

An Automatic Nuclei Segmentation on Microscopic Images using Deep Residual U-Net

Ramya Shree H P, Minavathi, Dinesh M S

Computer Science & Engineering
PES College of Engineering
Mandya, Karnataka, India

Abstract—Nuclei Segmentation is the preliminary step towards the task of medical image analysis. Nowadays, there exists several deep learning-based techniques based on Convolutional Neural Networks (CNNs) for the task of nuclei segmentation. In this study, we present a neural network for semantic segmentation. This network harnesses the strengths in both residual learning and U-Net methodologies, thereby amplifying cell segmentation performance. This hybrid approach also facilitates the creation of network with diminished parameter requirement. The network incorporates residual units contributes to a smoother training process and mitigate the issue of vanishing gradients. Our model is tested on a microscopy image dataset which is publicly available from the 2018 Data Science Bowl grand challenge and assessed against U-Net and several other state-of-the-art deep learning approaches designed for nuclei segmentation. Our proposed approach showcases a notable improvement in average Intersection over Union (IoU) gain compared to prevailing state-of-the-art techniques, by exhibiting a significant margin of 1.1% and 5.8% higher gains over the original U-Net. Our model also excels across various key indicators, including accuracy, precision, recall and dice-coefficient. The outcomes underscore the potential of our proposed approach as a promising nuclei segmentation method for microscopy image analysis.

Keywords—Nuclei segmentation; convolutional neural networks; neural networks; U-Net; deep learning; semantic segmentation; 2018 data science bowl

I. INTRODUCTION

Microscopic image analysis continues to serve as a benchmark in the diagnosis and prognostication of various types of cancer. Segmenting nuclei is the initial phase in the analysis of microscopic images, as it directly influences the outcomes. This task is very challenging because the image acquisition is associated with color variation which is due different the use of different staining methods [1], artifacts, large variation in size, shape and texture of the cell nuclei [2], touching and overlapping nuclei [3], which is an obstacle for computer-aided diagnosis (CAD) segmentation algorithms. Numerous investigations have concentrated on nuclei detection because of its importance in the diagnosis of cancer. while traditional image processing approaches are being utilized for this task, could not achieve optimized performance due to inherent diversity involved within the images [4]. Past decade has witnessed a substantial progress in deep learning. Techniques relying on deep neural networks have attained state-of-art performance in automatic medical image segmentation [5]. These methods have demonstrated superior

outcomes compared to traditional approaches, showcasing the capability of harnessing deep learning techniques for the task of medical image segmentation.

Numerous studies have explored deep learning architectures for cell nuclei segmentation, each presenting unique methods and approaches to address this critical challenge. Despite these advancements, the demand for more precise nuclei segmentation persists. The rapid evolution of deep learning in the domain of medical image analysis has led to the development of various approaches, with many relying on the U-Net architecture [11], which has become the standard for medical image segmentation. Training exceptionally deep architectures introduces challenges that are primarily related to the problem of vanishing gradients. However, due to low resolution and blurry boundary of medical images, it is still a challenging task to design new models that can effectively capture more fine-grained details. In light of the ongoing need for more accurate nuclei segmentation, inspired by the success of U-Net [11] and deep residual learning [21], we propose Deep Residual U-Net that combines the strengths of residual learning and the U-Net architecture. This integration simplifies the training process, ensures smooth information propagation through the use of skip connections and mitigate the issue of vanishing gradients.

The significance of the Deep Residual U-Net is described below:

- Propose Deep Residual U-Net for semantic segmentation. This approach integrates deep residual network with standard U-net architecture to extract robust discriminative features from the input images.
- The evaluation of this proposed model on publicly available microscopy image dataset, specifically the 2018 Data Science Bowl grand challenge, has revealed notable improvements in various performance metrics.
- Proposed model achieves better segmentation masks in comparison with other baseline models, the especially in complex images with diverse cell sizes and shapes, where overlapping nuclei are prevalent.

Rest of the paper is organized as follows: Section II provides an overview of the related work in the field of nuclei segmentation. In Section III, we describe the overall methodology. Section IV describes the dataset, evaluation metrics and the experimental setup used for experimentation. In Section V, the results obtained and the performance

evaluation is discussed. In the last, we summarize the paper and discuss the future work.

II. RELATED WORK

Numerous deep learning architectures have been proposed for cell nuclei segmentation. Song et al. [6] introduced an approach based on CNN to segment cervical nuclei and cytoplasm. Their methodology involved in employing a CNN for nuclei detection, followed by coarse segmentation based on Sobel edge operator, morphological operations and thresholding. Xing et al. [7] generated probability maps for nuclei by applying Two-class CNN to digitized histopathology images. And to address the challenge of overlapping nuclei, the robust shape model (dictionary of nuclei shapes) was constructed and repulsive deformable model at local level was utilized. Kumar et al. [8] proposed Three-class CNN to predict the nuclei, background, and the boundary of each nucleus. This approach yielded notably better outcomes when compared to Two-class problem but the post-processing step was time consuming. The first FCN for semantic segmentation was presented by Long et al. [9]. Their results showed that the FCN can achieve state-of-the art performance in the realm of segmentation. Further, the inference step associated with this method is significantly faster to obtain the corresponding segmentation mask. For the task of nuclei segmentation in histopathology images, Naylor et al. [10] used FCN to obtain the nuclei probability map, then watershed method was applied to split the touching nuclei, but when comparing the nuclei boundaries predicted by this approach to the ground truth images was not accurate.

Investigation in the domain of deep learning is increasing rapidly, development of new architectures is under process at significantly fast speed. Accounting the importance of cell nuclei segmentation, several approaches have been proposed to address this issue, most of them are relying on U-Net [11]. U-Net is the prevailing architecture used for medical image segmentation. Several approaches based on U-Net have been presented to resolve the issue of nuclei segmentation. Cui et al. [12] have proposed a method, inspired by U-Net, to predict nuclei and their contours simultaneously in H&E-stained images. By predicting contour of each nucleus, applying a sophisticated weight map in the loss function they were able to split touching and overlapping nuclei accurately with simple and parameter free post-processing step. Caicedo et al. [4] trained U-Net model to predict the nuclei and their boundaries, giving the loss function with weight which is 10 times more to the boundary class. The first best solution by [ods.ai] topcoders [13], used encoder-decoder type architecture based on U-Net, initializing encoders with pretrained weights. Kong et al. [14] have used Two-stage stacked U-Nets, where stage1 for nuclei segmentation and stage2 to tackle the problem of overlapping nuclei. Zhao et al. [15] used U-Net++, which is a modification to the U-Net [11] architecture, which combined U-Nets of different depths. Pan et al. [16] proposed AS-Unet, an extension of the U-Net architecture, is structured around three fundamental components: encoder module, decoder module and an atrous convolutional module. Alom et al. [17] applied a technique which relied on the RCNN for nuclei segmentation tasks. J. Cheng et al. [18] used FCANet, which is based on U-Net that captures long-range and short-range

distance features and use attention module to refine the features. Chen et al. [19] proposed Dense-Res-Inception Net (DRINET) for the task of segmentation on medical images and compared their results with FCN, U-Net and ResUNet. Ibtehaz et al. [20] enhanced the U-Net architecture and proposed an advanced MultiResUNet architecture for medical image segmentation. They conducted comprehensive comparison with U-Net using diverse medical image segmentation datasets, revealing that their proposed method achieved enhanced accuracy compared to U-Net.

From the aforementioned review of relevant studies, it is evident that there have been substantial efforts invested in the advancement of deep Convolution Neural Networks (CNN) architectures in effectively segmenting both natural and medical images. Recent research has indicated that better performance can achieved through deeper network. However, training exceptionally deep architectures poses the challenges due to the issue of vanishing gradients. To handle this problem, He et al. [21] proposed the deep residual learning framework to facilitate the training process by utilizing an identity mapping [22]. Instead of using skip connection in Fully Convolutional Networks (FCNs) [9], Ronneberger et al. [11] proposed U-Net that amalgamate feature maps from various hierarchical levels, resulting in enhanced segmentation accuracy. By merging low-level intricate details and high-level semantic insights, U-net has demonstrated remarkable performance in biomedical image segmentation tasks [11]. Inspired by deep residual learning [21] and U-Net [11], we have integrated the residual network to the U-Net architecture. This integration allows us to harness the strengths of both residual learning and U-Net framework, resulting in a unified approach that maximizes the benefits derived from each approach. We have replaced plain neural units by residual units in the U-Net architecture which simplifies the training process and incorporation of robust skip connections within the network has enabled the smooth propagation of information without experiencing degradation.

III. METHODOLOGY

A. Overview of Deep Residual U-Net Architecture

- *U-Net*: Semantic segmentation is a task of dividing an image into segments or regions, where each segment corresponds to a meaningful object or part of a scene. In semantic segmentation, utilization of low-level details while preserving high-level semantic information [9, 11] holds immense significance, as it is crucial in attaining precise results. However, especially when we are with limited training samples, training such deep neural networks becomes very challenging. Two possible ways to address this problem. First, by using a pre-trained network later fine-tuning it on the target dataset as in [9]. Two, by utilizing data augmentation strategy which is done in U-Net [11]. Along with the data augmentation, we believe that U-Net architecture also contributes to alleviate the training problem. The idea behind this is that, copying low level features to its equivalent high levels creates a path to propagate information which allows signal to propagate seamlessly between low levels and high

levels by facilitating backward propagation during training process and also compensating low level finer details to its corresponding high level semantic features. This concept pertains to residual neural network [21].

- Residual unit: Adding more layers can improve the performance of the multi-layer neural network by increasing the trainable parameters which may lead to redundant computation and may cause degradation problem [21]. To handle this problem, He et al. [21] proposed the deep residual learning framework, aiming to alleviate training difficulties and effectively mitigate the degradation problem. This leads to improvement in network's performance without the need for deeper network or pre-trained networks. A deep residual network comprises a sequence of layered residual blocks, with each individual residual block is defined as:

$$y_i = h(x_i) + F(x_i, W_i), \quad (1)$$

$$x_{i+1} = f(y_i) \quad (2)$$

Where x_i is the input and x_{i+1} output of the i^{th} residual block, $F(\cdot)$ is the residual function, $f(y_i)$ is the activation function, $h(x_i)$ represents the identity mapping function and W_i is the weight vector of the feature map within the i^{th} residual block. The difference between plain neural unit and residual block illustrated in Fig. 1. Within each residual block, a composition of batch normalization (BN), ReLu activation function and convolution layers is present. He et al. [22] has discussed the impacts on using different combinations and suggested the full pre-activation design as depicted in Fig. 1(b). In our work, we have used full pre-activation residual block to build our deep residual U-Net.

- Deep Residual U-Net: Here we propose a deep residual U-Net, a neural network designed for semantic segmentation which takes the advantage of both U-Net and residual neural network. With this combination, the residual blocks in the network contributes to smoother training process and the skip connections within these blocks, as well as between the network's low levels and high levels ensures smooth flow of information without degradation. Also, it allows to design the neural network with relatively few parameters and could achieve comparable better performance in semantic segmentation tasks.

In this study, we have utilized 9-level architecture of deep residual U-Net for nuclei segmentation as depicted in Fig. 2. The network consists three parts: Encoding, Bridge and Decoding paths. The first part, involves in capturing high level features from the input image by reducing its spatial dimensions while preserving the important spatial information through skip connections. The last part, involves in upsampling the encoded feature maps to reconstruct the segmentation mask with the same spatial resolution as that of the input image. The middle part, serves as a bridge in connecting the encoding path and decoding path. All segments are constructed using residual blocks, each comprising two 3x3 convolutional blocks and an identity mapping. Each convolutional block has a BN layer, a ReLu activation layer and a convolution layer in it. The identity mapping connects both input and output of the block.

The encoder path has four residual blocks. In each block, to downsample the feature map size, instead of using maxpooling operation, a stride of 2 is implemented on the first convolution block to reduce the feature map by half. Likewise, the decoder pathway comprises four residual blocks. Every residual block entails the upsampling of feature maps from lower levels, coupled with the concatenation of feature maps derived from the corresponding encoder path. Finally, 1x1 convolution layer is employed succeeded by a sigmoid activation layer. The role of sigmoid activation function is to project the multi-channel feature maps into the intended segmentation map. Overall, the network encompasses 28 convolution layers. The parameters and the resultant output sizes of each step are provided in Table I.

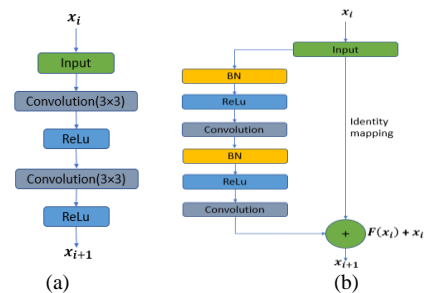


Fig. 1. Building blocks of neural networks. (a) Plain neural unit used in U-net and (b) residual block with identity mapping used in the proposed deep residual U-Net.

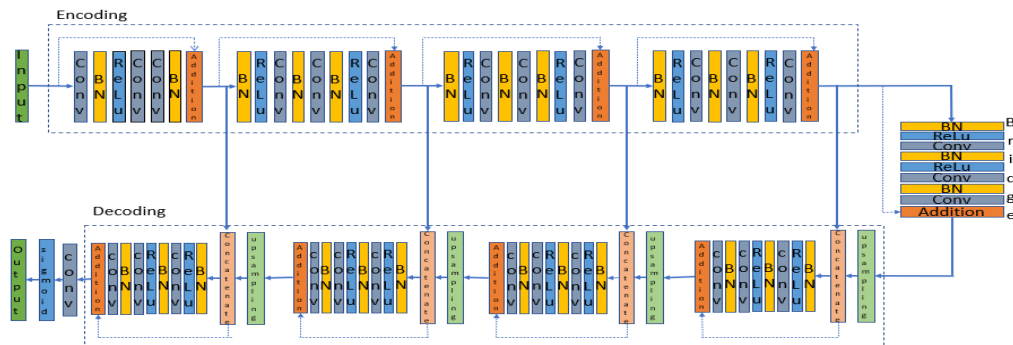


Fig. 2. The architecture of deep residual U-Net.

TABLE I. THE NETWORK STRUCTURE OF DEEP RESIDUAL U-NET

	Level	Conv Layer	Filter	Stride	Output Size
Input					$256 \times 256 \times 3$
Encoding	Level 1	Conv 1	$3 \times 3/16$	1	$256 \times 256 \times 16$
		Conv 2	$3 \times 3/16$	1	$256 \times 256 \times 16$
		Conv 3	$3 \times 3/16$	1	$256 \times 256 \times 16$
	Level 2	Conv 4	$3 \times 3/32$	2	$128 \times 128 \times 32$
		Conv 5	$3 \times 3/32$	1	$128 \times 128 \times 32$
		Conv 6	$3 \times 3/32$	1	$128 \times 128 \times 32$
	Level 3	Conv 7	$3 \times 3/64$	2	$64 \times 64 \times 64$
		Conv 8	$3 \times 3/64$	1	$64 \times 64 \times 64$
		Conv 9	$3 \times 3/64$	1	$64 \times 64 \times 64$
	Level 4	Conv 10 Conv 11 Conv 12	$3 \times$	2	$32 \times 32 \times 128$
			$3/128$		
			$3 \times$	1	$32 \times 32 \times 128$
$3/128$					
$3 \times$	1	$32 \times 32 \times 128$			
$3/128$					
Bridge	Level 5	Conv 13 Conv 14 Conv 15	$3 \times$	2	$16 \times 16 \times 256$
			$3/256$		
			$3 \times$	1	$16 \times 16 \times 256$
			$3/256$		
$3 \times$	1	$16 \times 16 \times 256$			
$3/256$					
Decoding	Level 6	Conv 16 Conv 17 Conv 18	$3 \times$	1	$32 \times 32 \times 128$
			$3/128$		
			$3 \times$	1	$32 \times 32 \times 128$
			$3/128$		
	$3 \times$	1	$32 \times 32 \times 128$		
	$3/128$				
	Level 7	Conv 19 Conv 20 Conv 21	$3 \times 3/64$	1	$64 \times 64 \times 64$
			$3 \times 3/64$	1	$64 \times 64 \times 64$
			$3 \times 3/64$	1	$64 \times 64 \times 64$
	Level 8	Conv 22 Conv 23	$3 \times 3/32$	1	$128 \times 128 \times 32$
			$3 \times 3/32$	1	$128 \times 128 \times 32$
	Level 9	Conv 24 Conv 25 Conv 26 Conv 27	$3 \times 3/32$	1	$128 \times 128 \times 32$
$3 \times 3/16$			1	$256 \times 256 \times 16$	
$3 \times 3/16$			1	$256 \times 256 \times 16$	
Output		Conv 28	1×1	1	$256 \times 256 \times 1$

IV. EXPERIMENTS

Within this section, we delve into the details of the dataset, the evaluation metrics, the experimental setup and the data augmentation techniques utilized to validate our proposed model.

A. Dataset

In our research, we utilized the dataset provided by Kaggle 2018 DSB challenge. The dataset includes 871 images with 37, 333 manually annotated nuclei. The images represent 31 experiments with 22 cell types, 15 different resolutions and 5 groups of images which are visually indistinguishable. This dataset includes 2D light microscopy images with different staining methods including DAPI, Hoechst or H&E and cells of different sizes which display the structures from variety of organs and animal model. Out of 31 experiments, 16 are for training (670 samples), first-stage evaluation (65 samples) and 15 for second-stage evaluation (106 samples). This dataset is readily accessible to the public through the Broad Bioimage Benchmark Collection.

The dataset comprises of 670 training samples accompanied by their corresponding masks. For our experiments, we have allocated 80% of the dataset for training

purpose, while 10% for validation and remaining 10% was reserved for testing. Additionally, we have evaluated our trained model on `stage1_test` dataset provided by the challenge. This dataset comprises of 65 samples, each equipped with ground truth masks.

B. Evaluation Metrics

The evaluation of the proposed model is based on several metrics, including the Sørensen–Dice coefficient (DSC), also referred to as F1-Score, the Intersection over Union (IoU), commonly known as the Jaccard Index (JI), Precision and Recall. DSC assesses the similarity between predicted and ground truth masks, while IoU quantifies the overlap between the two masks, Precision measures the portion of pixels that are correctly classified as nuclei pixels out of all the pixels that are classified as nuclei pixels and Recall measures the portion of pixels that are correctly classified as nuclei pixels out of all the pixels that are actually nuclei pixels in the image. These indices are expressed in equations (3-6). The terms TP, FP, TN and FN correspond to True positive, False positive, True negative and False negative [16].

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Dice coefficient (DSC)} = \frac{2TP}{2TP+FP+FN} \quad (5)$$

$$\text{IoU} = \frac{TP}{TP+FP+FN} \quad (6)$$

C. Experimental Setup

The proposed model is implemented using Keras framework [23] with Tensorflow 2.7.0 as backend, OpenCV library and python 3.7. The number of kernels in the encoder were set to 16, 32, 64, 128 and 256, and the kernels in the decoder were set to 128, 64, 32, 16 and 1. The input images were resized to dimensions of 256×256 pixels. We have employed binary cross-entropy as the loss function and an Adam optimization technique, aiming to minimize the loss function with the batch size of 16 and a learning rate of $1e-4$. A training procedure was conducted for a span of 100 epochs, with criteria such as early stopping and ReduceLROnPlateau. To prevent the model overfitting, data augmentation techniques such as horizontal flipping, rotation and zooming has been adopted on training dataset in our experiment. The experiment was performed on Nvidia GeForce RTX2080 Ti with 11GB RAM.

V. RESULTS AND DISCUSSION

Within this section, we present the results and compare with state-of-art methods. U-Net is still considered as baseline for diverse medical image segmentation tasks. In the interest of comprehensive comparison, we also trained U-Net, U-Net++ and HR-Net which are well-regarded techniques for nuclei segmentation with the same experimental setup. The learning curve of our proposed model is presented in Fig. 3. Notably, our model demonstrates the convergence after 30 epochs, exhibiting a validation loss of 0.069 and IoU score of 0.8213, respectively.

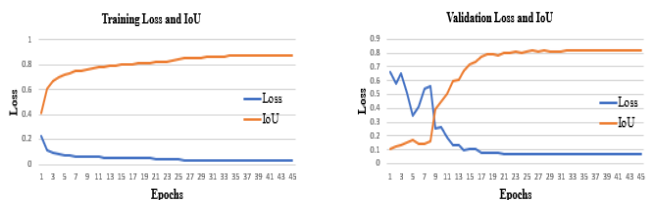


Fig. 3. Loss v/s IoU curve of training (left-side) and validation (right-side) trained on our proposed model.

Fig. 4 depicts the loss curve during training and validation process. When we observe the curve during training, the loss value is decreasing after each epoch. During validation, the loss value reduces unevenly and later it becomes smooth.

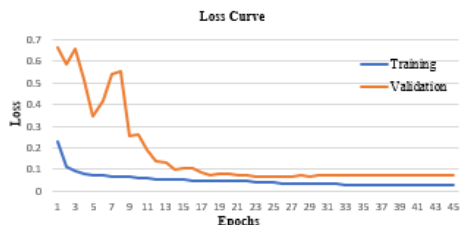


Fig. 4. Training v/s validation loss curve.

Fig. 5 illustrates a comparative analysis of our proposed model in contrast to other baseline models in terms of loss values and IoU scores throughout training and validation phases. In Fig. 5(a), which portrays training loss curve, as well as in Fig. 5(b) which depicts validation loss curve, the loss value of our proposed model is lowest during training and remains competitive with other models during validation process. Fig. 5(c) shows the IoU curve obtained during training and Fig. 5(d) shows the IoU curve observed during validation for all the models. It is evident from Fig. 5(c) and 5(d), our proposed model attains a better IoU in comparison to other models during both training and validation process. Notably, U-Net has achieved the lowest IoU, while U-Net++ and HR-Net yield nearly identical IoU scores. Overall, from Fig. 5, our proposed model showcases better performance when compared to other models.

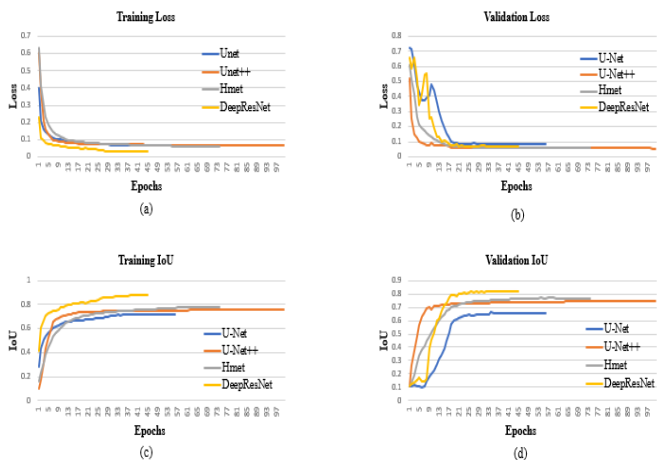


Fig. 5. Comparison of loss curve and IoU curve of our proposed model with other models during training and validation process. (a) Training loss, (b) validation loss, (c) Training IoU and (d) Validation IoU.

The outcomes of our proposed method are presented in Table II alongside those of baseline methods, assessed using various evaluation metrics. It is evident that our proposed method surpasses the initial U-Net [11] by 1.77% in DSC and 1.09% in IoU. Furthermore, when compared with U-Net++ [15] and HR-Net [24], our method showcases substantial improvement across DSC, IoU and Precision metrics. The recall of the U-Net++ has a slight increase of 0.01% compared to U-Net.

TABLE II. QUANTITATIVE RESULTS ON THE EXPERIMENTAL DATASET WITH BASELINE METHODS

Model	Accuracy	DSC	IoU	Recall	Precision
U-Net [11]	0.978±0.0 21	0.908±0.0 87	0.842±0.1 21	0.911±0.1 13	0.917±0.0 87
U-Net++ [15]	0.979±0.0 22	0.898±0.0 74	0.827±0.1 10	0.912±0.1 05	0.895±0.0 65
HR-Net [24]	0.967±0.0 28	0.852±0.1 21	0.757±0.1 42	0.880±0.1 36	0.836±0.1 46
DeepResNet	0.977±0.0 22	0.910±0.1 02	0.853±0.1 26	0.911±0.1 15	0.918±0.0 84

The comparison of our proposed method with other state-of-art techniques is summarized in Table III. The data within the table clearly underscores better performance of our model in relation to other methods.

TABLE III. QUANTITATIVE RESULTS ON THE EXPERIMENTAL DATASET WITH STATE-OF-ART METHODS

Model	Accuracy	DSC	IoU	Recall	Precision
SegNet [25]	--	0.738±0.1 34	0.620±0.1 35	--	0.820±0.1 32
DeepLabV3+ [26]	--	0.741±0.3 19	0.674±0.2 60	--	0.818±0.4 01
DANet [27]	--	0.616±0.1 61	0.564±0.3 00	--	0.761±0.1 26
FCANet [18]	--	0.897±0.0 80	0.814±0.1 36	--	0.895±0.0 51
DoubleU-Net [28]	0.941±0.0 68	0.903±0.0 89	0.833±0.1 29	0.865±0.1 31	0.957±0.0 39
MSAU-Net [29]	0.944±0.0 66	0.907±0.0 39	0.842±0.1 28	0.893±0.1 22	0.938±0.0 69
TransU-Net [30]	0.954±0.0 47	0.895±0.0 99	0.821±0.1 36	0.906±0.1 21	0.900±0.1 01
OAU-Net [31]	0.9677	0.8992	0.8235	0.9008	0.9096
DeepResNet	0.977±0.0 22	0.910±0.1 02	0.853±0.1 26	0.911±0.1 15	0.918±0.0 84

Fig. 6 displays the segmentation outcomes of each model. Through visual examination, it becomes apparent that the segmentation mask generated by our model is better than those produced by other models.

Our proposed model, along with other models, underwent testing and evaluated using stage1_test dataset which includes 65 samples, each accompanied by a ground truth mask provided by the organizers. Quantitative outcomes on

stage1_test dataset, comparing our proposed method to other techniques are tabulated in Table IV across various evaluation metrics. Upon inspecting Table IV, it apparent that our model surpasses the original U-Net by 5.8% in DSC and 6.0% in IoU.

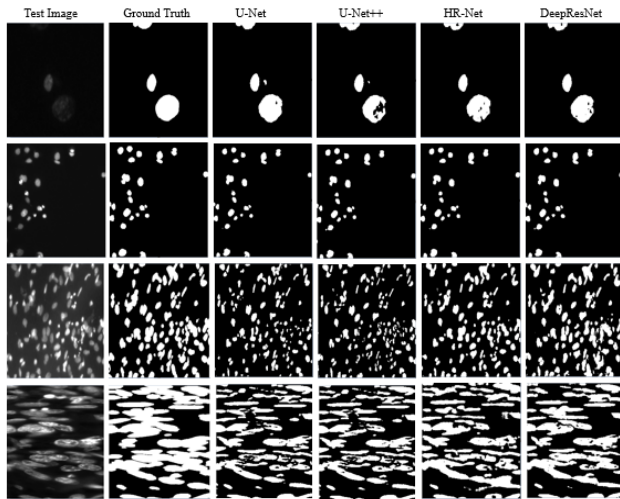


Fig. 6. Visualization of segmentation results on test dataset.

TABLE IV. QUANTITATIVE RESULTS ON THE STAGE1_TEST DATASET

Model	Accuracy	DSC	IoU	Recall	Precision
U-Net [11]	0.932±0.075	0.780±0.211	0.679±0.234	0.739±0.243	0.904±0.149
U-Net++ [15]	0.943±0.056	0.826±0.140	0.724±0.172	0.790±0.176	0.892±0.140
HR-Net [24]	0.941±0.063	0.824±0.150	0.724±0.179	0.803±0.167	0.869±0.163
DeepResNet	0.946±0.053	0.838±0.129	0.739±0.162	0.817±0.150	0.886±0.145

Notably, the model’s precision (0.886) falls short of that achieved by U-Net++. However, recall rate remains competitive when compared to other models. Overall, our model demonstrates strong performance across multiple metrics.

To visualize the segmentation outcomes in detail, we examine samples from test set that encompass cells of varying sizes. The qualitative comparison between ground truth images, our proposed model and other models on stage1_test dataset is depicted in Fig. 7. In the Fig. 7, the first column represents the actual image, second column displays its corresponding ground truth mask and remaining columns represents the prediction masks of different models. Among the prediction masks, the red represents the part where the model predicts the background as the target area (FP). For the microscopic images with few of cells, where the nuclei can easily be discriminated from the background, all the models demonstrate satisfactory segmentation outcomes. However, for the complex images such as third, fourth and fifth rows of images with nuclei of different sizes and shapes, the segmentation mask produced by our model is better than those of other models. Also, our model exhibits fewer false positives and enhanced detection accuracy in comparison to the other models.

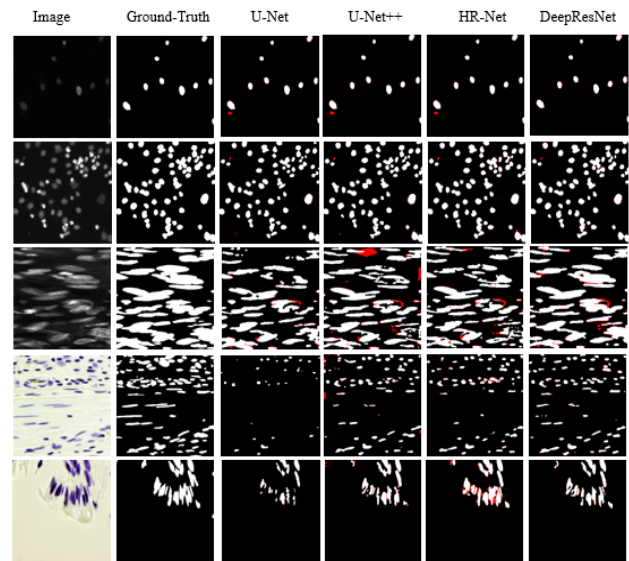


Fig. 7. Visualization of segmentation results on stage1_test dataset.

VI. CONCLUSION

Addressing the requirement for more precise nuclei segmentation task, we propose a semantic segmentation neural network that harnesses the combined strength of residual learning and U-Net. The residual block within the network makes the training process easier, while the skip connections within and between the residual block at low and high levels of the network will propagate the information both in forward and backward computations. Also, this property allows us design a simple powerful neural network with fewer number of parameters compared to U-Net. Our model’s efficacy was evaluated on publicly available microscopy image dataset from 2018 Data Science Bowl grand challenge. The outcomes of our experiments revealed an average IoU improvement of 1.1 and 5.8 (for the stage1_test set) over original U-Net. Across images with smaller number of cells, where nuclei were distinct, all models performed well. However, when faced with complex images containing cells of diverse sizes and shapes, our proposed model consistently generated better segmentation masks compared to its counterparts. The evaluation conclusively demonstrates that our model excels in terms of accuracy, precision, recall and dice-coefficient when compared to U-Net and other prominent models.

Subsequently, we incorporated watershed segmentation as a post-processing step to tackle the challenges associated with touching nuclei and overlapping/clustered nuclei. This approach proved successful in effectively segmenting touching nuclei. In the future, we plan to conduct experimentation on larger and more diverse dataset that can help validate our model’s performance and generalizability. Further, we continue to explore and develop most effective post-processing methods to address the challenge of overlapping nuclei.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to PES College of Engineering, Mandya, for generously providing essential computational resources in form of workstation (Nvidia GeForce RTX2080 Ti with 11GB RAM) under TEQIP III. This support played a pivotal role in successful completion of our research work.

REFERENCES

- [1] Khoshdeli, M., and Parvin, B.: Deep leaning models delineates multiple nuclear phenotypes in H&E-stained histology sections, arXiv preprint arXiv:1802.04427, (2018).
- [2] Liu, Y., Zhang, P., Song, Q., Li, A., Zhang, P., Gui, Z.: Automatic segmentation of cervical nuclei based on deep learning and a conditional random field, *IEEE Access*, vol. 6, pp. 53 709–53 721, (2018). doi: 10.1109/ACCESS.2018.2871153
- [3] Xing, F., Xie, Y., Yang, L.: An automatic learning-based framework for robust nucleus segmentation, *IEEE transactions on medical imaging*, Feb;35(2):550-66, (2016). doi: 10.1109/TMI.2015.2481436
- [4] Caicedo, J. C., Roth, J., Goodman, A., Becker, T., Karhohs, K. W., Broisin, M., Csaba, M., McQuin, C., Singh, S., Theis, F. et al. "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images," *BioRxiv*, p. 335216, (2019). doi: 10.1002/cyto.a.2386
- [5] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., C. I. S'anchez. A survey on deep learning in medical image analysis, *Medical image analysis (MedIA)*, vol. 42, pp. 60–88, (2017).
- [6] Song, Y., Zhang, L., Chen, S., Ni, D., Li, B., Zhou, Y., Lei, B., Wang, T.: A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei, *Engineering in Medicine and Biology Society (EMBC), 36th annual international conference of the IEEE, IEEE*, pp. 2903–2906, (2014). doi: 10.1109/EMBC.2014.6944230
- [7] Xing, F., Xie, Y., Yang, L.: An automatic learning-based framework for robust nucleus segmentation, *IEEE transactions on medical imaging*, Feb;35(2):550-66, (2016). doi: 10.1109/TMI.2015.2481436
- [8] Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: 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). doi: 10.1109/TMI.2017.2677499
- [9] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, (2015). doi: 10.1109/TPAMI.2016.2572683
- [10] Naylor, P., Lae, M., Reyal, F., Walter, T.: Nuclei segmentation in histopathology images using deep neural networks, *14th International Symposium on Biomedical Imaging (ISBI 2017)*. *IEEE*, pp. 933–936, (2017). doi: 10.1109/ISBI.2017.7950669
- [11] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241, (2015). doi: 10.1007/978-3-319-24574-4_28
- [12] Cui, Y., Zhang, G., Liu, Z., Xiong, Z., Hu, J.: A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images, arXiv preprint arXiv:1803.02786, (2018). doi: 10.1007/s11517-019-02008-8
- [13] [ods.ai] topcoders, 1st place solution, <https://www.kaggle.com/c/data-science-bowl-2018/discussion/54741>
- [14] Yan, K., Georgi, Z. G., Wang, X., Zhao, H., Lu, H.: Nuclear segmentation in Histopathological Images using Two-Stage Stacked U-Nets with Attention Mechanism, *Front. Bioeng. Biotechnol.*, 26 October (2020). doi: 10.3389/fbioe.2020.573866
- [15] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J.: UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation, *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, June (2020). doi: 10.1109/TMI.2019.2959609
- [16] Pan, X., Li, L., Yang, D., He, Y., Liu, Z., Yang, H.: An Accurate Nuclei Segmentation Algorithm in Pathological Image Based on Deep Semantic Network, *IEEE Access*, vol. 7, pp. 110674–110686, (2019). doi: 10.1109/ACCESS.2019.2934486
- [17] Alom, MZ., Yakopcic, C., Hasan, M., Taha, TM., Asari, VK.: Recurrent residual U-Net for medical image segmentation, *J. Med. Imag.* 6(1), 014006 (2019), doi: 10.1117/1.JMI.6.1.014006
- [18] Cheng, J., Tian, S., Yu, L., Lu, H., Lv, X.: Fully convolutional attention network for biomedical image segmentation, *Artificial Intelligence in Medicine* 107 (4):101899, June (2020). doi: 10.1016/j.artmed.2020.101899
- [19] Chen, L., Bentley, p., Mori, K., Misawa, K., Fujiwara, M., Rueckert D.: Drinet for medical image segmentation, *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2453–2462, (2018).
- [20] Ibtihaz, N., Rahman, MS.: Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation, *Neural Networks*, vol. 121, pp. 74–87, (2020). doi: 10.1109/TMI.2018.2835303
- [21] He, K., Zhang, H., Ren, S., Sun, J.: Deep residual learning for image recognition, in *CVPR*, pp. 770–778, (2016). doi: 10.1109/CVPR.2016.90
- [22] He, K., Zhang, H., Ren, S., Sun, J.: Identity mappings in deep residual networks, in *ECCV*, pp. 630–645, (2016). arXiv:1603.05027v3
- [23] Chollet F, et al. Keras: The python deep learning library 2015; 2019.
- [24] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., Xiao, B. Deep High-Resolution Representation Learning for Visual Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 3349–3364, Oct. (2021). doi: 10.1109/TPAMI.2020.2983686
- [25] Badrinarayanan, V., Kendall, A., Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39(12):2481–95. <https://doi.org/10.17863/CAM.17966>.
- [26] Chen L, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV) 2018*:833–51. https://doi.org/10.1007/978-3-030-01234-2_49
- [27] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019*:3146–54. <https://doi.org/10.1109/CVPR.2019.00326>.
- [28] Jha D, Riegler MA, Johanseny D, Halvorsen P, Johanseny HD, Met S (2006) Double U-Net: a deep convolutional neural network for medical image segmentation. *Electr Eng Syst Sci—Image Video Process*, Cornell Univ, June (2020). <https://doi.org/10.1109/CBMS49503.2020.00111>
- [29] Xu, Q., Duan, W. An Automatic Nuclei Image Segmentation Based on Multi-Scale Split-Attention U-Net, *Proceedings of the MICCAI Workshop on Computational Pathology*, PMLR 156:236-245, (2021).
- [30] Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y., Lu L., Yuille A.L., Zhou Y. Transunet: Transformers make strong encoders for medical image segmentation (2021). arXiv preprint arXiv:2102.04306
- [31] Song H., Wang Y., Zeng S., Guo X., Li Z. OAU-Net: Outlined Attention U-net for biomedical image segmentation, *Biomedical Signal Processing and Control*, Volume 79, January (2023). <https://doi.org/10.1016/j.bspc.2022.104038>