HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Organ-Agnostic Whole Slide Image Analysis Using Self-Supervised Transfer Learning

Jiwoo Sim[1] and Bruce Kohl[#]

[1]Cranbrook Schools, USA
[#]Advisor

ABSTRACT

Traditional methods for pathology image analysis are well-known for their time-consuming and labor-intensive nature, often relying on the expertise of pathologists. In recent years, numerous research studies have been proposed to develop automated systems using machine learning approaches to address these challenges. While these systems have demonstrated promising performance, they often exhibit bias towards specific organs, cells, or tasks, limiting their ability to provide generalized solutions for pathology image analysis. To address this issue, I propose an organ-agnostic pathology image analysis system that leverages a self-supervised transfer learning approach. The proposed system comprises two stages: self-supervised representation learning and transfer learning. In the self-supervised representation learning phase, a machine learning model is trained to consistently extract essential features encapsulating the characteristics of diverse pathological images such as visual patterns of tumors. Subsequently, in the transfer learning phase, these well-pretrained models are utilized to train downstream tasks, such as tumor type classification or cancer area segmentation. The proposed approach outperforms all existing state-of-the-art supervised methods in multiple public pathology image benchmarks.

Introduction

Pathology image analysis, often referred to as digital pathology or histopathology image analysis, is a field of medical science and computer science that involves the application of advanced image processing and machine learning techniques to the analysis of pathological images. These images typically come from tissue samples, biopsies, or other specimens obtained from patients. The main goal of pathology image analysis is to assist pathologists and healthcare professionals in the diagnosis, prognosis, and treatment planning of various diseases and conditions, including cancer, infectious diseases, autoimmune disorders, and more.

Traditionally, pathologists have relied on their expertise to meticulously examine tissue specimens, making critical judgments based on visual inspection. However, this manual and subjective process is not without its limitations—often marked by time-consuming analyses, potential inter-observer variability, and the burden of ever-increasing caseloads.

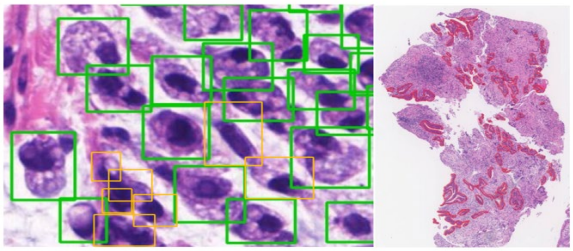


Figure 1. Example of pathology images. (Left): signet ring cell, (right): lesion segmentation from colon tumor

Lately, machine learning approaches that have demonstrated remarkable performance in numerous computer vision challenges are being rapidly applied to the field of pathology image analysis. Cruz-Roa et al. have demonstrated promising results by employing convolutional neural networks to detect regions of invasive ductal carcinoma tissue within whole slide images of breast cancer (Cruz-Roa et al. 2014). Šarić et al. present a fully automated method for the detection of lung cancer in whole slide images of lung tissue samples (Šarić et al. 2019). Their approach is developed using two well-known convolutional neural networks, VGG (Simonyan et al. 2014) and ResNet (He et al. 2016). Wang et al. proposed a patch-based convolutional neural network to perform the category prediction (normal or tumor) on the patches extracted from the 60 liver tumor whole slide images (Wang et al. 2021). While numerous research studies have been proposed, they often exhibit biases toward specific organ cells or tasks. This necessitates the retraining of the model for each specific task or organ, highlighting a significant demand for the development of a unified pathology image analysis system.

In this research, I introduce an organ-agnostic pathology image analysis system that utilizes a self-supervised transfer learning approach. The system consists of two stages: self-supervised representation learning and transfer learning. In the self-supervised representation learning phase, a machine learning model is trained to consistently extract important features that encapsulate the diverse characteristics of pathological images, including visual tumor patterns. Following this, in the transfer learning phase, these well-pretrained models are applied to train downstream tasks, such as tumor type classification or cancer area segmentation.

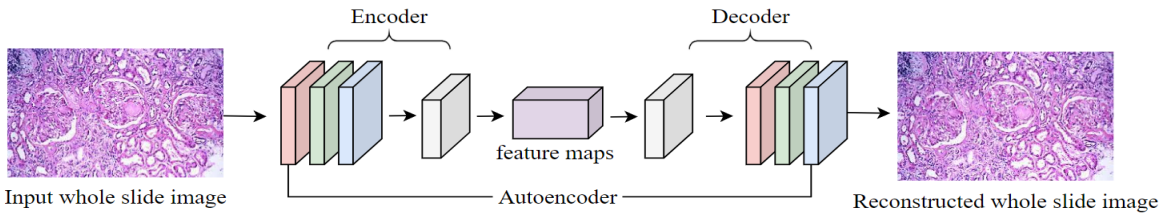
The primary contributions of this research paper are as follows:

1. To the best of my knowledge, this is the first attempt to develop the unified and organ-agnostic pathology image analysis system.

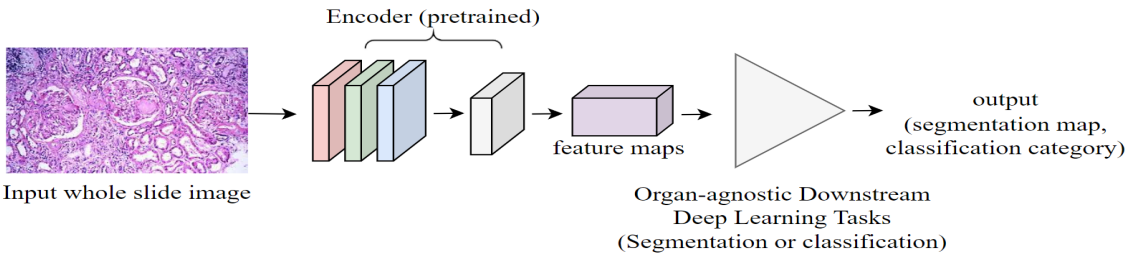
2. Through an extensive series of experiments, the efficacy of the two-stage approach (comprising representation learning and transfer learning) has been thoroughly examined.

Chapter 2 provides an in-depth exploration of the proposed pathology image analysis system, Chapter 3 presents a comprehensive overview of the experimental results, and Chapter 4 offers a summary of the research.

Proposed Whole Slide Image Analysis System



(a)



(b)

Figure 2. Architecture of the proposed pathology image analysis system. (a): Pathology image representation learning, and (b): Transfer learning for downstream task

In this chapter, I offer a comprehensive and detailed account of the proposed pathology image analysis system, including an overview of the system architecture, the data processing procedure, and implementation. Figure 2 illustrates the architecture of the proposed system. In Figure 2(a), I depict the pathology image representation learning stage, which is designed to train the network to consistently extract essential features encapsulating the visual characteristics of various tumor types and human organs. In Figure 2(b), I provide an overview of the transfer learning stage, demonstrating how the pretrained network is leveraged to train downstream tasks, such as tumor segmentation or cancer type classification.

Pathology Image Representation Learning

The objective of the proposed pathology image representation learning is to train an autoencoder, consisting of an encoder and a decoder, to consistently extract essential image features. Initially, the pathological whole slide image is input into the encoder, where it is transformed into feature maps. These feature maps are subsequently passed to the decoder to reconstruct the original input slide image. Throughout this process, the encoder is compelled to extract vital image features, encompassing global organ cell information and visual tumor cell patterns. The decoder reconstructs the original inputted image by learning to reverse the encoding process, mapping the features in the feature space back into the same format as the input data.

For the encoder, I employed Resnet34 (He et al., 2016), a popular convolutional neural network model that shows remarkable performance across various computer vision tasks. In constructing the decoder, I replaced the downsampling layers in Resnet34 with upsampling layers to restore the input image to its original dimensions.

To train the proposed network, I employed the L1 loss function, a commonly used method for training autoencoder architectures. The mathematical formulation of the L1 loss function is elucidated in Equation 1.

Equation 1: L1 loss function

$$L_1 = \frac{1}{HW} \sum_{x}^{W} \sum_{y}^{H} \left| I_{(x,y)} - \widehat{I_{(x,y)}} \right|$$

Here, W and H denote the width and height of the input whole slide image. $I(x,y)$ and $I\widehat{(x,y)}$ denotes pixel intensity of the input image and reconstructed image, respectively. This function calculates the mean pixel-wise error between the original input whole slide image and the reconstructed image.

For implementation details, I train the network using Adam optimizer (Kingma et al. 2014) for 200 epoches. The initial learning rate is set to 0.0001 and is reduced by a factor of 0.1 at the 110th and 160th epochs to prevent loss saturation.

Transfer Learning

In this chapter, I explain the utilization of the pretrained network for a range of organ-agnostic downstream pathological analysis tasks. As illustrated in Figure 2(b), the encoder, having been trained during the representation learning phase, serves as the foundational starting point for each of these pathological analysis downstream tasks. The pretrained encoder takes various types of whole slide images as input and generates feature maps that are supplied to downstream networks, including image classification and segmentation networks. In this study, I trained two distinct classification networks and a nuclei segmentation network. The comprehensive details about each task presented in Chapter 3. For both classification networks, I implemented two linear layers with a hidden size of 128. In the case of the segmentation network, I utilized an inverted Resnet-34 architecture. I replaced the downsampling layers with upsampling layers to reconstruct the original dimensions of the whole slide image.

To train both the classification and segmentation networks, I employed the cross-entropy loss function, as depicted in Equation 2. This loss function is a standard choice for training both classification and image segmentation tasks due to its effectiveness.

Equation 2: Cross-entropy loss function

$$loss = -\frac{1}{N}\sum_{i}^{N}(y_i \log x_i - (1 - y_i)\log(1 - x_i))$$

Here, N denotes the number of samples in the training dataset. The variables $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ denote the network's prediction and its ground truth, respectively. Regarding the training hyperparameters, both tasks were trained using the Adam optimizer, consistent with the representation learning phase, for a duration of 80 epochs, with a fixed learning rate of 0.0001.

Experimental Results

Dataset

To train and assess the proposed whole slide image analysis system, I leveraged three publicly available datasets: PatchCamelyon metastatic dataset (Veeling et al. 2018), MicroSatelliteInstable (MSI) and MicroSatelliteStable (MSS) dataset (Mahbod et al. 2023), and CryoNuSeg segmentation dataset (Mahbod et al. 2021).

The PatchCamelyon metastatic dataset comprises 327,680 color images. These images are extracted from histopathologic scans of lymph node sections and are accompanied by binary labels denoting the presence or absence of metastatic tissue. Approximately 42.8% of these samples correspond to positive cases of metastatic tissue, while the remainder represent normal tissue.

The MSI and MSS dataset consists of 192,312 image patches extracted from histological images of patients with colorectal and gastric cancers. The image patches are categorized into two groups: MSI and MSS. MSI-type cancer can be effectively treated with immunotherapy. Given the efficacy of immunotherapy in treating MSI-type cancer, precise classification becomes essential and imperative.

The CryoNuSeg segmentation dataset comprises frozen H&E-stained histological images, featuring a collection of 30 image patches representing 10 different human organs. This dataset serves as a valuable resource for training and validating algorithms designed for nuclei instance segmentation tasks. Figure 5 shows a snapshot of the dataset, where the first column displays the pathology images, while the subsequent columns depict the corresponding ground truth segmentation maps.

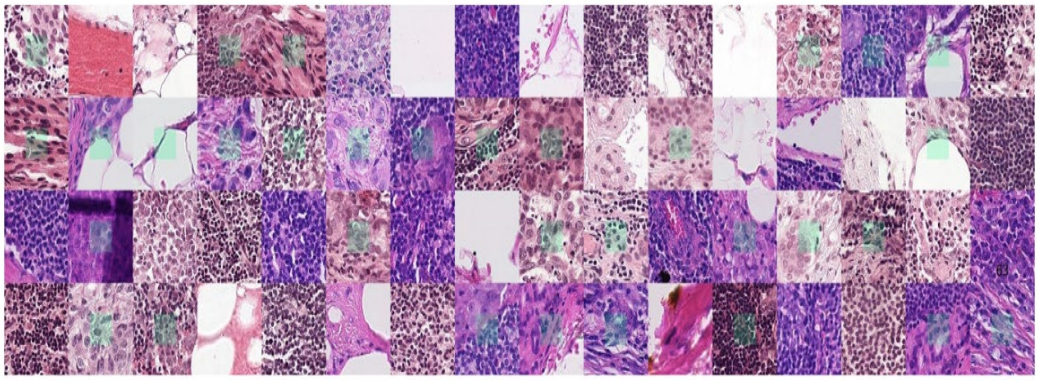


Figure 3. Snapshot of PatchCamelyon metastatic dataset (Veeling et al. 2018)

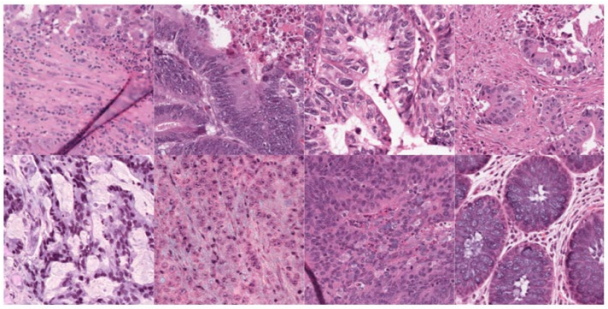


Figure 4. Snapshot of MSI and MSS dataset (Mahbod et al. 2023)

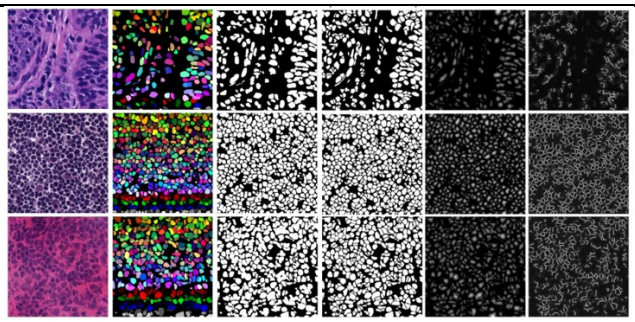


Figure 5. Snapshot of CryoNuSeg segmentation dataset (Mahbod et al. 2021)

Evaluation Metrics

In this chapter, I introduce the evaluation metrics employed to assess the performance of the proposed method.

For the classification task, the metrics utilized include accuracy, recall, precision, and F1-score, which are widely recognized for their effectiveness in classification assessment. Additionally, I employ the confusion matrix and Receiver Operating Characteristic (ROC) curve to further analyze the model's performance. A confusion matrix is a table used to evaluate the performance of a classification algorithm, particularly in binary classification problems. It provides a detailed breakdown of the model's predictions and actual outcomes, allowing for a more in-depth analysis of the model's performance. On the other hand, a ROC curve is a graphical representation that visualizes the trade-off between the true positive rate, also known as sensitivity or recall, and the false positive rate as the classification threshold changes.

Metastatic Tissue Classification

Table 1. Performance comparison on PatchCamelyon metastatic dataset.

Method	Accuracy	Recall	Precision	F1-Score
AlexNet (Krizhevsky et al. 2012)	0.9272 (±0.0014)	0.9427 (±0.0012)	0.9116 (±0.0007)	0.9265 (±0.0011)
VGG19 (Simonyan et al. 2014)	0.9338 (±0.0008)	0.9486 (±0.0006)	0.9176 (±0.0010)	0.9321 (±0.0014)
MobileNetV2 (Sandler et al. 2018)	0.9346 (±0.0012	0.9519 (±0.0013)	0.9173 (±0.0008)	0.9342 (±0.0006)
EfficientNet-B7 (Tan et al. 2019)	0.9356 (±0.0013	0.9517 (±0.0004)	0.9191 (±0.0007)	0.9347 (±0.0011)
HRNet-40 (Wang et al. 2020)	0.9416 (±0.0016	0.9577 (±0.0009)	0.9260 (±0.0011)	0.9414 (±0.0016)
Resnet-34 (He et al. 2016)	0.9444 (±0.0014)	0.9591 (±0.0011)	0.9251 (±0.0008)	0.9431 (±0.0010)
Proposed Method (Resnet-34 based)	**0.9644 (±0.0005)**	**0.9800 (±0.0007)**	**0.9480 (±0.0010)**	**0.9637 (±0.0009)**

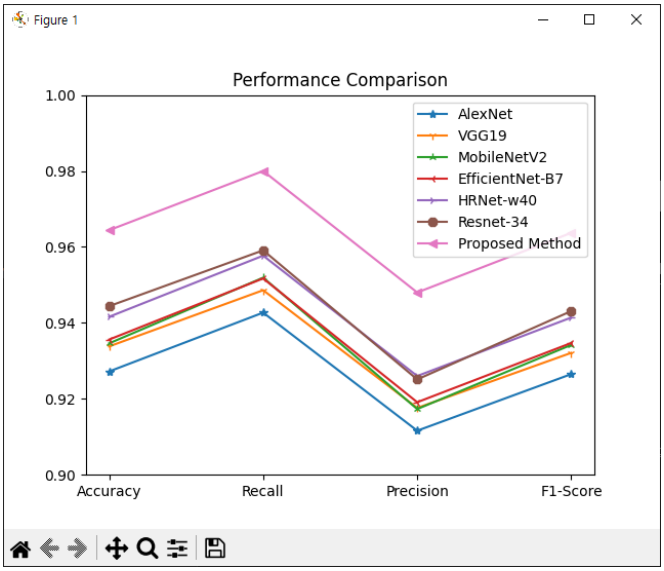
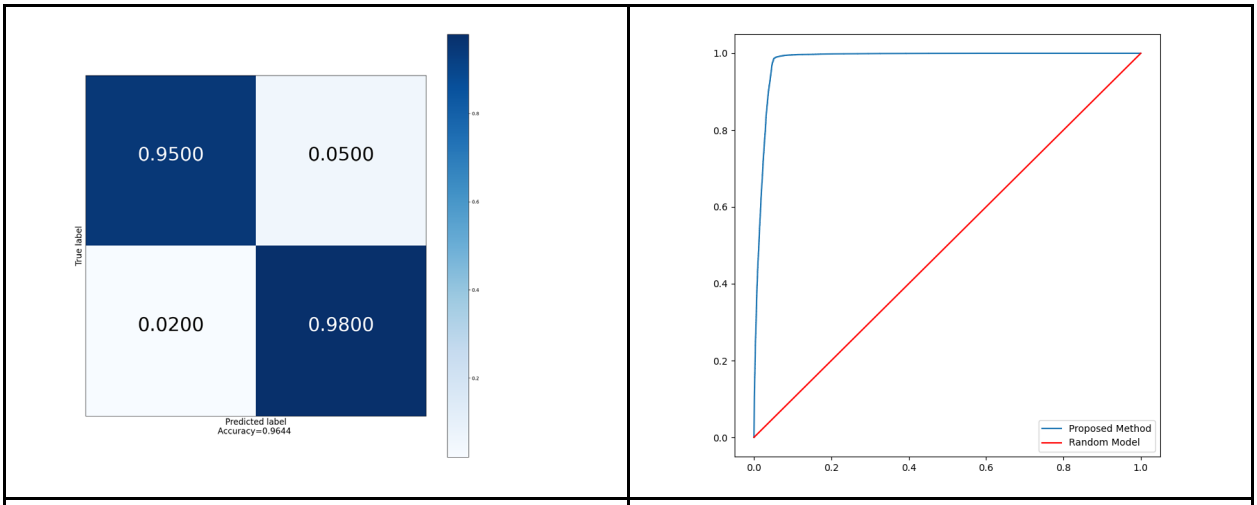


Figure 6. Performance comparison on PatchCamelyon metastatic dataset (line graph)

Table 1 and Figure 6 present a performance comparison between the proposed method and state-of-the-art classification techniques on the PathCamelyon metastatic dataset. The comparison includes well-known models such as AlexNet (Krizhevsky et al. 2012), VGG19 (Simonyan et al. 2014), MobileNetV2 (Sandler et al. 2018), EfficientNet-B7 (Tan et al. 2019), HRNet-w40 (Wang et al. 2020), and Resnet-34 (He et al. 2016), all of which have demonstrated competitive performance in classification tasks.

Interestingly, VGG19, MobileNetV2, and EfficientNet-B7, characterized by relatively shallow network architectures, exhibited subpar results. In contrast, HRNet-w40 and Resnet-34, which have deeper networks, delivered superior performance. Notably, the proposed method outperformed all of the aforementioned comparison methods, clearly establishing its superiority. This superiority can be attributed to the novel representation learning approach employed, enabling the trained model to effectively capture essential visual characteristics better than other methods.

Figure 7 depicts both the confusion matrix and ROC curve for the proposed method. The diagonal elements in the confusion matrix signify the method's ability to classify input samples accurately and consistently. Furthermore, the ROC curve clearly demonstrates the superiority of the proposed method, as indicated by its substantial area under the curve.



(a) (b)

Figure 7. Evaluation result on PatchCamelyon metastatic dataset. (a): confusion matrix, and (b): ROC curve

MSI and MSS Classification

In this chapter, another classification experiment is introduced to showcase the transferability of the proposed method. The experimental protocol mirrors that of the study conducted in Chapter 3.4, with the sole difference being the replacement of the dataset, which is now the MSI and MSS classification dataset.

Table 2. Performance comparison on MSI and MSS classification dataset.

Method	Accuracy	Recall	Precision	F1-Score
AlexNet (Krizhevsky et al. 2012)	0.9064 (±0.0010)	0.8853 (±0.0011)	0.9216 (±0.0012)	0.9024 (±0.0013)
VGG19 (Simonyan et al. 2014)	0.9108 (±0.0009)	0.8907 (±0.0010)	0.9285 (±0.0009)	0.9067 (±0.0008)
MobileNetV2 (Sandler et al. 2018)	0.9139 (±0.0012)	0.8912 (±0.0011)	0.9275 (±0.0012)	0.9096 (±0.0009)
EfficientNet-B7 (Tan et al. 2019)	0.9251 (±0.0014)	0.9023 (±0.0007)	0.9422 (±0.0008)	0.9211 (±0.0009)
HRNet-40 (Wang et al. 2020)	0.9270 (±0.0008)	0.9055 (±0.0007)	0.9410 (±0.0010)	0.9236 (±0.0012)
Resnet-34	0.9338 (±0.0012	0.9121 (±0.0009)	0.9477 (±0.0010)	0.9302 (±0.0011)

(He et al. 2016)				
Proposed Method (Resnet-34 based)	**0.9510 (±0.0008)**	**0.9300 (±0.0009)**	**0.9656 (±0.0011)**	**0.9474 (±0.0013)**

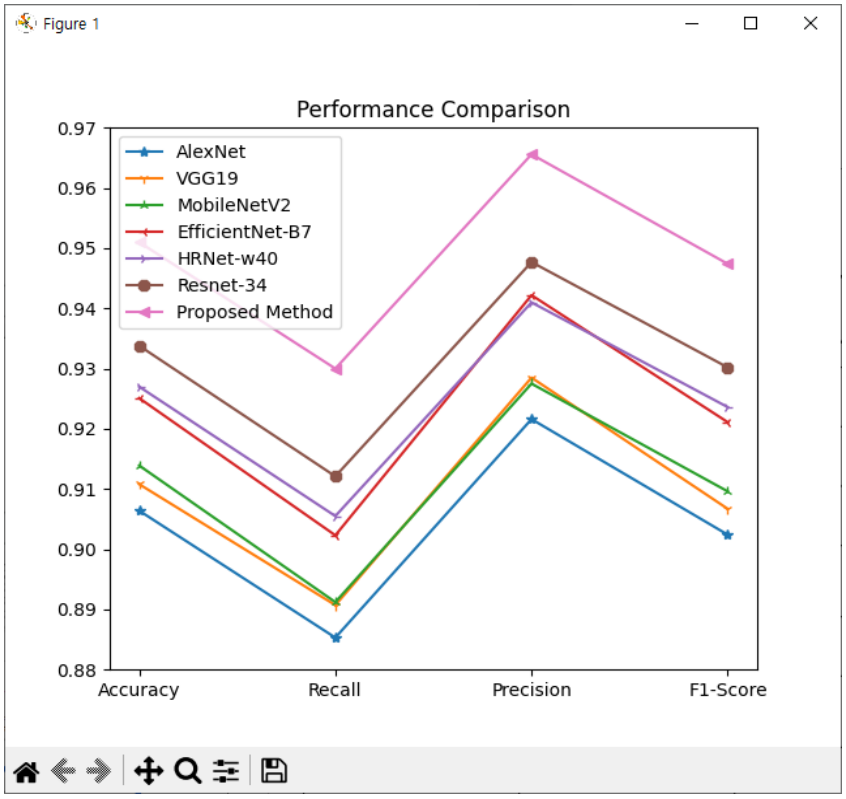
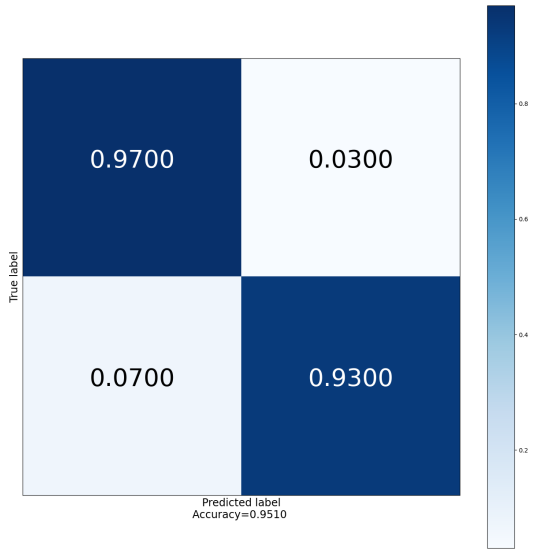
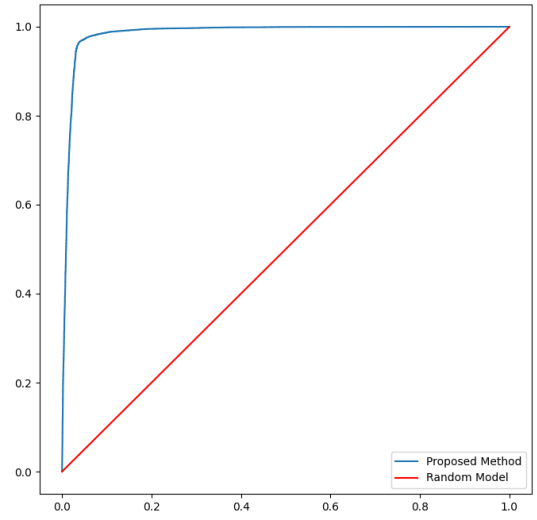


Figure 8. Performance comparison on MSI and MSS dataset (line graph)

Table 2 and Figure 8 provide a summary of the evaluation results on the MSI and MSS dataset. Much like the findings from the initial classification experiment, it's evident that shallow convolutional neural networks (AlexNet, VGG19, MobileNetV2, and EfficientNet-B7) yielded poor results, whereas deeper networks (HRNet-w40 and Resnet-34) demonstrated superior performance. Notably, the proposed method outperformed all compared methods with a significant performance margin. This second experiment further reinforces the effectiveness of the proposed representation learning-based approach over other supervised methods.





(a) (b)

Figure 9. Evaluation result on MSI and MSS dataset
(a): confusion matrix, and (b): ROC curve

Figure 9 illustrates the confusion matrix and ROC curve of the results obtained from the second experiment conducted with the proposed method. Similar to the first experiment, the diagonal elements in the confusion matrix represent the method's consistent and accurate classification of input samples.

Nuclei Segmentation

Finally, I introduce a nuclei segmentation experiment using the CryoNuSeg segmentation dataset to offer a broader perspective on understanding the proposed method. In this particular experiment, I assessed the mean IOU values for comparison. The comparison methods chosen for this task include PSPNet (Zhao et al. 2017), IDW-CNN (Wang et al. 2017), Multipath-RetineNet (Lin et al. 2017), Resnet-38-MS-COCO (Wu et al. 2019), and DeepLabv3 (Chen et al. 2017), all of which have demonstrated competitive results in image segmentation tasks.

Table 3. Performance comparison on CryoNuSeg segmentation dataset.

Method	**mIOU**
PSPNet (Zhao et al. 2017)	77.9
IDW-CNN (Wang et al. 2017)	78.5
Multipath-RetineNet	80.2

(Lin et al. 2017)	
Resnet-38-MS-COCO (Wu et al. 2019)	83.9
DeepLabv3 (Chen et al. 2017)	84.8
DeepLabv3+ (Chen et al. 2017)	85.4
Proposed Method	**88.7**

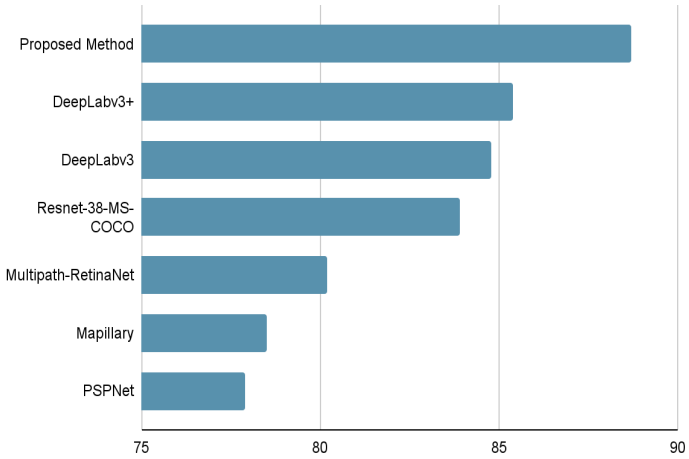


Figure 10. Performance comparison on CryoNuSeg segmentation dataset

Table 3 and Figure 10 present the results of the segmentation experiment comparison. The proposed method outperforms previous supervised segmentation approaches. Remarkably, the proposed method exhibits a significant performance advantage over Resnet-38-MS-COCO, despite both having similar layer depths and architectural components. This analysis underscores the superiority and effectiveness of the proposed pathology representation learning approach.

Conclusion

In this study, I introduce a pathology image analysis system that is agnostic to specific organs and utilizes a self-supervised transfer learning approach. The system is structured into two distinct stages: self-supervised representation learning and transfer learning. To assess the performance of the proposed methods, I conducted three distinct experiments on three different datasets. The outcomes of two pathology image classification experiments and a nuclei segmentation experiment clearly demonstrate the effectiveness and superiority of the proposed method.

In the future, I plan to develop an application service to deliver the research findings to real-world pathology cases, thereby contributing to the field of healthcare.

Acknowledgments

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

References

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. https://doi.org/10.48550/arXiv.1706.05587

Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., ... & Madabhushi, A. (2014, March). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In Medical Imaging 2014: Digital Pathology (Vol. 9041, p. 904103). SPIE. https://doi.org/10.1117/12.2043872

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Kaczmarzyk, J. R., Gupta, R., Kurc, T. M., Abousamra, S., Saltz, J. H., & Koo, P. K. (2023). ChampKit: a framework for rapid evaluation of deep neural networks for patch-based histopathology classification. Computer Methods and Programs in Biomedicine, 107631. https://doi.org/10.1016/j.cmpb.2023.107631

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1925-1934). https://doi.org/10.48550/arXiv.1611.06612

Mahbod, A., Schaefer, G., Bancher, B., Löw, C., Dorffner, G., Ecker, R., & Ellinger, I. (2021). CryoNuSeg: A dataset for nuclei instance segmentation of cryosectioned H&E-stained histological images. Computers in biology and medicine, 132, 104349. https://doi.org/10.1016/j.compbiomed.2021.104349

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520). https://doi.org/10.48550/arXiv.1801.04381

Šarić, M., Russo, M., Stella, M., & Sikora, M. (2019, June). CNN-based method for lung cancer detection in whole slide histopathology images. In 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech) (pp. 1-4). IEEE. https://doi.org/10.23919/SpliTech.2019.8783041

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR. https://doi.org/10.48550/arXiv.1905.11946

Veeling, B. S., Linmans, J., Winkens, J., Cohen, T., & Welling, M. (2018). Rotation equivariant CNNs for digital pathology. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11 (pp. 210-218). Springer International Publishing. https://doi.org/10.48550/arXiv.1806.03962

Wang, G., Luo, P., Lin, L., & Wang, X. (2017). Learning object interactions and descriptions for semantic image segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5859-5867). https://doi.org/10.1109/CVPR.2017.556

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3349-3364. https://doi.org/10.48550/arXiv.1908.07919

Wang, J., Xu, Z., Pang, Z. F., Huo, Z., & Luo, J. (2021). Tumor detection for whole slide image of liver based on patch-based convolutional neural network. Multimedia Tools and Applications, 80, 17429-17440. https://doi.org/10.1007/s11042-020-09282-x

Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognition, 90, 119-133. https://doi.org/10.48550/arXiv.1611.10080

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2881-2890). https://doi.org/10.48550/arXiv.1612.01105