Deep Learning-Based Approaches for Nucleus Segmentation

Duy Cuong Bui*, Myungsik Yoo°

ABSTRACT

The accurate identification of cell nuclei is a critical aspect of various analyses, given that human cells, numbering around 30 trillion, contain DNA as their genetic code. In this research paper, we provide a comprehensive overview of deep learning-based techniques for nucleus segmentation. We have replicated and assessed the state-of-the-art methods using datasets like FCN, SegNet, U-net, and DoubleU-net, with a focus on the Data Science Bowl 2018 dataset comprising 670 training data folders and 65 testing data folders. Our experimental findings reveal that DoubleU-Net surpasses U-Net and other baseline models, yielding more precise segmentation masks. This promising outcome suggests that DoubleU-Net could serve as a robust model for addressing various challenges in medical image segmentation.

Key Words : Deep Learning, CNN, Nuclei Segmentation, Image Segmentation, U-Net

Ⅰ. Introduction

The segmentation of cell nuclei is a crucial step in the analysis of biomedical microscopy images[1]. It serves as the foundation for various medical analyses like cell counting[2] and cell type classification[3]. However, this task is quite challenging due to variations in staining, tissue types, and the diverse visual characteristics of different cell types, all of which affect how nuclei appear[1]. Manual segmentation of nuclei is not only time-consuming but also becomes impractical when dealing with large datasets. Furthermore, the accuracy of segmentation results relies heavily on the expertise of individuals and is often not reproducible. Consequently, there is a significant demand for automated methods for instance-based nuclei segmentation in microscopy images.

In the literature, various techniques have been proposed for automating the segmentation of nuclei, spanning from basic background subtraction to more advanced methods. These include approaches like the Otsu-based method[4], the watershed method[5], Grab Cut[6], and active contour[7]. However, these conventional methods have their limitations. They are often sensitive to the choice of parameters, and their effectiveness is typically limited to specific categories of structured nuclei.

In contrast, deep learning-based techniques have gained significant traction in the field of medical imaging, being applied to various applications such as medical image super-resolution[8], classification[9], and notably, medical imaging segmentation[10,11]. These approaches have also been extensively utilized for cell and nuclei segmentation, with numerous deep learning-based methods dedicated to this task[12]. For instance, Pan et al. introduced a deep semantic network designed for nuclei segmentation in pathological images[13]. Vuola et al. employed MaskRCNN to segment nuclei[14]. Van et al.
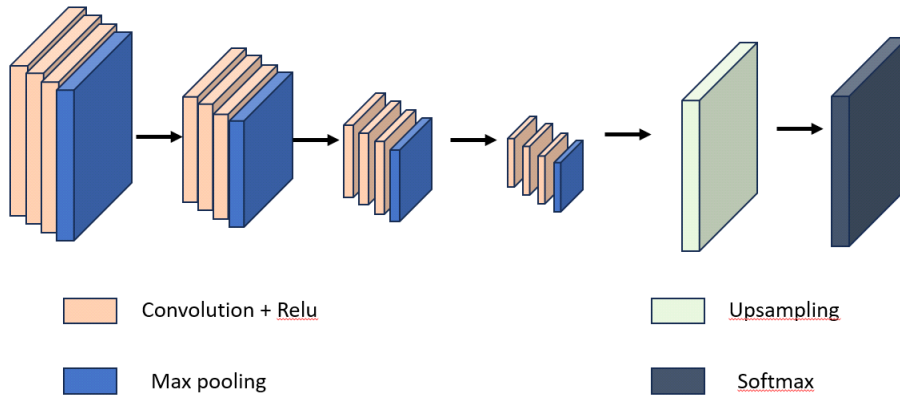
Fig. 1. FCN architecture

presented DeepCell, a method tailored for analyzing cells in live-cell imaging[15]. Ronneberger et al. proposed the U-Net model, which has become a popular choice for nuclei segmentation[16]. Zeng et al. developed a modified version of U-Net for nuclei segmentation, incorporating features like residual blocks and channel attention mechanisms[17]. Zhou et al. introduced CIA-Net, which utilizes two separate decoders for distinct tasks and incorporates a multilevel information aggregation module to capture dependencies between nuclei and their contours[18].

In the current study, we conduct a feasibility study on deep learning-based approaches for nuclei images convolution Network (FCN)[19], SegNet[20], UNet[16], DoubleUNet[21]. We demonstrate the utility of such models by evaluating them for cell segmentation and recommend which one is most suitable for nuclei segmentation tasks.

The paper is organized as follows: in Section II, we review several deep learning models for image segmentation. Section III presents the comparative results when apply deep learning reviewed in Section II. Conclusions of this paper are given in Section IV.

II. Materials and Methods

2.1 Fully Convolutional Neural Network

Long et al. introduced the Fully Convolutional Neural Network (FCN) for addressing semantic image segmentation challenges[22]. This architecture has since been extended to tackle various other segmentation tasks, including ventricle segmentation[19]. As presented in Fig.1, the network includes 15 convolution layers, 3 max pooling layers, followed by upsampling layers and a classifier layer.

The network is divided into two main parts, contracting path (also called encoder) and expanding path (called decoder). The encoder consists of convolutional layers used to preserve the spatial structure of the feature map and max pooling layers to reduce the resolution. In the improved version of FCN by Tran[19] each convolution layer is followed by a rectified linear unit (ReLU) and a mean-variance normalization (MVN). The purpose of MVN is to normalize the intensity distribution of the feature map so that the feature map pixel values have zero mean and unit variance. The decoder, symmetric to the encoder, consists of transposed convolution layers, and upsampling layers. The feature map in the decoder is combined with the corresponding feature map in the encoder to preserve spatial information that might be lost during pooling operations. Finally, a classifier such as softmax is used to produce class probabilities for each pixel in the image to be segmented.

2.2 SegNet

The SegNet architecture was proposed by Badrinarayanan et al.[20] for semantic pixel-wise segmentation. Similar to the FCN architecture, the
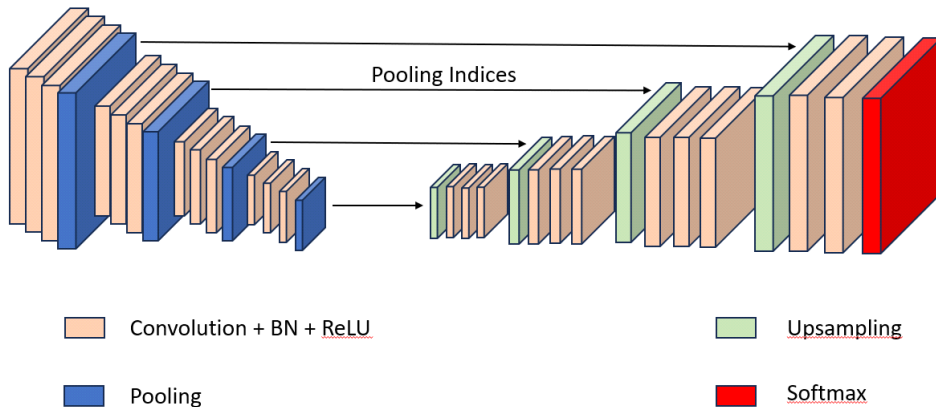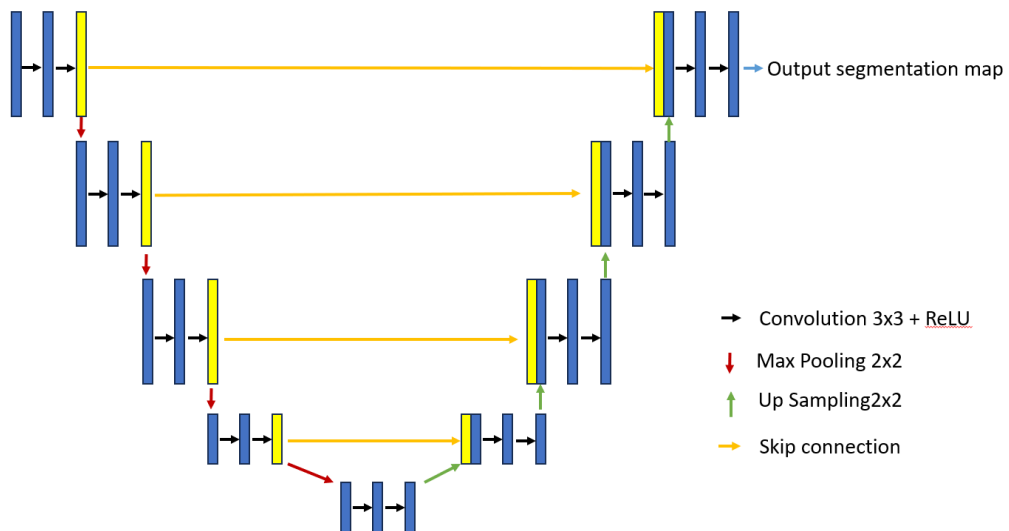
Fig. 2. SegNet architecture

Fig. 3. U-Net architecture

SegNet consists of an encoder path account for downsampling with convolution and max-pooling layers, and a decoder path used for upsampling the feature maps to the same size of the input image. To retain information lost due to pooling, Ninh et al.[23], proposed the skip connection mechanism to fuse between the encoder to decoder of the SegNet. In addition, compared to the original SegNet model, their model has also fewer learned parameters since the downsampling and upsampling layers are reduced. The improved version of the SegNet architecture is presented in Fig. 2.

2.3 U-Net

The U-Net neural network architecture, introduced by Ronneberger and colleagues in reference[16], has become widely recognized as the standard approach for medical image segmentation tasks. It draws inspiration from the Fully Convolutional Network (FCN)[19] and can be conceptually divided into two main sections: the contracting part (encoder) and the expanding part (decoder). The contracting part employs a combination of convolutional layers and max-pooling operations to achieve downsampling, while the expanding part involves upsampling and convolutional layers. To preserve crucial spatial information that might be lost during downsampling, U-Net employs skip connections, where feature

maps from the encoder are concatenated with their counterparts in the decoder at the same spatial resolution. This architectural design facilitates accurate segmentation of medical images.

2.4 DoubleU-Net

The DoubleU-Net has recently been proposed by Jha et al.[21] shown in Fig 4. As can be seen from this figure, DoubleU-Net starts with an VGG-19 pre-trained network as an encoder in the first sub-network, and then follows by an astrous spatial pyramid pooling (ASPP) block and finally is the decoder. The input image is passed through the first sub-network, then the first output is created. This output then multiplies with the input, the result becomes the input of the second sub-network and the second output is created. Finally, these two outputs are concatenated and become the output of DoubleU-Net. What makes DoubleU-Net differ from U-Net is that there are two subnetworks. Squeeze-and-Excitation block is used inside to enhance the power of convolution layer. Moreover, the skip connection is performed from the first encoder to the first decoder and from either the first encoder or the second encoder to the second decoder, which maintains the spatial resolution and enhances the quality of the output feature maps.
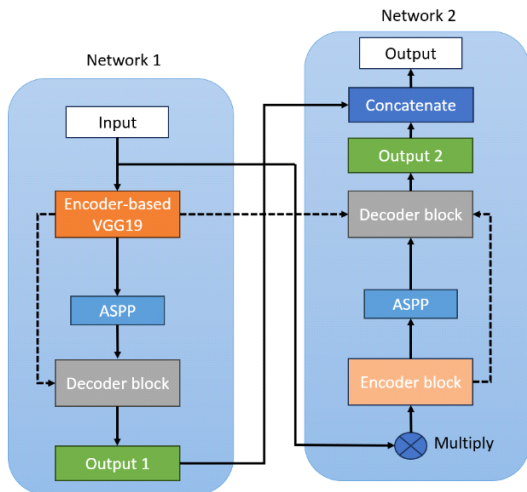

Fig. 4. DoubleU-Net architecture

2.5 Loss function

For deep learning-based nuclei image segmentation, the binary cross-entropy function and Dice loss are widely used to train the networks. Let y be the predicted map by the network, $\hat{y}$ be the corresponding label map, and N be the number of pixels of the maps. The Binary Cross Entropy loss is express as:

$$L_{BCE} = -\frac{1}{N}\sum(\hat{y}\,logy + (1-\hat{y})\log(1-y)\,)$$

(1)

The Dice loss is computed as

$$L_{Dice} = 1 - 2\frac{\sum \hat{y}y}{\sum \hat{y} + \sum y}$$

(2)

In our paper, we evaluate a loss function which is conjunction between the Binary Cross Entropy loss and Dice loss. While Dice loss captures the result between pixel overlap, BCE loss is associated with models that output probabilities. The combination of these two loss functions will help the optimization process become more general and converge better. Our loss function is express as

$$L = L_{BCE} + \frac{1}{2}L_{Dice}$$

(3)

Ⅲ. Experimental Results

3.1 Datasets

The Data Science Bowl 2018 (DSB2018) dataset presented a global challenge to scientists, tasking them with the automatic identification and segmentation of cells within a collection of microscopic images. The primary objective was to develop image segmentation techniques that could be universally applied across multiple experiments without the need for additional human intervention. This approach aimed to decrease the time required for image quantification, enabling future researchers to readily apply and evaluate various experiments for both research and clinical purposes.

The Data Science Bowl 2018 (DSB2018) dataset comprises a total of 670 training pairs and 65 testing
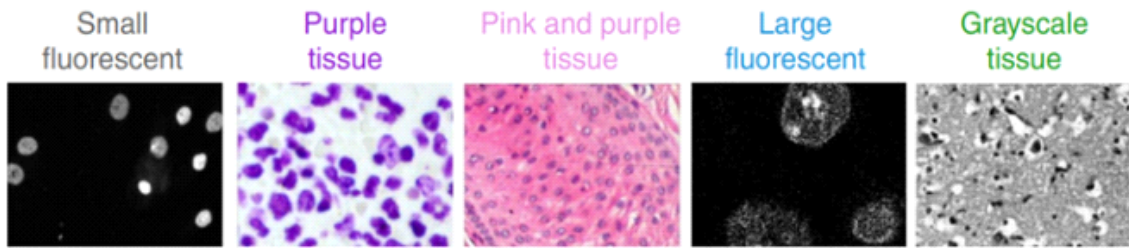
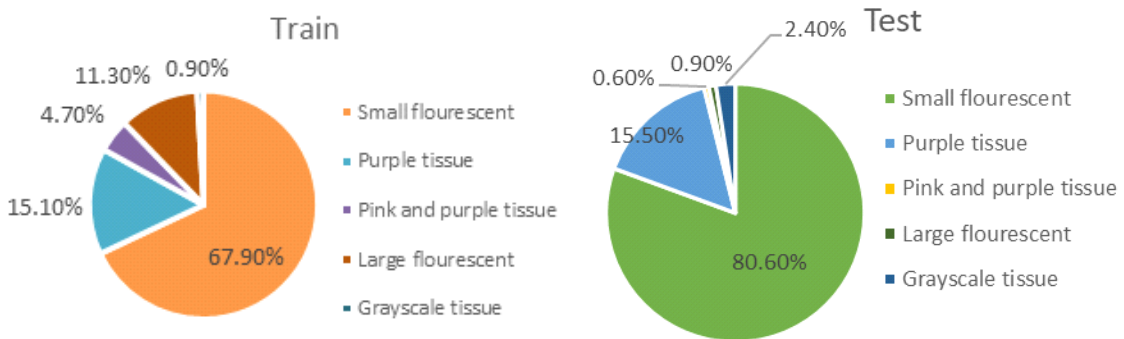Fig. 5. Five types of nuclei images in Data Science Bowl 2018 dataset

Fig. 6. Distribution of the dataset for training and testing.

pairs. Each pair consists of an image and its corresponding masks. Notably, the DSB2018 dataset encompasses five distinct types of cell images, including Small Fluorescent, Purple Tissue, Pink and Purple Tissue, Large Fluorescent, and Grayscale Tissue. These images exhibit variations in their content and characteristics, making the dataset diverse and challenging to work with, as depicted in Fig 5.

The distribution of data of nucleus types of the train and test sets of the DSB2018 dataset are given in Fig. 6.

3.2 Training

We implemented our model and utilized the Adam algorithm for optimizing the trainable parameters of the model with original learning rate 3×10^{-4}. The training process is looped on the dataset for 200 epochs with batch size 16 and the data is augmented after every epoch. Early stopping and ReducelronPlateau have already been used.

3.3 Evaluation metrics

We use Dice Similarity Coefficient (DSC) to evaluate the segmentation performance by the neural network. The DSC metric is defined as

$$DSC = \frac{2TP}{FN+FP+2TP} \qquad (4)$$

where TP, FN, FP represent for True Positive, False Negative and False Positive, respectively.

In addition to DSC, we also use the Intersection over Union (IoU) index as an alternative evaluation measure, defined as

$$IoU = \frac{TP}{TP+FP+FN} \qquad (5)$$

where TP, FN, FP represent for True Positive, False Negative and False Positive, respectively.

3.4 Results and validation

In this research paper, we undertake an investigation into the practicality of employing deep learning-based methods for the segmentation of nucleus images. Our focus is on evaluating the performance of various cutting-edge deep learning models on the Data Science Bowl 2018 (DSB2018)

dataset. The models we assess encompass the FCN, SegNet, U-Net, and DoubleU-Net architectures. To quantitatively assess their segmentation results, we present a comprehensive evaluation in Tab 1, which includes metrics like the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU).

As depicted in Table 1, it becomes evident that the DoubleU-Net model outshines its counterparts, achieving the highest DSC and IoU scores. In comparison to its predecessor model, U-Net, DoubleU-Net demonstrates remarkable improvements, surpassing it by 1.8% in DSC and 2.5% in IoU. Moreover, DoubleU-Net also exhibits substantial advancements in Recall and Precision scores, surpassing U-Net by 7.9% and 2.3%, respectively.

Furthermore, when pitted against the FCN and SegNet models, DoubleU-Net exhibits superior performance in terms of DSC, IoU, and Precision. These findings position DoubleU-Net as the standout model, delivering the best overall performance among all the state-of-the-art methods we evaluated for nucleus image segmentation.

However, the pursuit of higher performance inevitably introduces greater complexity into the model. In the case of DoubleU-Net, it comprises two pairs of encoder-decoder structures, and each of these encoder-decoder pairs is essentially a U-Net model. While this complexity is instrumental in achieving the impressive segmentation results mentioned earlier, it does come with the trade-off of reduced computational efficiency. In contrast to U-Net and other models like FCN and SegNet, DoubleU-Net's computational speed is notably slower. This means that while DoubleU-Net excels in accuracy and segmentation quality, it may not be the ideal choice in scenarios where real-time or near-real-time processing is a critical factor, as its computational demands are more substantial due to its intricate architecture.

In Table 2, we delve into a detailed experiment to compare the output of the two subnetworks in DoubleU-Net, when every network can be considered as an U-Net model. We can observe that the output from Network 2 is more accurate than the output from Network 1, which confirms the preeminence of DoubleU-Net in the Tab 1 when developed from a single U-Net network. This all-encompassing approach allows us to draw robust and well-founded conclusions regarding the

Table 1. Quanlitative results on Data Science Bowl 2018 dataset (The bests are in bold)

Methods	DSC	IoU	Recall	Precision	Time (ms)
FCN	90.2%	82.2%	77.3%	90.0%	25.4
SegNet	90.0%	81.9%	**89.1%**	88.9%	20.1
U-Net	89.5%	81.5%	78.7%	92.2%	**16.6**
DoubleU-net	**91.3%**	**84.0%**	86.6%	**94.5%**	29.3

Table 2. Qualitative results of Output 1 and Output 2 in the DoubleU-Net architecture.

DoubleU-Net	DSC	IoU	Recall	Precision
Output 1	90.4%	82.6%	82.0%	91.1%
Output 2	91.3%	84.0%	86.6%	94.5%

Table 3. Comparison between different loss functions. (DSC score)

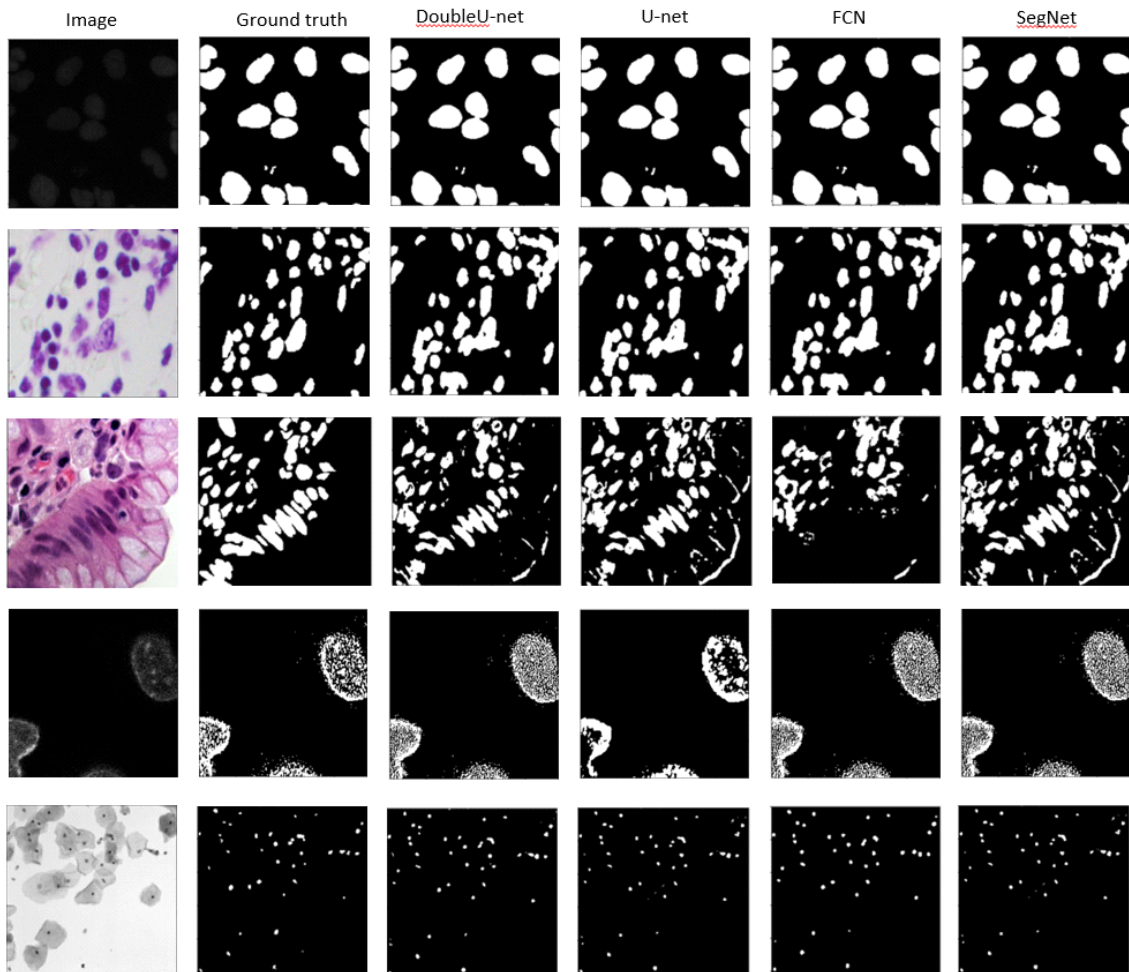| Methods	FCN	SegNet	U-Net	DoubleU-net
Dice loss	86.5%	85.9%	84.1%	88.3%
BCE loss	90.0%	89.6%	88.9%	91.0%
Our	90.2%	90.0%	89.5%	91.3%

Fig. 7. Quantitative segmentation results by state of the arts neural networks on the DSB2018 dataset

DoubleU-Net's capability to handle nuclei image segmentation tasks, helping to improve model performance and provide more accurate predictions.

In Table 3, we showcase the results achieved through the fusion of the Binary Cross-Entropy Loss and Dice Loss functions, resulting in improved model performance. Through a thorough analysis of the enhancements observed in each model, we can make informed conclusions regarding the appropriateness and effectiveness of the selected loss function. It's worth noting that the BCE loss function utilized in the study by Jha et al.[21] may encounter challenges related to sigmoid saturation, wherein the model's output reaches extreme values close to 0 or 1. Consequently, the incorporation of

the Dice loss serves as a countermeasure to alleviate this effect. The results are shown in Table 3.

To provide better evidence of higher performance, we create visual representations of segmented results produced by the aforementioned models in Fig 7. Upon careful scrutiny of this figure, it becomes apparent that the segmentation masks generated by DoubleU-Net display the closest resemblance and alignment with the ground truth annotations. In other words, when comparing the model's output to the actual reference data, DoubleU-Net consistently demonstrates a superior ability to accurately identify and delineate the nuclei in the images. This visual evidence further reinforces the notion that

DoubleU-Net outperforms other models in terms of segmentation quality and precision.

Ⅳ. Conclusion

In this study, we have explored the feasibility of utilizing deep learning-based methods for the segmentation of nuclei images. We conducted experiments using several state-of-the-art models on the Data Science Bowl 2018 dataset. Our results and evaluations revealed that the DoubleU-Net model achieved superior segmentation performance when compared to other state-of-the-art alternatives. Looking ahead, our future research will be directed towards reducing the training time or enhancing the accuracy of the DoubleU-Net for nuclei image segmentation, as well as extending its application to other segmentation tasks.

Additional strategies to enhance the approach's performance may involve considering alternatives like switching to a different pre-trained encoder, adopting a more contemporary neural network architecture, or seeking out an improved loss function.

References

[1] J. C. Caicedo, J. Roth, A. Goodman, T. Becker, K. W. Karhohs, C. McQuin, S. Singh, and A. E. Carpenter, "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images," *Cytometry A*, vol. 95, no. 9, pp. 952-965, 2019. (https://doi.org/10.1002/cyto.a.23863).

[2] T. Falk, et al., "U-Net: Deep learning for cell counting detection and morphometry," *Nature Meth.*, vol. 16, no. 1, pp. 67-70, 2019. (https://doi.org/10.1038/s41592-018-0261-2).

[3] Y. Liu and F. Long, "Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning," in *CNMC Challenge: Classification in Cancer Cell Imaging*, pp. 113-121, 2019. (https://doi.org/10.1101/580852).

[4] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans SMC*, vol. 9, pp. 62-66, 1979. (https://doi.org/10.1109/TSMC.1979.4310076).

[5] C. Wählby, I.-M. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson, "Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections," *J. Microscopy*, vol. 215, pp. 67-76, 2004. (https://doi.org/10.1111/j.0022-2720.2004.01338.x).

[6] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM TOG*, vol. 23, pp. 309-314, 2004. (https://doi.org/10.1145/1015706.1015720).

[7] T. Hayakawa, V. B. Surya Prasath, H. Kawanaka, B. J. Aronow, and S. Tsuruoka, "Computational nuclei segmentation methods in digital pathology: A survey," *Archives of Computational Meth. Eng.*, vol. 28, pp. 1-13, 2021. (https://doi.org/10.1007/s11831-019-09366-4).

[8] T. J. Jebaseeli. C. A. D. Durai, and J. D. Peter, "Retinal blood vessel segmentation from diabetic retinopathy images using tandem PCNN model and deep learning based SVM," *Optik Int. J. Light and Electr. Optics*, vol. 199, no. 163328, 2019. (https://doi.org/10.1016/j.ijleo.2019.163328).

[9] A. Esteva, B. Kuprel, R. A. Novoato, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115-118, 2017. (https://doi.org/10.1038/nature21056).

[10] V.-T. Pham, T.-T. Tran, P.-C. Wang, and M.-T. Lo, "Tympanic membrane segmentation in otoscopic images based on fully convolutional network with active contour loss," *Signal, Image and Video Process.*, vol. 15, no. 3, pp. 519-527, 2021. (https://doi.org/10.1007/s11760-020-01772-7).

[11] V.-T. Pham, T.-T. Tran, P.-C. Wang, P.-Y. Chen, and M.-T. Lo, "EARUNet: A deep

learning-based approach for segmentation of tympanic membranes from otoscopic images," *Artificial Intell. in Med.*, vol. 115, pp. 1-12, 2021.
(https://doi.org/10.1016/j.artmed.2021.102065).

[12] C. Sommer, C. Straehle, U. Köthe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit," *IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, pp. 230-233, 2011.
(https://doi.org/10.1109/ISBI.2011.5872394).

[13] X. Pan, L. Li, D. Yang, Y. He, Z. Liu, and H. Yang, "An accurate nuclei segmentation algorithm in pathological image based on deep semantic network," *IEEE Access*, vol. 7, pp. 110674-110686, 2019.
(https://doi.org/10.1109/ACCESS.2019.293448 6).

[14] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-RCNN and U-Net ensembled for nuclei segmentation," *IEEE 16th ISBI 2019*, pp. 208-212, 2019.
(https://doi.org/10.48550/arXiv.1901.10170).

[15] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. A. Ashley, and M. W. Covert, "Deep learning Automates the11 quantitative analysis of individual cells in live-cell imaging experiments," *PLoS Comput. Biol.*, vol. 12, no. e1005177, 2016.
(https://doi.org/10.1371/journal.pcbi.1005177).

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Int. Conf. Med. Image Computing and Computer-Assisted Intervention*, pp. 234-241, Springer, 2015.
(https://doi.org/10.48550/arXiv.1505.04597).

[17] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-Unet: An improved neural network based on Unet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, pp. 21420-21428, 2019.
(https://doi.org/10.1109/ACCESS.2019.289692 0).

[18] Y. Zhou, O. F. Onder, Q. Dou, E. Tsougenis, H. Chen, and P. A. Heng, "CIA-net: Robust nuclei instance segmentation with contour-aware information aggregation," *Int. Conf. Inf. Process. in Med. Imaging*, pp. 682-693, 2019.
(https://doi.org/10.48550/arXiv.1903.05358).

[19] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 39, no. 12, pp. 2481-2495, 2017.
(https://doi.org/10.1109/TPAMI.2016.2644615)

[20] D. Jha, M. Riegler, D. Johansen, P. Halvorsen, and H. Johansen, "Doubleu-net: A deep convolutional neural network for medical image segmentation," *IEEE 33rd Int. Symp. CBMS*, pp. 558-564, 2020.
(https://doi.org/10.1109/CBMS49503.2020.001 11)

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. CVPR*, pp. 3431-3440, 2015.
(https://doi.org/10.48550/arXiv.1411.4038)

[22] Q. C. Ninh, T. T. Tran, T. T. Tran, T. A. X. Tran, and V. T. Pham, "Skin lesion segmentation based on modification of segnet neural networks," in *Proc. 2019 6th NICS*, Hanoi, pp. 575-578, 2020.
(https://doi.org/10.1109/NICS48868.2019.9023 862)

[23] A. Tversky, "Features of similarity," *Psychol. Rev.*, vol. 84, no. 4, p. 327, 1977.
(https://doi.org/10.1037/0033-295X.84.4.327).

[24] J. C. Caicedo, et al., "Nucleus segmentation across imaging experiments: the 2018 data science bowl," *Nature Meth.*, vol. 16, no. 12, pp. 1247-1125, 2019.
(https://doi.org/10.1038/s41592-019-0612-7)

Duy Cuong Bui

2020 : B.S. degree, Hanoi University of Science and Technology

<Research Interests> Computer vision, deep learning.

Myungsik Yoo

1989 : B.S. degree, Korea University, Seoul, South Korea

1991 : M.S. degree, Korea University, Seoul, South Korea

2000 : Ph.D. degree, State University of New York at Buffalo, NY, USA

<Research Interests> Visible light communications, cloud computing, and Internet protocols.