (0°, 90°, 180°, 270°) and random flips, to enhance model generalization.

The framework was implemented in PyTorch, and models were trained on an NVIDIA GeForce RTX 4090 GPU. The Adam optimizer [10] was used with an initial learning rate of 0.001. A learning rate scheduler reduced the rate by a factor of 0.5 after 5 epochs with no improvement, while early stopping with a patience of 15 epochs helped prevent overfitting.

Each model was trained for up to 100 epochs with a batch size of 64, randomly sampling 6400 patches per epoch to ensure balance between positive and negative examples. The Asymmetric Unified Focal++ loss function [21], with parameters $\delta = 0.8$, $\gamma = 0.5$, and $\gamma^+ = 2$, was used to address class imbalance, providing stronger penalties for misclassifying metastatic regions while maintaining efficient learning for non-metastatic regions.

4.2 Evaluation Metrics

Model performance was assessed using multiple metrics. The primary metric was the Free Response Operating Characteristic (FROC) curve [5], computed using the evaluation code from the CAMELYON16 dataset [12]. The FROC curve plots the true positive rate against the average number of false positives per image, with the FROC value averaged across six predefined false positive rates (0.25, 0.5, 1, 2, 4, 8 per slide). Additionally, the Dice coefficient was calculated to measure spatial overlap between predicted and ground truth segmentations, accounting for both false positives and negatives. For computational efficiency, we measured average inference time per slide and the number of patches processed, enabling comparison of the computational demands and efficiency gains of the cascade approach.

4.3 Cascade Hyperparameter Search

The cascade pipeline architecture required extensive experimentation to identify the optimal combination of resolution levels, patch sizes, and decision thresholds. We evaluated three resolution level combinations: (20×, 10×), (20×, 5×), and (10×, 5×). For each combination, we further assessed four patch size configurations: both levels using 256×256 pixels, both levels using 512×512 pixels, low resolution using 256×256 and high resolution using 512×512, and the inverse configuration.

For each configuration, we conducted a grid search over the percentile parameter p (ranging from 70 to 95 in increments of 5) and the threshold parameter t (ranging from 0.1 to 0.9 in increments of 0.05), resulting in 216 distinct cascade configurations per resolution and patch size combination.

Initial screening was performed on a balanced subset of the test data to identify promising candidates, and the best configurations were further evaluated on the full test set to determine the optimal cascade pipeline configuration.

5 Results

5.1 Individual Resolution Model Performance

The experimental evaluation commenced with an assessment of individual resolution-level models to establish baseline performance metrics. Table 1 summarizes the quantitative results for models trained at each of the four resolution levels (0-3) using the standard 256×256 pixel patch size, as well as the effect of increasing FOV via larger patch size at test time. Figure 4 shows qualitatively how each of the four resolution levels are affected by the spatial context, cellular structure trade off and the effect that has on the model.

Table 1. Performance of single-resolution models, and the effect of patch size at inference on model performance.

Resolution Level	Patch Size	FROC Score	Dice	Inference Time (s)
Level 0 (40x)	256×256	0.159	0.202	479.3
Level 1 (20x)	256×256	0.590	0.505	176.5
	512×512	0.625	0.512	297.0
Level 2 (10x)	256×256	0.607	0.556	48.6
	512×512	**0.633**	**0.565**	87.3
Level 3 (5x)	256×256	0.519	0.533	14.7
	512×512	0.541	0.543	29.8

Contrary to the common assumption that higher resolution levels lead to better performance [1,5], our experimental results show that Level 2 (10× magnification) achieves the highest performance across both primary evaluation metrics. This suggests that an intermediate resolution strikes the optimal balance between cellular detail and contextual information for metastasis detection in the CAMELYON16 dataset. The significant performance gap between Level 0 and the other resolution levels—particularly the 74% reduction in the FROC score compared to Level 2—indicates that the highest resolution may capture excessive detail at the expense of the broader contextual information necessary for accurate detection. This finding is especially surprising considering that much of the prior work [5,7,11,19] on this dataset has exclusively used the highest resolution level, and it suggests that overly high magnification might introduce noise and variability, potentially impairing model generalization.

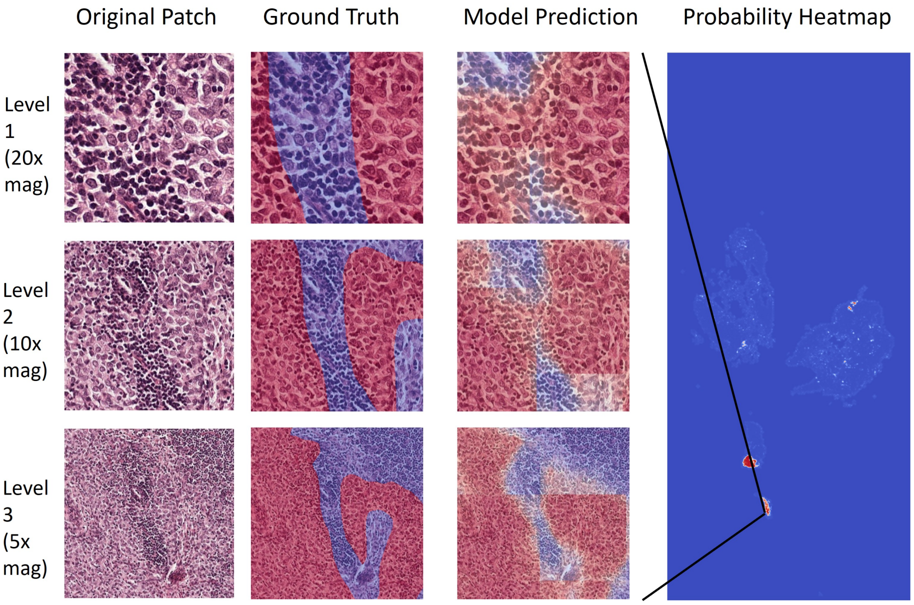


Fig. 4. Qualitative comparison of effect the various resolution levels have on the model output.

5.2 Impact of Increased FOV at Inference

To investigate the influence of contextual information on model performance, we evaluated each resolution-level model using an enlarged patch size of 512×512 pixels during inference without retraining the models using this increased patch size. Table 1 presents a comparative analysis of performance metrics between standard and enlarged patch sizes across resolution levels.

Table 2. Performance of cascade inference pipeline.

Cascade	PS Configuration	p & t	FROC Score	Dice	Inference Time (s)
Level 2→1	256→256	0.95 & 0.5	0.602	0.524	62.7
	256→512	0.95 & 0.5	0.651	0.536	74.1
Level 3→1	256→256	0.95 & 0.15	0.618	0.534	64.9
	256→512	0.95 & 0.15	**0.661**	**0.545**	76.5

The results demonstrate consistent performance improvements across all resolution levels when using the enlarged patch size, with increases in FROC scores ranging from 4.2% to 5.9% and more modest gains in Dice coefficients. This uniform improvement suggests that increased contextual information contributes positively to metastasis detection accuracy regardless of resolution level. The most substantial improvement was observed at Level 2, where the FROC score increased from 0.607 to 0.633, further solidifying this resolution level's effectiveness.

However, these performance gains come at a significant computational cost, with inference time increasing by approximately 68-103% across the different resolution levels. This trade-off between performance and computational efficiency motivates the exploration of cascade approaches that could potentially capture the benefits of enlarged patches while maintaining reasonable computational demands.

5.3 Cascade Pipeline Evaluation

To gain deeper insights into the cascade pipeline's behaviour, we conducted comprehensive ablation studies on the percentile threshold (p) and confidence threshold (t) parameters across different resolution level combinations. The ablation studies revealed that the optimal parameter values varied depending on the specific resolution levels employed in the cascade. Generally, higher percentile values (90-95) combined with relatively low threshold values (0.10-0.20) yielded the best performance across configurations. This pattern suggests that examining any regions that had even a low confidence of relevance at higher resolution levels is sufficient to capture most true positive findings while increasing computational efficiency.

Further analysis of patch size combinations revealed that utilizing larger patch sizes (512×512) at higher resolution levels provided greater performance benefits than at lower resolution levels. This observation aligns with the intuition that increasing the amount contextual information is particularly valuable when examining detailed cellular structures at higher magnifications, while lower magnifications inherently capture broader context even with smaller patch sizes.

The cascade pipeline approach was designed to leverage the complementary strengths of different resolution levels while optimizing computational efficiency. We systematically evaluated various resolution level combinations, patch size configurations, and selection parameter values. Table 2 presents the performance of the top-performing cascade configurations compared to the best single-resolution models.

The optimal cascade configuration utilized Level 3 (5× magnification) with 256×256 pixel patches for initial screening, followed by Level 1 (20× magnification) with 512×512 pixel patches for detailed analysis of regions of interest. This configuration employed a percentile value (p) of 95 and a threshold value (t) of 0.15 for region selection, achieving an FROC score of 0.661. This represents a 4.4% improvement over the best single-resolution model (Level 2 with 512×512 pixel patches) while reducing inference time by 12.4%.

The experimental results demonstrate three key findings. First, contrary to conventional practices in histopathology analysis, intermediate resolution levels (10× magnification) consistently outperform the highest resolution (40× magnification), suggesting an optimal balance between cellular detail and contextual information. Second, increased field of view at inference time yields universal performance improvements regardless of resolution level, indicating the importance of contextual information in metastasis detection. Finally, the cascade approach combining Level 3 (5× magnification) for initial screening with Level 1 (20× magnification) for detailed analysis achieves both superior performance (FROC score of 0.661) and improved computational efficiency (12.4% faster than the best single-resolution model).

These findings challenge the prevailing assumption in computational pathology that higher resolution always yields better results, and demonstrate that strategic multi-resolution approaches can simultaneously improve both accuracy and efficiency. The potential clinical impact of these improvements is significant, as reduced processing time could accelerate diagnosis while maintaining or even enhancing detection reliability.

5.4 Limitations and Future Work

While our cascade framework demonstrates promising results, several limitations warrant discussion. First, our study aims to assess the feasibility and potential benefits of a cascading approach in whole slide image (WSI) analysis. We evaluate it using the widely adopted U-Net architecture and the CAMELYON16 dataset, providing a good comparison and benchmark. While this ensures reproducibility, future work is needed to test generalizability across diverse models and datasets. Second, the achieved FROC score of 0.661, though representing a 4.4% improvement over single-resolution approaches, remains below state-of-the-art ensemble methods (0.807) [5]. This gap suggests that architectural improvements and multi-model ensembles could further enhance performance. Additionally, our modular cascade design, while offering deployment flexibility, may suffer from error propagation between stages compared to end-to-end approaches [16]. Future work should explore applying our multi-resolution insights to modern architectures and validating across diverse pathology tasks.

6 Conclusion

This study explored the effectiveness of multi-resolution approaches for cancer detection and segmentation in whole slide images, addressing key questions about resolution, field of view, and cascaded inference pipelines. Our findings challenge conventional assumptions in computational pathology.

We demonstrated that intermediate resolution levels (10× magnification) can outperform the highest resolution (40× magnification) for metastasis detection in the CAMELYON16 dataset, suggesting that the optimal balance between cellular detail and tissue context lies at intermediate magnifications. Additionally, we found that expanding the field of view during inference improves performance across all resolution levels, emphasizing the importance of broader contextual information. Lastly, our cascade inference pipeline, combining a low-resolution model (5× magnification) for region proposal and a higher-resolution model (20× magnification) for detailed analysis, achieved superior performance (FROC score of 0.661) while reducing computational demands.

Our study utilized the U-Net architecture, a widely-used model in medical imaging. However, we believe that our findings are applicable to other architectures as well, offering valuable insights for more efficient and effective histopathological analysis. By combining multiple resolution levels, we can improve both accuracy and efficiency, potentially accelerating the clinical adoption of automated metastasis detection systems.

7 Acknowledgement

This work was funded by EPSRC grant number EP/S021892/1 and Cancer Research Wales. For the purpose of open access the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

1. Abdel-Nabi, H., Ali, M., Awajan, A., Daoud, M., Alazrai, R., Suganthan, P.N., Ali, T.: A comprehensive review of the deep learning-based tumor analysis approaches in histopathological images: segmentation, classification and multi-learning tasks. Cluster Computing (2023). https://doi.org/10.1007/s10586-022-03951-2
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T., Nishino, K. (eds.) Computer Vision – ECCV 2022 Workshops. pp. 205–218. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-25066-8_9
3. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M.P., Zhang, S., Xing, L., Lu, L., Yuille, A., Zhou, Y.: TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024). https://doi.org/10.1016/j.media.2024.103280

4. Dong, N., Kampffmeyer, M., Liang, X., Wang, Z., Dai, W., Xing, E.: Reinforced Auto-Zoom Net: Towards Accurate and Fast Breast Cancer Segmentation in Whole-Slide Images. In: Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.M.R., Bradley, A., Papa, J.P., Belagiannis, V., Nascimento, J.C., Lu, Z., Conjeti, S., Moradi, M., Greenspan, H., Madabhushi, A. (eds.) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 317–325. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_36
5. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A.W.M., and the CAMELYON16 Consortium: Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA **318**(22), 2199–2210 (2017). https://doi.org/10.1001/jama.2017.14585
6. Elmore, J.G., Longton, G.M., Carney, P.A., Geller, B.M., Onega, T., Tosteson, A.N.A., Nelson, H.D., Pepe, M.S., Allison, K.H., Schnitt, S.J., O'Malley, F.P., Weaver, D.L.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA **313**(11), 1122–1132 (2015). https://doi.org/10.1001/jama.2015.1405
7. Guo, Z., Liu, H., Ni, H., Wang, X., Su, M., Guo, W., Wang, K., Jiang, T., Qian, Y.: A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images. Scientific Reports **9**(1), 882 (2019). https://doi.org/10.1038/s41598-018-37492-9
8. Ho, D.J., Yarlagadda, D.V.K., D'Alfonso, T.M., Hanna, M.G., Grabenstetter, A., Ntiamoah, P., Brogi, E., Tan, L.K., Fuchs, T.J.: Deep Multi-Magnification Networks for multi-class breast cancer image segmentation. Computerized Medical Imaging and Graphics **88**, 101866 (2021). https://doi.org/10.1016/j.compmedimag.2021.101866
9. Khened, M., Kori, A., Rajkumar, H., Krishnamurthi, G., Srinivasan, B.: A generalized deep learning framework for whole-slide image segmentation and analysis. Scientific Reports **11**(1), 11579 (2021). https://doi.org/10.1038/s41598-021-90444-8
10. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)
11. Lin, H., Chen, H., Graham, S., Dou, Q., Rajpoot, N., Heng, P.A.: Fast ScanNet: Fast and Dense Analysis of Multi-Gigapixel Whole-Slide Images for Cancer Metastasis Detection. IEEE Transactions on Medical Imaging **38**(8), 1948–1958 (2019). https://doi.org/10.1109/TMI.2019.2891305
12. Litjens, G., Bandi, P., Ehteshami Bejnordi, B., Geessink, O., Balkenhol, M., Bult, P., Halilovic, A., Hermsen, M., van de Loo, R., Vogels, R., Manson, Q.F., Stathonikos, N., Baidoshvili, A., van Diest, P., Wauters, C., van Dijk, M., van der Laak, J.: 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. GigaScience **7**(6), giy065 (2018). https://doi.org/10.1093/gigascience/giy065
13. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979). https://doi.org/10.1109/TSMC.1979.4310076
14. van Rijthoven, M., Balkenhol, M., Siliņa, K., van der Laak, J., Ciompi, F.: HookNet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. Medical Image Analysis **68**, 101890 (2021). https://doi.org/10.1016/j.media.2020.101890

15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Sun, S., Yuan, H., Zheng, Y., Zhang, H., Jiang, Z.: Cancer Sensitive Cascaded Networks (CSC-Net) for Efficient Histopathology Whole Slide Image Segmentation. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 476–480 (2020). https://doi.org/10.1109/ISBI45749.2020.9098695
17. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–9 (2015). https://doi.org/10.1109/CVPR.2015.7298594
18. Verghese, G., Lennerz, J.K., Ruta, D., Ng, W., Thavaraj, S., Siziopikou, K.P., Naidoo, T., Rane, S., Salgado, R., Pinder, S.E., Grigoriadis, A.: Computational pathology in cancer diagnosis, prognosis, and prediction – present day and prospects. The Journal of Pathology **260**(5), 551–563 (2023). https://doi.org/10.1002/path.6163
19. Wang, D., Khosla, A., Gargeya, R., Irshad, H., Beck, A.H.: Deep Learning for Identifying Metastatic Breast Cancer (2016). https://doi.org/10.48550/arXiv.1606.05718
20. Yan, J., Chen, H., Wang, K., Ji, Y., Zhu, Y., Li, J., Xie, D., Xu, Z., Huang, J., Cheng, S., Li, X., Yao, J.: Hierarchical Attention Guided Framework for Multi-resolution Collaborative Whole Slide Image Segmentation. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 153–163. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_15
21. Yeung, M., Sala, E., Schönlieb, C.B., Rundo, L.: Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics **95**, 102026 (2022). https://doi.org/10.1016/j.compmedimag.2021.102026
22. Zeng, L., Tang, H., Wang, W., Xie, M., Ai, Z., Chen, L., Wu, Y.: MAMC-Net: an effective deep learning framework for whole-slide image tumor segmentation. Multimedia Tools and Applications **82**(25), 39349–39369 (2023). https://doi.org/10.1007/s11042-023-15065-x