Influence of magnification in deep learning aided image segmentation in histological digital image analysis

McCombe, K. D., Craig, S. G., James, J. A., & Gault, R. (2022). Influence of magnification in deep learning aided image segmentation in histological digital image analysis. In *Proceedings of the 24th Irish Machine Vision and Image Processing Conference* (pp. 161-164). Irish Pattern Recognition & Classification Society. https://doi.org/10.56541/RAKL2135

Published in:
Proceedings of the 24th Irish Machine Vision and Image Processing Conference

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Influence of Magnification in Deep Learning Aided Image Segmentation in Histological Digital Image Analysis

Kris D. McCombe[1], Stephanie G. Craig[1], Jacqueline A. James[1,] Richard Gault[2]

1. Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Northern Ireland
2. School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Northern Ireland

Abstract

The use of digital pathology has grown significantly for both healthcare and research purposes in recent years. With this comes opportunity to develop systems supported by computer vision (CV) and artificial intelligence (AI), with the potential to improve patient management and quality of care. The accessibility of CV and AI toolboxes have resulted in the rapid application of image analysis in this domain driven by accuracy related metrics. However, in this short paper we illustrate common pitfalls in the field through a semantic segmentation task, specifically how magnification can influence training data quality and demonstrate how this can ultimately affect model robustness.

Keywords: Digital Pathology, Segmentation, Medical Imaging

1 Introduction

Traditional pathology methods involve the examination of a patient tissue sample under a light microscope by a pathologist. Advances in computing hardware, image capture technology, and image compression algorithms have made it possible to scan, digitise, and store images of these patient samples at scale. The examination of computerised images has been termed Digital Pathology [Naizi et al., 2019]. Digital pathology holds many advantages over its traditional counterpart. For example, the digitisation of histopathology images allows for global remote access, which became particularly important over the past three years due to the COVID-19 pandemic [Browning et al., 2020]. Digital image analysis (DIA) can also be performed on these images for research and clinical biomarker quantification [Lara et al., 2021]. The rise of DIA has enabled the creation of more automated workflows through computer vision (CV) and Artificial Intelligence (AI) [Salto-Tellez et al, 2019]. Cancer Research UK estimates suggest that half of the UK population will get cancer in their lifetime [CRUK, 2015]. With increasing consultation requests and a decrease in qualified pathologists to assess them [George et al., 2020], the opportunity to implement an AI-assisted diagnostic workflow is prominent. Ultimately, these AI-driven tools have the potential to shorten time to diagnosis and improve patient outcomes [Steiner et al, 2020]. In this short paper we outline a workflow using semantic segmentation to automate annotation of histological structures in digitised cancer images. In addition, we examine how the choice of magnification can influence training data, and how this can impact on overall model performance and the potential for clinical translation.

2 Methods

This work aims to semantically segment complex structures from digital pathology images across multiple cancer types to examine the impact the magnification of the images can have on overall model performance and potential future clinical translation. Our hypothesis is that increased model performance may not always translate to the most applicable approach for clinical translation.

2.1 Data Description and Ground-truth annotation

Haematoxylin and Eosin-stained slides were obtained with ethical approval from the Northern Ireland Biobank (NIB21/0008) from cancer patients spanning four cancer types: Oropharyngeal, Lung, Breast and Oesophageal Cancer (151 patients in total). The slides from were digitised and saved using the Leica Aperio AT2 at x40 magnification (0.25 μm / pixel). For the purposes of this case study, Tertiary Lymphoid Structures (TLS), an ectopic immune structure that can occur during clinical scenarios of chronic inflammation, including cancer, were chosen as the target for segmentation. TLS have shown prognostic and predictive potential in cancer therapy, and are present across multiple cancer types [Schumacher et al 2022]. Identification of these complex structures is often time-consuming, providing motivation to develop an automated segmentation tool. TLS were manually annotated using QuPath (v0.3.2) [Bankhead et al. 2017] to generate ground-truth binary masks.

2.2 Image Tiling and Extraction

Whole Slide Images (WSIs) are very large often spanning 80k x 60k pixels making them prohibitive to perform deep-learning based image analysis at a whole image level. A common approach in the field to overcome this challenge is to apply tiling; breaking down these images into segments of manageable size that can be handled by commercial computers. This raises questions regarding wider image context and appropriate magnification for histological structures. To investigate this, we extracted tiles across 20x, 10x, 4x and 1x magnification, mimicking a physical microscope, to explore how this affects model training and performance. Tiles were extracted at 50% overlap with a resolution of 512px by 512px. This equated to a "real" tile size of 26.83mm^2, 1.68mm^2, 0.27mm^2 and 0.07mm^2 for 1x, 4x, 10x and 20x respectively. Only tiles that exhibited a TLS were used for training (2208 tiles at 1X, 7675 at 4X, 12841 at 10X and 51275 at 40X). Tile extraction was performed using a bespoke QuPath/Groovy script which can be found here: https://github.com/KDM-Echo/IMVIP22-Magnification. Patients were divided into a 70:15:15 ratio for training, validation and independent test sets. The training, test and validation datasets are split at a patient level to avoid any potential data leakage.

2.3 Model and Training

The semantic segmentation model was built using the PyTorch framework (v1.11.0 + CUDA 11.6) through the segmentation-models-pytorch python package. The model consists of a UNet++ Architecture with an efficientnet-b0 encoder. UNet++ was chosen due to its regular use in digital pathology image segmentation, and the relatively lightweight nature of the model. The model was initiated with pretrained weights from ImageNet. The model was trained for 5 epochs. The Dice Loss function from the segmentation-models-pytorch was used. The optimiser was ADAM set at a learning rate of 0.001. A batch size of 8 was used as this maximised available memory. The model was trained on an Octane VI laptop with 32GB RAM, an intel Core i7-9700 CPU, and RTX 2070 8GB GPU. Models which achieved the highest Intersect over Union (IoU) score in the validation set were brought forward for analysis in the independent test cohort.

2.4 Model Evaluation

Models were evaluated by accuracy, precision, recall, Dice score, IoU score and Matthew's Coefficient in the independent patient set, both at a tile level, which were guaranteed to contain a TLS, and at a whole slide level for all four magnifications. The time taken to run inference on these independent, whole slide test images was also recorded.

3 Results

A summary of the validation results can be seen in Table 1. At a WSI level, the 4X magnification achieved the highest Dice, IOU score, pixel accuracy and Matthew's Coefficient score. In contrast, 1X and 20X models achieved the worst IOU and Matthew's Coefficient. At a WSI level, recall increased with magnification, with precision

decreasing at 10X and 20X magnification. The times taken to segment the 23 whole slide images in the test set were 61 seconds at 1X (22.8 slides/min) 19 minutes 44 seconds at 4X (1.16 slides/min), 105 minutes 16 seconds at 10X (0.22 slides per minute) and 534 minutes 56 seconds at 20x (23.25 min/slide).

Magnification	Accuracy	Dice	IOU	Recall	Specificity	Precision	Matthew's Coefficient
1X (Tile)	0.991	0.611	0.571	0.459	**0.998**	0.725	0.573
4X (Tile)	**0.956**	0.610	0.550	0.709	0.976	0.704	**0.682**
10X (Tile)	0.858	0.579	0.487	0.827	0.868	0.685	0.657
20X (Tile)	0.836	**0.816**	**0.735**	**0.930**	0.676	**0.830**	0.641
1X (WSI)	0.996	0.804	0.788	0.459	**0.999**	**0.700**	0.565
4X (WSI)	**0.997**	**0.935**	**0.931**	0.709	0.998	0.579	**0.639**
10X (WSI)	0.986	0.844	0.842	0.823	0.987	0.201	0.403
20X (WSI)	0.900	0.657	0.656	**0.903**	0.900	0.035	0.166

Table 1. Modelling results for the independent test set. Best result for each category is highlighted in bold. At a tile level, TLS are guaranteed to be in the image whilst the WSI level looks at the entire patient.

4 Discussion

4.1 Correct Magnification is Crucial in Histological Object Detection

In this study, we examined the influence of magnification on quality of complex semantic segmentation. When acting on the individual tiles that were guaranteed to have a TLS, it was found that the model trained at 20X magnification yielded the best Dice and IOU scores. In general, precision and recall both tend to increase with magnification when inferring on these tiles. This is likely due to the changes in negative to positive pixel ratio within the training set between magnifications (88:1 in 1x, 15.4:1 in 4x, 3.5:1 in 10x, 0.74:1 in 20x). With considerably more positive pixels to learn from, it is understandable why these trends exist, with less potential for false positives due to larger proportion of positive pixels in the ground truth images. In real-world scenarios however, these models will not have specifically curated images to evaluate. For this reason, we applied each model across the whole image for the independent test set. All models achieved marked increases in IoU scores however, this is likely due to the relatively large imbalance between non-TLS and TLS labelled pixels indicating that tile-based evaluation is important to assess the model's ability in the given segmentation task. Interestingly, the highest IoU scores were achieved in the 4X magnification (Table 1), a contrast to the best performing 20X images at the tile level. Where recall again tended to increase with magnification, precision appeared to fall, indicating increasing false positives. While this may be down to the increase positive pixel ratio as mentioned before, this may also be the result of a loss of architectural context. Figure 1 shows the same TLS at the different magnifications. These results make it clear that identifying the correct magnification per structure is imperative in model success. The average size of a TLS in our images was 0.21mm^2, close to the 10x tile size of 0.27mm^2, where the negative-positive pixel ratio increases drastically. It may also be advantageous to develop a model that utilises multiple magnifications, simulating the zooming in and out on a physical microscope [Rijthoven et all, 2021].

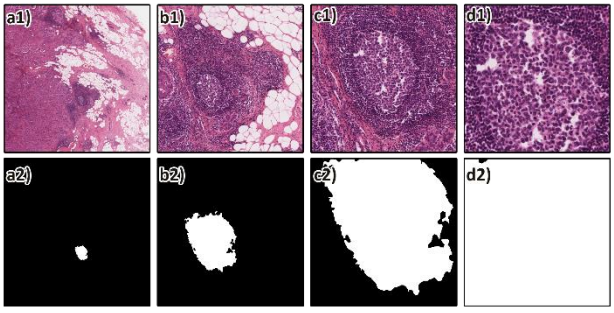

Figure 1. Raw image and corresponding mask of the same TLS at 1x (a) 4x (b), 10x (c) and 20x (d).

4.1 Making a Clinically Relevant Tool

It is important to take into account additional time and hardware costs for each approach. A higher magnification results in a greater number of tiles to examine meaning a longer inference time. For a tool to be clinically relevant, it should at least match the time taken for a pathologist to manually examine the slide in the same manner. Inference at 20X magnification took approximately 9 hours for 23 whole slide images compared to with 20 minutes at 4X for

the same images. For context, review and digital annotation of these images took approximately 5-10 minutes per image. Ultimately there remains somewhat of a knowledge and understanding gap between digital pathology image analysis and the more traditional pathology aspects, meaning the majority of DIA AI research is not translated to clinical use [Steiner, 2020]. It is unreasonable to expect pathologists to develop a background in coding to utilise these deep learning techniques, just as it is to expect computer scientists to become experts in cancer pathology. The development of robust AI models is a collective endeavour and efforts must be made to bridge the knowledge gap between the two disciplines. The development of open-source, graphical user interfaces (GUI) offering a code-free, point-and-click experience, may be a potential remedy. By providing a GUI we can remove the intimidation of coding, and by keeping the software open source, we provide a pathway for community driven development.

5 Conclusion

In this study, we demonstrate the importance of selecting the correct magnification and appropriate context to perform semantic segmentation of digital pathology images. When developing deep learning models for DIA, it is important to understand how changing magnification can impact training data and tissue architectural context. We demonstrate why it is important to extrapolate models to whole patient images for evaluation to reflect real world scenarios and how this can influence optimal model selection. The results show (Figure 1 and Table 1) that increased metrics do not necessarily translate to a robust clinically relevant model. Finally, we show it is important to consider time as a factor, as these tools ultimately need to be designed to improve times to diagnosis as well as impact on patient experience.

References

[Bankhead et al. 2017] Bankhead P., Loughrey M., Fernández J. et al (2017) QuPath: Open source software for digital pathology image analysis. *Scientific Reports (7: 16878)*

[Browning et al., 2020] Browning L., Colling R., Rakha E. et al (2020) Digital pathology and artificial intelligence will be key to supporting clinical and academic cellular pathology through COVID-19 and future crises: the PathLAKE consortium perspective. *Journal of Clinical Pathology (74: 443-447)*

[CRUK, 2015] CRUK 1 in 2 people in the UK will get cancer, *Accessed July 15th 2022* (https://news.cancerresearchuk.org/2015/02/04/1-in-2-people-in-the-uk-will-get-cancer/)

[George et al., 2020**]** George J., Gkousis E., Feast A. et al. (2020). Estimating the cost of growing the NHS cancer workforce in England by 2029, *CRUK*

[Lara et al., 2021] Lara H., Li Z., Abels E., et al. (2021) Quantitative Image Analysis for Tissue Biomarker Use: A White Paper From the Digital Pathology *Association Appl Immunohistochem Mol Morphol (29(7):479-493)*

[Naizi et al., 2019] Niazi M. et al. (2019) Digital pathology and artificial intelligence, *Lancet Oncol (20: 253-261)*

[Rijthoven et all, 2021] Rijthoven M. et al. (2021) HookNet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images *Medical Image analysis (68(9):101890)*

[Salto-Tellez et al, 2019] Salto-Tellez M., Maxwell P., Hamilton P. (2019) Artificial intelligence-the third revolution in pathology *Histopathology (3:372-376)*

[Schumacher et al, 2022] Shumacher, T., Thommen, D. Tertiary Lymphoid Structures in Cancer *Science (375(6576): eabf9419)*

[Steiner et al, 2020] Steiner D., Chen P., Mermel C. (2020) Closing the translation gap: AI applications in digital pathology *Biochimica et Biophysica Acta – Reviews on Cancer (1875(1):188452)*