Hybrid Deep Learning Architectures for Histological Image Segmentation

Md Jahid Hasan
Faculty of Engineering
Multimedia University
Cyberjaya, Malaysia
sawikot@gmail.com

Wan Siti Halimatul Munirah Wan Ahmad
Center for Visual Computing,
Faculty of Computing and Informatics;
Faculty of Engineering, Multimedia University
Cyberjaya, Malaysia
wan.sitihalimatul@mmu.edu.my

Mohammad Faizal Ahmad Fauzi
Faculty of Engineering
Multimedia University
Cyberjaya, Malaysia
faizal1@mmu.edu.my

Fazly Salleh Abas
Faculty of Engineering
and Technology
Multimedia University
Cyberjaya, Malaysia
fazly.salleh.abas@mmu.edu.my

***Abstract*—In histopathology image analysis, accurate segmentation of nuclei holds immense significance, particularly in the early detection and treatment of diseases like breast cancer. Nuclei segmentation is a fundamental but challenging task because of the intricate variations in nuclear shapes, sizes, densities, and overlapping instances. In this paper, we evaluate eight convolutional neural network (CNN) models, two of them existing models namely U-Net, SegNet, and six hybrid models by combining U-Net and SegNet modify decoder with ResNet, VGG and DenseNet (ResNet-UNet, ResNet-SegNet, VGG-UNet, VGG-SegNet, DenseNet-UNet, and DenseNet-SegNet. This experiment aims to identify the best deep-learning model for segmenting hematoxylin and eosin (H&E) stain images using a publicly available dataset called MoNuSeg. From the experimented work, we found that VGG-UNet outperforms other models with an F1 score of 0.8452 and IoU of 0.6929 respectively. This research will serve as a foundation for the future construction of more complex deep learning models with cascade or any combination of the models studied.**

***Keywords*—Cancer, Histopathology Image, Deep Learning, Nuclei Segmentation, H&E**

I. Introduction

Cancer is a deadly disease that affects millions of people worldwide [1]. It's not a single illness but rather a collection of diseases characterized by the abnormal growth of cells that have the potential to invade or spread to other parts of the body. From breast and lung to skin and blood cancers, this condition encompasses a vast array of types, each with its unique characteristics and challenges. Digital pathology plays a crucial role in the diagnosis, prognosis, and treatment of cancer [2]. By digitizing tissue samples and utilizing advanced imaging techniques, digital pathology enables pathologists to analyze and interpret these samples with unprecedented precision and efficiency. Moreover, digital pathology helps personalize treatments by detailing tumor characteristics and guiding tailored therapies. Overall, it transforms cancer understanding, diagnosis, and treatment.

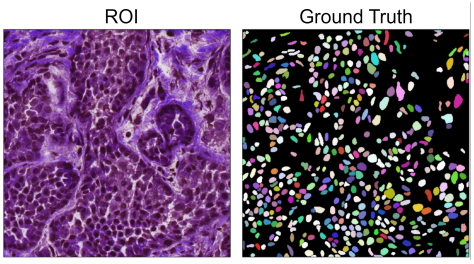

Fig. 1. Sample images from MonuSeg dataset: the stain images and their corresponding ground truth labels.

Histopathology is a specialized field within pathology that involves the microscopic examination of tissues to study the manifestations of diseases [3]. Typically performed on tissue samples obtained through procedures such as biopsies or surgical resections, histopathology is integral to the accurate diagnosis and characterization of various medical conditions, including cancer. Through the analysis of cellular and tissue structures, histopathologists gain crucial insights into the nature and extent of diseases, aiding in the formulation of effective treatment strategies. Hematoxylin and Eosin (H&E) staining, is a fundamental and widely used technique in histopathology. H&E staining is essential for providing contrast in histological sections, allowing pathologists to observe tissue architecture, identify different cell types, and evaluate the presence of abnormalities or diseases. It serves as the cornerstone of histopathological examination, playing a pivotal role in routine pathology reporting and aiding in the comprehensive understanding of various medical conditions. In summary, histopathology, utilizing techniques like H&E staining, unravels disease complexities at cellular and molecular levels. It guides tailored treatment based on tissue and cell specifics. The combination of morphological and

molecular information derived from histopathological analyses is paramount in advancing our understanding of diseases and improving patient care. In Figure 1, an illustration is provided depicting a H&E image alongside its corresponding ground truth.

Segmentation within digital pathology is a critical step that delineates specific regions or structures within tissue samples, playing a pivotal role in cancer diagnosis and treatment. Preceding the application of machine learning or deep learning algorithms for classification [4] and the automated calculation of cancer status scores [5], a critical step involves nuclei segmentation. By precisely identifying areas of interest such as tumor boundaries, cell types, or areas of significance like necrosis, segmentation provides pathologists and algorithms with a clear view of the intricate details within the sample. This process enables quantitative analysis, offering insights into tumor characteristics, grading, and molecular profiles that are essential for personalized treatment plans. Automated segmentation algorithms streamline this process, enhancing efficiency and expediting decision-making. Moreover, Segmentation helps extract vital data for new diagnostic methods and therapies, advancing our fight against cancer.

The framework proposed by the authors in [6] introduces a method for weakly supervised segmentation based on partial points annotation. Their two-stage approach commences with semi-supervised training of a detection model, utilizing partially labeled nuclei locations. The subsequent stage focuses on weakly supervised segmentation, employing coarse labels derived from detected nuclei locations. Naylor et al. [7] presented a deep learning-based approach for segmenting nuclei in H&E stained tissue slides. This method formulates the segmentation task as a regression of intra-nuclear distance maps, implicitly incorporating a shape prior into the segmentation network. The approach surpasses previous methods and includes a meticulously annotated dataset for nuclei segmentation. Xie et al. [8] proposed an instance segmentation approach that integrates Deep Convolutional Neural Networks (CNN) with Marker-controlled Watershed to handle overlapping nuclei in histopathology images. Their Deep Interval-Marker-Aware Network addresses multiple segmentation tasks and utilizes the interval between overlapping nuclei to refine segmentation results. Another contribution on CNN by [9] introduces fully convolutional networks with multiple encoders and deep fusion for enhanced segmentation. The multi-scale architectures outperform single-scale U-Nets, benefiting from both local and global contexts. In [10], the authors advocate for incorporating the geometric structure of the special Euclidean motion group SE(2) in CNN for translation and rotation equivariance. This framework undergoes evaluation in three histopathology image analysis tasks, including mitosis detection, nuclei segmentation, and tumor detection. Leveraging the unique optical properties of H&E staining images, [11] proposes a Hematoxylin-aware CNN model for nuclei segmentation. The Triple U-net structure integrates RGB, Hematoxylin, and Segmentation branches with a feature aggregation strategy, outperforming state-of-the-art methods. Wang et al. [12] introduce RCSAU-Net, a novel segmentation method utilizing an improved U-Net architecture and GAN-based learning to enhance nuclei segmentation accuracy and address over- and under-segmentation issues. The integration of adversarial training enhances contextual semantics, resulting in superior performance on MoNuSeg and PanNuke datasets, showcasing robust generalization capabilities.

This study is an extension of our recent research in [13], wherein we introduced two methods, ResNet-SegNet and ResNet-UNet, derived from the well-known SegNet and UNet approaches. Our earlier findings revealed that ResNet-UNet outperformed in offline evaluations on ER-IHC stain images. We developed a real-time application based on our proposed models, employing PR-IHC stain images despite training our models with ER-IHC stain images. We conducted evaluations in both offline (using the test set) and online (using real-time setup) scenarios, presenting our findings accordingly. In this work, we leverage both SegNet and UNet methods to propose our own hybrid approaches using VGG, DenseNet and ResNet, for histological image segmentation. The output model will be named as VGG-UNet, VGG-SegNet, DenseNet-UNet, DenseNet-SegNet, ResNet-UNet, and ResNet-SegNet. The models are trained using the publicly available H&E dataset (Monuseg [14]), expanding on our previous paper, which solely focused on the ER-IHC dataset.

The paper follows this structure: Section II furnishes a comprehensive overview of the methodology utilized. Section III delves into the presentation of results and discussions. Lastly, Section IV encapsulates the conclusions derived from our findings.

II. Methodology

The flowchart in Fig. 2 illustrates the sequential steps involved in the training and testing workflow. Initially, the dataset images undergo patch to a standardized dimension of 500 by 500 pixels from 1000 by 1000. Then the image resizes to 512 by 512 pixels. The eight deep learning models, U-Net, SegNet, VGG-UNet, VGG-SegNet, DenseNet-UNet, DenseNet-SegNet, ResNet-UNet, and ResNet-SegNet architectures, are then trained on the prepared datasets. Following the training phase, the models are subjected to evaluation using the test set derived from the image database. Two distinct types of evaluations are conducted to assess the performance of the trained models. The entire process is aimed at optimizing the models for accurate and effective image segmentation.

A. Dataset

The proposed model was evaluated using a well-known publicly available dataset called MoNuSeg [14], and then compared with the existing state-of-the-art methods. The Multi-Organ (MoNuSeg) dataset comprises 30 H&E stained histopathology images for training, each of size 1000×1000, with annotations for 21,623 nuclear boundaries. Captured at a 40x magnification factor, these images feature patients selected from the TCGA website, originating in various hospitals and representing diverse cancer types. The MoNuSeg dataset is

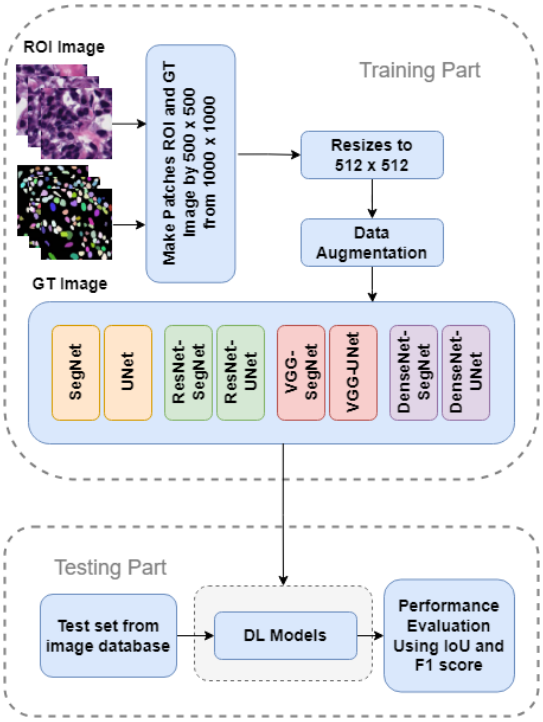

Fig. 2. Image segmentation workflow

particularly valuable for its diversity in nuclear appearances and practicality, encompassing images from seven organs (Breast, Stomach, Liver, Prostate, Kidney, Colon, and Bladder) belonging to 30 different patients. Fig. 1 illustrates examples of Regions of Interest (RoI) and Ground Truth for one sample image.

The test dataset used in the evaluation consists of images from organs that were not included in the training set, making it even more challenging than the training set. The dataset includes 14 RoIs data for testing additionally, with each cancer type represented by two RoIs. The testing data size is also the same as the training dataset around 1000 by 1000. During both training and testing procedures for semantic segmentation, we patch the images into 500x500 segments and resize them to 512x512.

B. Deep Learning Models

We use two existing models, U-Net and SegNet, and another six hybrid models by combining U-Net and SegNet modify decoder with ResNet, VGG and DenseNet (ResNet-UNet, ResNet-SegNet, VGG-UNet, VGG-SegNet, DenseNet-UNet, and DenseNet-SegNet) shown in Table I to train and test our dataset. In Fig. 3 shows how we produced hybrid VGG-UNet and VGG-SegNet from UNet and SegNet. Other architectures like ResNet-UNet, ResNet-SegNet, DenseNet-UNet, and DenseNet-SegNet are made in the same way. UNET and SegNet are both popular architectures in the field of semantic segmentation, a computer vision task that involves classifying and segmenting objects within an image. While they share the common goal of pixel-wise segmentation, they differ in their underlying architectures and design philosophies. UNET, introduced by Ronneberger et al. [15], is characterized by a U-shaped network structure, featuring an encoder path for capturing contextual information and a decoder path for precise localization. On the other hand, SegNet, proposed by Badrinarayanan et al. [16], employs an architecture based on an encoder-decoder framework, but it leverages a different strategy by using pooling indices from the encoder for upsampling in the decoder.

1) Overview of U-Net architecture: The U-Net architecture consists of an encoder, a bottleneck layer, a decoder, and an output layer. The encoder, forming the top part of the U shape, captures context and features from the input image through convolutional layers and max-pooling. A bottleneck layer refines these features, connecting the encoder to the decoder. The decoder, forming the bottom part of the U shape, upsamples the features to the original input size using transposed convolution operations. Skip connections play a vital role by linking corresponding layers in the encoder and decoder, aiding in the recovery of fine details and maintaining spatial information. The final output layer produces a segmentation mask, a pixel-wise classification map, with activation functions like softmax or sigmoid depending on the segmentation task. Skip connections, also known as residual connections, enhance the model's ability to precisely localize and segment objects in the image.

2) Overview of SegNet:: SegNet's encoder consists of multiple convolutional layers, similar to other neural networks, extracting hierarchical features from the input image. What sets SegNet apart is its use of max-pooling indices, recording locations of maximum values during pooling. These indices are stored and later used in the decoder for upsampling. Like U-Net, SegNet may have a bottleneck layer capturing abstract features. The decoder, responsible for upsampling, uses a unique "unpooling" operation with stored max-pooling indices, maintaining spatial information effectively. The output layer classifies pixels into predefined categories based on the segmentation task, using an appropriate activation function. Unlike U-Net, SegNet doesn't have skip connections; instead, it relies on max-pooling indices for precise upsampling, aiding in spatial detail recovery.

3) Overview of VGG-UNet and VGG-SegNet architecture: VGG16 is a popular neural network known for its effectiveness in classifying images. It was created by the Visual Geometry Group at the University of Oxford and gained recognition in the ImageNet competition [17]. In the VGG-UNet architecture, VGG acts as the encoder, extracting features from the input image and reducing its size. The decoder, following the U-Net design, uses transposed convolutions for upsampling and includes skip connections to combine information from both VGG and U-Net. Our implementation simplifies the U-Net decoder, using only one convolutional layer per block to reduce parameters and computation time. The final decoder layer creates a segmentation map by classifying each pixel. In VGG-SegNet, VGG serves as the encoder, extracting features for segmentation. The decoder follows SegNet, preserving

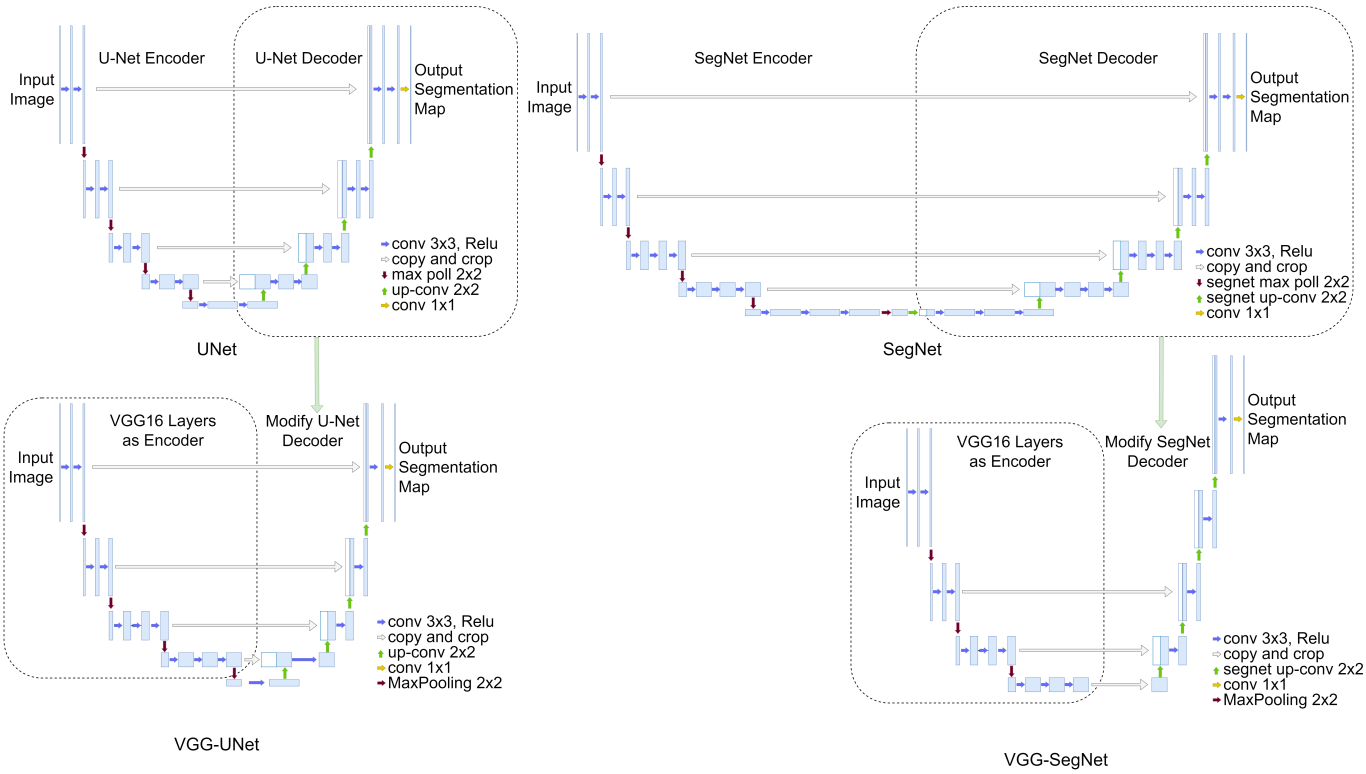

Fig. 3. Proposed models architecture.

spatial details with max-pooling indices. Our implementation maintains the SegNet-like structure, using one convolutional layer per block to reduce parameters. Unlike the original SegNet, our model has three skip layers instead of four. The final decoder layer generates the segmentation map by classifying pixels into categories using an appropriate activation function.

4) Overview of DenseNet-UNet and DenseNet-SegNet architecture: DenseNet-169, a variant of the DenseNet architecture renowned for its effectiveness in image classification, was developed by Gao Huang [18]. In the DenseNet-UNet setup, DenseNet acts as the encoder, extracting hierarchical features efficiently with dense connectivity. The decoder, inspired by U-Net, employs transposed convolutions and skip connections for upsampling, streamlining the process with a single convolutional layer per block for reduced parameters and enhanced computational efficiency. The final layer produces a segmentation map, categorizing each pixel. In the DenseNet-SegNet configuration, DenseNet serves as the encoder, capturing hierarchical features, while the decoder follows the SegNet architecture with transposed convolutions and max-pooling indices for accurate upsampling. Our implementation maintains a SegNet-like structure, using one convolutional layer per block and 3 skip layers, contributing to parameter reduction and computational efficiency. The decoder's output generates a segmentation map, employing an appropriate activation function for pixel classification.

5) Overview of ResNet-UNet and ResNet-SegNet architecture: ResNet-50, a key member of the ResNet family introduced by Kaiming He et al. in 2015 [19], has made a significant impact on deep learning and computer vision with its effectiveness in training deep neural networks. In the ResNet-UNet architecture, ResNet serves as the encoder, capturing hierarchical features through residual blocks with skip connections that address vanishing gradient challenges. The decoder, following the U-Net design, utilizes transposed convolutions and skip connections to link encoder and decoder layers. Our implementation simplifies the U-Net decoder by using a single convolutional layer per block, reducing parameters and computational overhead. Similarly, in ResNet-SegNet, the ResNet component excels as the feature extractor, leveraging residual blocks with skip connections for training deep networks. The decoder, inspired by SegNet, employs transposed convolutions and utilizes max-pooling indices for precise upsampling. Our model maintains a SegNet-like structure with one convolutional layer per block, reducing parameters and computational costs, and features 3 skip layers instead of 4. The final decoder layer classifies pixels into categories, producing the segmentation map with an appropriate activation function.

C. Data Augmentation

Data augmentation techniques are crucial for making models more versatile. There are three main approaches: randomly flipping and rotating images to help models recognize objects from various angles, adjusting brightness randomly to adapt to different lighting conditions, and adding Gaussian noise to

TABLE I
SEGMENTATION MODELS USED IN THE EXPERIMENT.

Model Name	Encoder	Decoder	Parameters (In Millions)	Training Time (1 epoch)
SegNet	Vanilla CNN	SegNet	33.3M	459s
U-Net	Vanilla CNN	U-Net	31.0M	484s
ResNet-SegNet	ResNet	SegNet	16.3M	136s
ResNet-Unet	ResNet	U-Net	16.3M	205s
VGG-SegNet	VGG	SegNet	12.3M	243s
VGG-Unet	VGG	U-Net	12.3M	209s
DenseNet-SegNet	DenseNet	SegNet	11.9M	346s
DenseNet-Unet	DenseNet	U-Net	11.9M	328s

simulate real-world imperfections and improve performance in noisy environments. Together, these strategies enhance a model's ability to handle diverse situations beyond its training data, ultimately improving its performance in tasks like image classification, segmentation, and object detection in real applications.

D. Evaluation Method

Evaluative methods such as precision, recall, F1 score, and Jaccard index play a crucial role in evaluating the performance.

Precision is the ratio of correctly predicted positive observations to the total predicted positives. It assesses the accuracy of positive predictions. Recall, also known as sensitivity or true positive rate, is the ratio of correctly predicted positive observations to all observations in the actual class. It measures the ability of the model to capture all the relevant cases. The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall, especially when there is an uneven class distribution.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision + Recall}}$$

The Jaccard Index, also known as the Jaccard similarity coefficient (IoU), is a measure of similarity between two sets. It's defined as the size of the intersection of the sets divided by the size of the union of the sets. Mathematically, it can be expressed as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

III. RESULTS AND DISCUSSIONS

The MoNuSeg dataset serves as a crucial asset for appraising the efficacy of diverse image segmentation models in precisely identifying objects within images. Our comprehensive assessment encompasses eight distinct models: SegNet, UNet, ResNet-SegNet, ResNet-UNet, VGG-SegNet, VGG-UNet, DenseNet-UNet, and DenseNet-SegNet. And three more existing models UCTransNet-pre [20], ATTransUNet [21], and CT-Net [22] where their author applies the same dataset we use. Our scrutiny is specifically directed toward key metrics, including the IoU, F1 score, precision, and recall. The evaluative methods results of all the models are detailed in Table II.

TABLE II
PERFORMANCE METRICS FOR VARIOUS MODELS FOR MONUSEG DATASET

Model Name	IoU	F1	Precision	Recall
SegNet [16]	0.5653	0.7530	0.7923	0.7380
UNet [15]	0.6076	0.7545	0.9203	0.6381
ResNet-SegNet	0.6499	0.7928	**0.9374**	0.6930
ResNet-UNet	0.6703	0.8213	0.8756	0.7858
VGG-SegNet	0.6745	0.8373	0.8643	**0.8200**
VGG-UNet	**0.6929**	**0.8452**	0.8890	0.8110
DensNet-UNet	0.6611	0.8050	0.9194	0.7230
DensNet-SegNet	0.6426	0.7945	0.9065	0.7195
UCTransNet-pre [20]	0.6550	0.7908	-	-
ATTransUNet [21]	0.6551	0.7916	-	-
CT-Net [22]	0.665	0.798	-	-

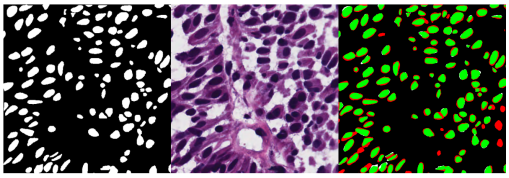

Fig. 4. Some highest segmentation result with RoI and Ground Truth for VGG-UNet in MoNuSeg dataset

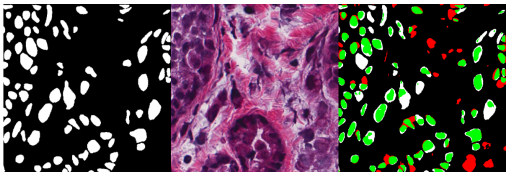

Fig. 5. Some lowest segmentation result with RoI and Ground Truth for VGG-UNet in MoNuSeg dataset

In comparing the performance of various semantic segmentation models, VGG-UNet stands out with the highest IoU of 0.6929, indicating excellent object localization. It achieves a high F1 score of 0.8452, showcasing a strong balance between precision (0.8890) and recall (0.8110). Following closely is VGG-SegNet, blending VGG16 with SegNet architecture, with an IoU of 0.6745 and a well-balanced F1 score of 0.8373. ResNet-UNet outperforms ResNet-SegNet, demonstrating improved object localization with an IoU of 0.6703 and a higher F1 score of 0.8213. DenseNet-UNet achieves IoU values of 0.6497 and 0.6611, respectively, while DenseNet-SegNet lags slightly with an IoU of 0.6426. SegNet and UNet exhibit the lowest performance, achieving an IoU of 0.5653 and 0.6076, along with F1 scores of 0.7530 and 0.7545 respectively. Overall, these models exhibit varying strengths in object localization, precision, and recall, highlighting the importance of architectural choices in semantic segmentation tasks. Additionally, the existing models - UCTransNet-pre, ATTransUNet, and CT-Net - also exhibit promising performance.

UCTransNet-pre and ATTransUNet demonstrate similar performance with an IoU of around 65.50 and F1 scores of around 79.08-79.16. CT-Net showcases an IoU of 0.665 and an F1 score of 0.798. These models contribute to the spectrum of available options for semantic segmentation on the MoNuSeg dataset, displaying competitive performance in comparison to the established architectures.

In summary, the discussion highlights that DenseNet-SegNet, DenseNet-UNet, and ResNet-UNet exhibit robust performance across various metrics, making them versatile choices for a wide range of segmentation tasks on this dataset. The VGG-UNet model emerges as the top performer on the MoNuSeg dataset, boasting the highest IoU, F1 score, precision, and a commendable recall rate. Figures in Fig. 4 showcase impressive scores, while Fig. 5 displays images with lower scores. However, the optimal model choice depends on specific task requirements and the desired precision-recall trade-off. Each model exhibits unique strengths and areas for improvement, offering valuable insights for selecting the most suitable model for image segmentation tasks within the MoNuSeg dataset.

IV. Conclusions

Segmenting nuclei is a demanding yet crucial task in computer-assisted histopathological analysis. This paper conducts a comparative analysis of deep learning models, forming the foundation for advanced and intricate deep model networks. The objective of this experiment is to determine the optimal deep-learning model for segmenting hematoxylin and eosin (H&E) stain images. The model is trained and evaluated using a publicly accessible dataset named MoNuSeg with H&E stain. In terms of performance, VGG-UNet demonstrates an F1 score of around 0.8452 and IoU of 0.6929. On the other hand, ResNet-UNet attains an F1 score of 0.8213 and IoU of 0.6703. These models hold potential for future endeavors in developing segmentation algorithms catering to various staining types and for real-time processing tasks such as cancer case classification, scoring, and grading.

Acknowledgment

This project received funding from the Ministry of Higher Education Malaysia (MOHE) under the Research Excellence Consortium (KKP-2020) for the project "Artificial Intelligence for Digital Pathology (AI4DP)".

References

[1] National Academies of Sciences, National Academies of Engineering, National Academies of Medicine *et al.*, "Guiding cancer control: A path to transformation," 2019.

[2] R. Pell, K. Oien, M. Robinson, H. Pitman, N. Rajpoot, J. Rittscher, D. Snead, C. Verrill, U. N. C. R. I. N. C.-M. P. C.-P. quality assurance working group, O. J. Driskell *et al.*, "The use of digital pathology and image analysis in clinical trials," *The Journal of Pathology: Clinical Research*, vol. 5, no. 2, pp. 81–90, 2019.

[3] X. Y. Liew, N. Hameed, and J. Clos, "A review of computer-aided expert systems for breast cancer diagnosis," *Cancers*, vol. 13, no. 11, p. 2764, 2021.

[4] W. S. H. M. W. Ahmad, M. J. Hasan, M. F. A. Fauzi, J. T. Lee, S. Y. Khor, L. M. Looi, and F. S. Abas, "Nuclei classification in er-ihc stained histopathology images using deep learning models," in *TENCON 2022-2022 IEEE Region 10 Conference (TENCON)*. IEEE, 2022, pp. 1–5.

[5] W. Ahmad, M. Fauzi, M. Hasan, Z. Rehman, J. Lee, S. Khor, L. Looi, F. Abas, A. Adam, E. Chan *et al.*, "Multi-configuration analysis of densenet architecture for whole slide image scoring of er-ihc," *IEEE Access*, 2023.

[6] H. Qu, P. Wu, Q. Huang, J. Yi, Z. Yan, K. Li, G. M. Riedlinger, S. De, S. Zhang, and D. N. Metaxas, "Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images," *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3655–3666, 2020.

[7] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 448–459, 2018.

[8] L. Xie, J. Qi, L. Pan, and S. Wali, "Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images," *Neurocomputing*, vol. 376, pp. 166–179, 2020.

[9] R. Schmitz, F. Madesta, M. Nielsen, J. Krause, S. Steurer, R. Werner, and T. Rösch, "Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture," *Medical image analysis*, vol. 70, p. 101996, 2021.

[10] M. W. Lafarge, E. J. Bekkers, J. P. Pluim, R. Duits, and M. Veta, "Roto-translation equivariant convolutional networks: Application to histopathology image analysis," *Medical Image Analysis*, vol. 68, p. 101849, 2021.

[11] B. Zhao, X. Chen, Z. Li, Z. Yu, S. Yao, L. Yan, Y. Wang, Z. Liu, C. Liang, and C. Han, "Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation," *Medical Image Analysis*, vol. 65, p. 101786, 2020.

[12] H. Wang, G. Xu, X. Pan, Z. Liu, R. Lan, and X. Luo, "Multi-task generative adversarial learning for nuclei segmentation with dual attention and recurrent convolution," *Biomedical Signal Processing and Control*, vol. 75, p. 103558, 2022.

[13] M. J. Hasan, W. S. H. M. W. Ahmad, M. F. A. Fauzi, J. T. H. Lee, S. Y. Khor, L. M. Looi, and F. S. Abas, "Real-time segmentation of ihc images from microscope using deep learning architecture," in *2023 IEEE 2nd National Biomedical Engineering Conference (NBEC)*. IEEE, 2023, pp. 25–30.

[14] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[15] O. Ronneberger, P. Fischer, and T. Brox, "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*. Springer, 2015, pp. 234–241.

[16] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[20] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, "Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 3, 2022, pp. 2441–2449.

[21] X. Li, S. Pang, R. Zhang, J. Zhu, X. Fu, Y. Tian, and J. Gao, "At-transunet: An enhanced hybrid transformer architecture for ultrasound and histopathology image segmentation," *Computers in Biology and Medicine*, vol. 152, p. 106365, 2023.

[22] N. Zhang, L. Yu, D. Zhang, W. Wu, S. Tian, X. Kang, and M. Li, "Ct-net: Asymmetric compound branch transformer for medical image segmentation," *Neural Networks*, 2023.