HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 4 (2024)

Target-Specific Segmentation in Hematoxylin and Eosin (H&E) Stained Pathology Images Using Conditional Segmentation Networks

Dae Young Leem[1], Jiyu Song[2] and Diane Yum[#]

[1]Taipei Fuhsing Private School, Republic of Korea
[2]Korea International School, Jeju, Republic of Korea
[#]Advisor

ABSTRACT

Digital pathology analysis is an advancement in medical diagnostics that leverages digital imaging to enhance the examination and interpretation of pathology slides. By converting traditional glass slides into high-resolution digital images, digital pathology enables pathologists to review and analyze samples with greater precision and efficiency. Machine learning-based techniques in digital pathology are attracting significant attention from pathologists and biologists due to their ability to reduce analysis time and provide objective, accurate results. In particular, semantic segmentation approaches have been widely adopted to isolate specific proteins or cell areas within digital pathology images. However, these methods often exhibit biases toward particular datasets which leads to models that can only process specific targets and lack general applicability. To address this limitation, we propose a target-specific segmentation approach using conditional segmentation networks. We introduce a one-hot vector to control the isolation of target proteins in a conditional manner. The proposed method achieved a pixel accuracy of 83.6% and an intersection over union of 0.7954 on a public pathology segmentation dataset.

Introduction

Biopsy is the process of extracting cells from a specific tissue, organ, or fluid of a human body, and examining them in the laboratory. Its main purpose is to detect cancer or observe cancer cell death in cancer-treating patients. WSI, short for Whole Slide Imaging, is a part of the biopsy that digitalizes physical pathology slides in order for pathologists to examine pathology images with higher resolution that can also be integrated with additional tools for clearer analysis.

Traditionally, pathologists had to manually examine the digital pathology images, which is time-consuming and delays diagnosis. Subjectivity also matters, since pathologists are humans and they have different views, human error and subjectivity are inevitable, which leads to inaccurate diagnosis. With fatigue due to labor-intensive work, accuracy drops further.

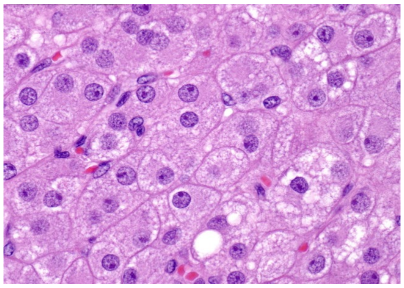

Figure 1. Example of a pathology image

To address this issue, image segmentation-based systems that assist pathologists have been developed. Image segmentation is a pixel-wise classification method that distinguishes and outlines objects within an image. By framing and highlighting images into distinct parts, it improves the analysis quality since it provides well-differentiated boundaries. As shown in Figure 2, image segmentation isolates the target protein areas from input pathology images. This advanced technology aids pathologists and biologists by reducing analysis time and increasing efficiency. This provides fast identification of target areas with less manual effort. However, this approach often exhibit biases toward specific datasets which results in models that are limited to processing particular targets and lack general applicability.

To overcome this limitation, we propose a target-specific segmentation approach using conditional segmentation networks. By introducing a one-hot vector to control the isolation of target proteins conditionally, the proposed method enhances flexibility and generalization. The remainder of this paper is organized as follows: Chapter 2 offers an overview of Whole Slide Imaging and Object Segmentation to provide essential context for the proposed approach. Chapter 3 details the specific steps of the proposed method, while Chapter 4 demonstrates its efficiency through various experimental results. Finally, Chapter 5 summarizes the findings and conclusions of the paper.

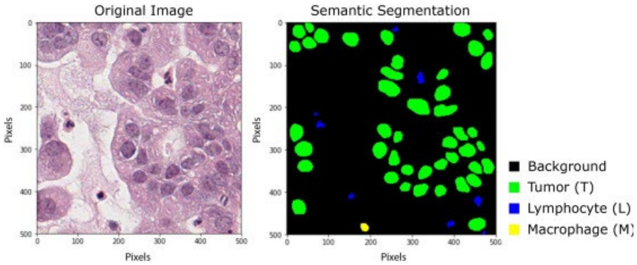

Figure 2. Two images with before (left) and after (right) image segmentation.

Related Work

Whole Slide Image

Whole slide imaging is a digital technique of converting physical slides into high-resolution digital images, which is widely used in pathology to identify malignant and cancerous cells for diagnosis and testing. The progress begins with the extraction of a microscopic tissue sample, which is then prepared for digital imaging. To safeguard the slide contamination, spoilage, or damage, a fixation process is employed. This process involves treating the tissue sample with a chemical solution, such as formalin, to preserve its cellular structure and prevent degradation. Fixation stabilizes proteins and other cellular components, ensuring that the tissue remains intact and accurate for subsequent staining and digital imaging. This step is important for maintaining the integrity of the sample and ensuring reliable diagnostic results.

After preservation and staining, the tissue slide is scanned to generate a high-resolution digital image that captures fine details. These digital images are typically stored in specialized file formats, such as SVS or NDPI, which are designed to preserve the image's integrity and resolution. Once digitized, pathologists review these images manually to identify abnormalities. Despite the advancements offered by WSI, it has significant drawbacks. The process remains highly labor-intensive, requiring extensive time and effort from even experienced pathologists. This labor-intensive nature, coupled with the complexity of the images, can lead to a diagnostic error rate of approximately 50%, highlighting the need for more efficient and accurate methods in digital pathology (Raab et al. 2005).

Object Segmentation

Semantic segmentation is a computer vision technique where the goal is to classify each pixel in an image into a specific category. Unlike object detection, which identifies and classifies objects in bounding boxes, semantic segmentation provides a pixel-level understanding of an image. Semantic segmentation is widely applied in pathology image analysis to facilitate the detailed examination of tissue samples. This enables pathologists to identify, classify, and quantify various structures and abnormalities within these images (Rashidi et al. 2019). Pathology images, typically obtained through microscopy of stained tissue sections, contain intricate details that require precise analysis at a cellular or sub-cellular level. Semantic segmentation aids in automating and enhancing this process.

Semantic segmentation is often implemented using the U-Net (Ronneberger et al. 2015) architecture, a type of convolutional neural network designed for tasks that require precise pixel-level predictions, like medical image segmentation. In this study, we utilized the U-Net architecture to develop a target-specific segmentation system. A detailed explanation of the proposed system is provided in Chapter 3.

Conditional Segmentation Network

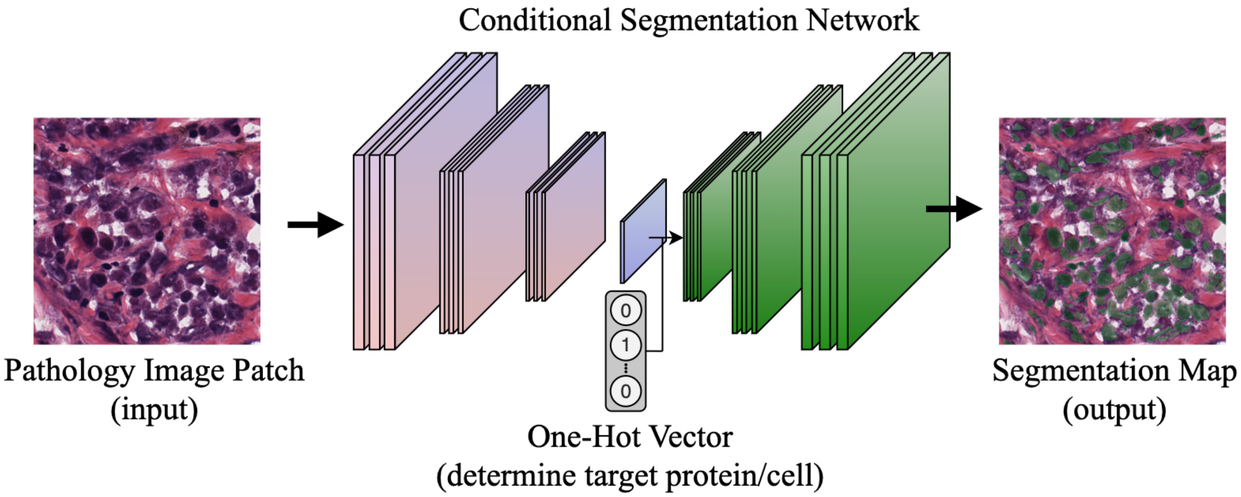

Figure 3. Architecture of the proposed conditional segmentation network

Figure 3 illustrates the architecture of the proposed conditional segmentation network. The proposed network is composed of three modules: an encoder, a decoder, and a one-hot vector input module. The encoder takes a high-resolution digital image of a pathology image patch stained with H&E as an input and generates a feature map using the convolutional neural network. As the feature maps in the encoder decrease in size, down sampling also progresses while creating the feature map. This feature map mathematically represents the important features of the imputed pathology images such as geometric patterns, shapes of cells, or colors. The one-hot vector controls which target protein to isolate in a binary manner. This vector is T-dimensional, where T represents the number of target proteins or cells. For example, if the user wants to isolate the first category, the first value in the one-hot vector is set to 1, while all other values are set to 0.

The generated feature map is concatenated with the one-hot vector and then fed into the decoder to produce the segmentation map. The decoder is an inverted version of the encoder, with the downsampling layers replaced by upsampling layers. This process is summarized in Equations 1 and 2.

Equation 1: Encoder

$$f = Encoder(I; W_{enc})$$

The variable ***I*** denotes the input image that the encoder will process, while W_{enc} represents the learnable parameters of the encoder.

Equation 2: Decoder

$$\widehat{Seg} = Decoder(f, c; W_{dec})$$

Here, *f* and *c* denote the feature maps and one-hot vectors, respectively. The variable W_{dec} represents the learnable parameters of the decoder. To train the proposed network, we utilized pixel-wise cross-entropy loss function as explained in Equation 3.

Equation 3: Pixel-wise cross-entropy loss function

$$Loss = -\sum_{i}\sum_{j}\sum_{t=1}^{T} Seg(i, j, t) \times log_e(\widehat{Seg}(i, j, t))$$

Here, *T* represents the total number of categories that users can select which allows the system to delineate the target. $\widehat{Seg}$ denotes the predicted probability matrix which indicates the likelihood that a specific pixel belongs to particular categories or cell types, such as ring cells, Ki-67, or metastatic cells. Seg refers to the corresponding ground truth probability matrix.

Then, the output probability matrix is put into the cross-entropy loss function, and hence gets multiplied by Seg, which represents the actual probability, either 1 or 0. For $i \times j \times T$ times, the system iterates this action and accumulate all the prediction actual answer values to calculate the overall loss. Ideally, the loss would converge towards 0, where $\widehat{Seg}$ values are all 1 or 0. To reach so, it is necessary to train the system.

Equation 4: Gradient descent (encoder)

$$W_{enc}^{t+1} \leftarrow W_{enc}^{t} - lr\frac{\partial Loss}{\partial W_{enc}^{t}}$$

Here, W_{enc}^{t}, short for weights, denotes the learnable parameters of the encoder at specific training time *t*. To update these parameters, we first compute the gradient of W_{enc}^{t} with respect to the aforementioned loss function. The gradient is then multiplied by the learning rate, which controls the speed and stability of the learning process.

To optimize the Wenct, the proposed system will iteratively subtract the derivativelearning rate from the original W_{enc}^{t}, and it is defined as W_{enc}^{t+1}.

Equation 5: Gradient descent (decoder)

$$W_{dec}^{t+1} \leftarrow W_{dec}^{t} - lr\frac{\partial Loss}{\partial W_{dec}^{t}}$$

Just like the gradient descent algorithm for encoder, the decoder undergoes the same process. The gradient of W_{dec}^{t} with respect to the loss function is computed, multiplied by learning rate, and gets subtracted from W_{dec}^{t} , which is defined as W_{dec}^{t+1}.

Experimental Results

Dataset

To train and evaluate the proposed system, two datasets were used. The PanNuke dataset (Gamper et al. 2020) contains 7,904 samples from various types of human organs. The other dataset, DeepLIIF dataset (Ghahremani et al. 2022) includes 1,624 samples with different protein stains such as Ki-67, LAP-2, and BCL-2. All samples from both datasets are stained with hematoxylin and eosin. For training and testing the proposed system, the two datasets were split into 80% for training and 20% for testing.

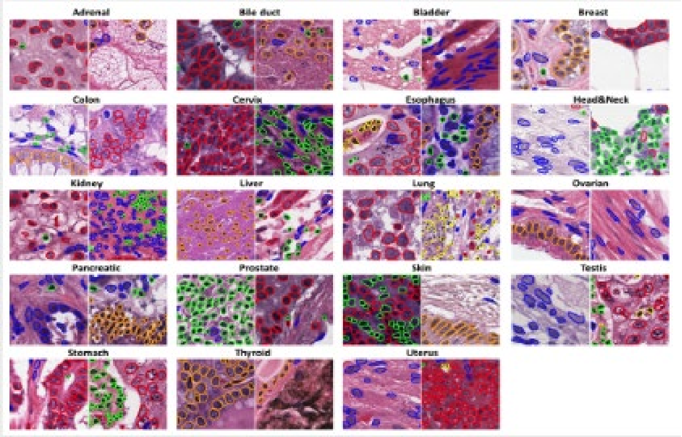

Figure 4. Snapshot of PanNuke dataset (Gamper et al. 2020)

Evaluation Metric

We utilized two evaluation metrics—pixel accuracy and Intersection over Union (IoU)—which are commonly used to assess the performance of semantic segmentation tasks.

Equation 6: Pixel Accuracy

$$Pixel\ Accuracy = \frac{number\ of\ true\ positive\ pixels}{total\ number\ of\ pixels}$$

Shown in Equation 6, pixel accuracy is the ratio of the number of correctly predicted pixels to the total number of pixels. IoU can quantify the segmentation performance of the trained model (Zhang et al. 2008). It is the value of ground truth and prediction's intersection area divided by their union area. When prediction and ground truth are overlapped, the intersection area is "true positive", prediction minus intersection is "false positive", and ground truth minus intersection is "false negative". As a result, we can redefine the equation as shown below.

Equation 7: IoU

$$IoU = \frac{TP}{TP + FP + FN}$$

Performance Comparison

Table 1: Performance comparison with different learning rate setups

LR: learning rate	Pixel Accuracy	IoU
UNet (LR 1e-3)	74.7 %	0.6629
UNet (LR-1e-4)	80.9%	0.7218
UNet (LR 1e-4 with schedule)	83.6 %	0.7954

As shown in Table 1 and Figure 5, when learning rate was set as 0.001, pixel accuracy was 74.7% and IoU was 0.6629. With 10 times less learning rate, pixel accuracy and IoU increased to 80.9% and 0.7218 respectively. Moreover, the optimal performance was achieved when we applied a learning rate schedule. For the first 250 epochs, we set the learning rate at 0.0001 while the last 250 epochs was set in half of the previous learning rate, 0.00005. With this hyperparameter setting, the proposed system achieved a pixel accuracy of 83.6% and an IoU of 0.7954.

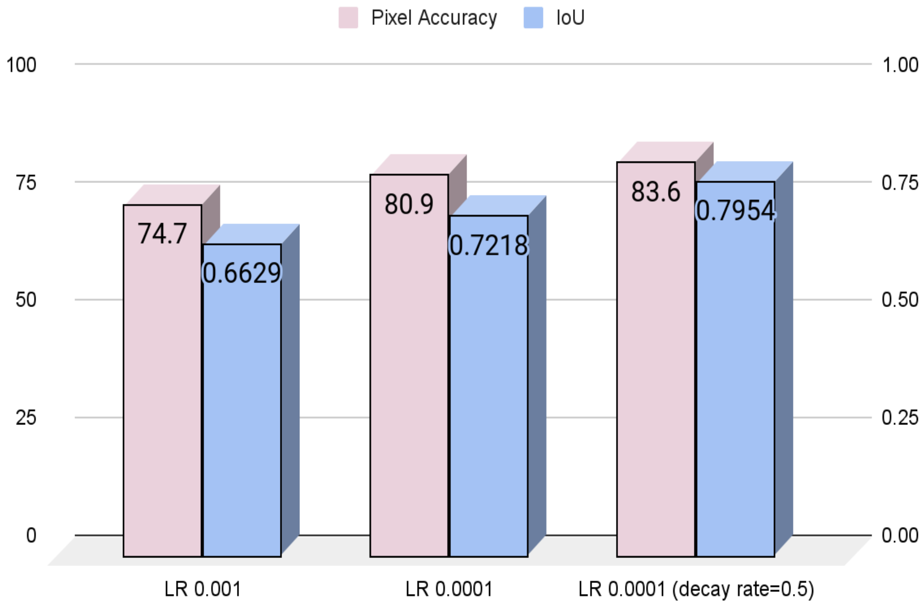

Figure 5. Performance comparison with different learning rate setups

Qualitative Experiement

We also conducted qualitative experiments. Figure 6 shows some of the samples we tested. The boundaries between segmented cells are clear and the gap of segmentation between true mask and predicted mask is kept in small scale.

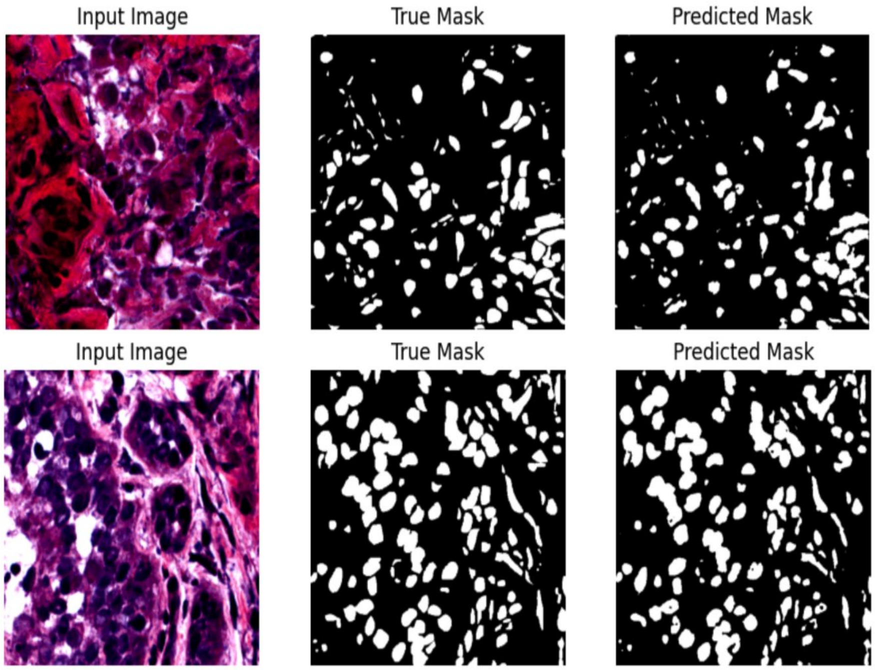

Figure 6. Qualitative experiment

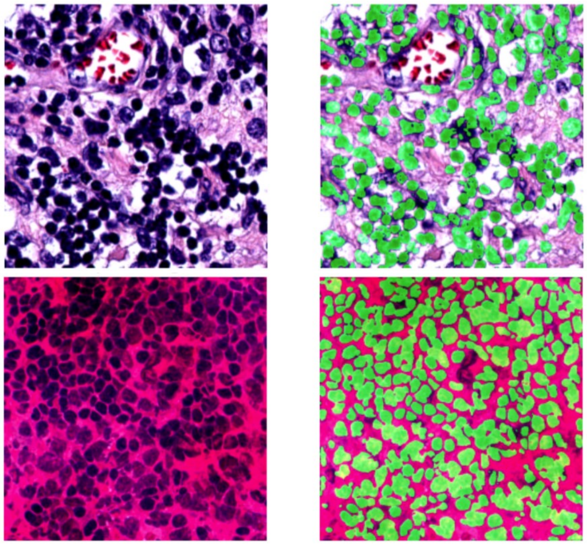

Figure 7. Ground truth alignment

Additionally, we performed a ground truth mask alignment experiment to visually assess the accuracy of the segmentation map generated by the proposed system. As shown in Figure 7, the system effectively isolates the target protein. This results clearly demonstrates the effectiveness of the proposed approach. By overlaying the segmented cells on the original pathology images, we expect that pathologists and biologists will be able to easily analyze and interpret the data which enhances the overall diagnostic process.

Conclusion

To provide a solution for the problem that occurs when doctors conduct the traditional WSI, we created a machine learning model that can segmentate the selected target proteins in the high-resolution WSI image. Other than the providings of time and labor sufficienting benefits, the model also showed high accuracy at a pixel accuracy of 83.6% and IoU of 0.7954. As targeting proteins are available, it was shown to be possible to segmentate different types of proteins without using different machines for each protein. This new method we suggest could be a solution for all issues that the traditional method contains: time-consuming, labor-intensity, and accuracy.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Gamper, J., Koohbanani, N. A., Benes, K., Graham, S., Jahanifar, M., Khurram, S. A., ... & Rajpoot, N. (2020). Pannuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778.

Gesson, K., Vidak, S., & Foisner, R. (2014, May). Lamina-associated polypeptide (LAP) 2α and nucleoplasmic lamins in adult stem cell regulation and disease. In Seminars in cell & developmental biology (Vol. 29, pp. 116-124). Academic Press.

Ghahremani, P., Li, Y., Kaufman, A., Vanguri, R., Greenwald, N., Angelo, M., ... & Nadeem, S. (2022). Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification. Nature machine intelligence, 4(4), 401-412.

Raab, S. S., Nakhleh, R. E., & Ruby, S. G. (2005). Patient safety in anatomic pathology: measuring discrepancy frequencies and causes. Archives of pathology & laboratory medicine, 129(4), 459-466.

Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. Academic pathology, 6, 2374289519873088.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

Zhang, H., Fritts, J. E., & Goldman, S. A. (2008). Image segmentation evaluation: A survey of unsupervised methods. computer vision and image understanding, 110(2), 260-280.