

Comparison of UNet, UNet++, Res-Unet,  
SegNet and Seg-Unet for Segmentation of Cells'  
Nuclei image

Meng Zhaoyi, Zhao jiayi, Liu Yanjie, Chen Tianyu

20 November 2020

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Literature Review</b>	<b>4</b>
<b>3</b>	<b>Background</b>	<b>7</b>
3.1	Unet . . . . .	7
3.2	Unet++ . . . . .	7
3.3	SegNet . . . . .	8
3.4	Seg-Unet . . . . .	9
3.5	Res-Unet . . . . .	10
3.5.1	U-Net . . . . .	10
3.5.2	Residual Network . . . . .	11
3.5.3	Res-Unet . . . . .	11
<b>4</b>	<b>Experiment and Results</b>	<b>12</b>
4.1	Dataset . . . . .	12
4.2	Evaluation Measures . . . . .	13
4.3	Hyperparameter Setup . . . . .	14
4.4	Results . . . . .	14
<b>5</b>	<b>Conclusion</b>	<b>17</b>

## Abstract

Nowadays, medical images are widely used in medical treatment, among which medical image segmentation is the most important part. With the continuous improvement of the accuracy of deep learning models, this method has gradually replaced the method of manually extracting features through machine learning.

In this paper, we compared five models: Unet, Unet++, SegNet, Seg-Unet and Res-Unet. Among them, Unet, Unet++, and ResUnet are considered to be good fits, and there is no sign of overfitting or underfitting. On the contrary, the effects of SegNet and Seg-Unet are not ideal. SegNet has over-fitting. Although Seg-Unet has no signs of over-fitting, the training and validation losses are too large. What's more, the dice coefficient of Unet++ and Res-Unet on the testing dataset are higher than 90%, which shows that these two models are the best.

## 1 Introduction

Medical image is an essential auxiliary means in modern medical therapy. For example, magnetic resonance imaging (MRI), computed tomography (CT), and digital mammography are widely used to diagnose diseases and map objects in anatomy.[1] The collected image can be either 2-dimensional (such as microscope images) or 3-dimensional (CT images). There is no obvious difference between these two types of image processing for human beings, while for machines, the 2D image operation is much simpler.

With the development of deep learning, image processing technology also makes much progress. Although image processing has been studied in the past using various traditional computer vision and machine learning techniques, the deep learning revolution is a historic step in this field because the image segmentation method, which is tackled using deep architectures, is surpassing other approaches by a large margin in terms of accuracy and sometimes even efficiency.[2]

The application of image segmentation also includes medical image segmentation. Because of the high accuracy and efficiency, this deep learning method replaces extracting hand-crafted features of machine learning approaches and has become increasingly prevalent.[3]

Considering the importance of medical images to diagnose patients' situation, few doctors can conclude without these images. It is obvious there are no high-quality medical resources in underdeveloped areas, and many diseases cannot be diagnosed quickly and receive adequate medical treatment, resulting in high mortality. While with the help of medical image segmentation even an inexperienced doctor can draw accurate conclusions to patients' situations, which may save more lives.[4]

However, challenges with Training Deep Models can not be ignored. Overfitting and gradient vanishing are two inevitable problems. The solutions to these problems are closely related to the specific model. At the same time, training time and memory should also be taken into consideration.

Therefore, this project intends to know the performance of different model which has been used in the medical image field and compares five typical models that include Unet, Unet++, SegNet, Seg-Unet and Res-Unet.

The paper is structured as follows. Section 2 describes related work in the area of image segmentation, especially medical image segmentation. In Section 3, the five models which have been compared are described. Section 4 describes the dataset, the experiments and discusses the results. The conclusion and future work are provided in Section 5.

## 2 Literature Review

FCN opens a new era of image segmentation, which classifies images at the pixel level. Before FCN, even though deep learning methods (such as CNN) have been applied in the image segmentation field, details of the image cannot be separated or recognized through the model. After discovering FCN, the input of the image can be any size, and use the deconvolution layer to upsample the feature map of the last convolution layer to restore it to the same size as the input image. [5]

Unet is one of the networks which is improved based on FCN. Olaf Ronneberger, Philipp Fischer, and Thomas Brox came up with this model and won the ISBI cell tracking challenge 2015. This model includes two parts, a feature extraction part and an upsampling part, and the whole model looks like U. The Unet model is especially fit for medical image segmentation and can be trained end-to-end from very few images.[6]

Deformable U-Net is a developed Unet model, which has a deformable convolutional block rather than the original convolutional layer. The deformable convolutional block introduces deformable convolutional layers and deformable ROI pooling layers into the traditional neural networks. Through this method, the network can adapt to different scales, shapes, orientations of the input.[4]

Unet++ concatenate many Unet networks of different sizes into a single model. The architecture of Unet++ is a deeply-supervised encoder-decoder network, and the sub-networks of encoder and decoder are connected by skip pathways, which are used to reduce the semantic gap between the feature maps.[7] This model alleviates the problems of the unknown depth of the whole network when assembling many U-Net networks of varying depths.[8]

SegNet is another model evolved from FCN. Compared with other models, the SegNet network is an efficient model in time and memory usage. The encoder of SegNet has the same topology as convolutional layers in VGG16.[9] While the fully connected layer is removed, in addition, the decoder uses the max-pooling indices received from the corresponding encoder to perform nonlinear upsampling of the input feature map.[10]

Nhu-Tai Do et al.[11] proposed the Seg-Unet model. This model combines SegNet and Unet together to improve the limited data

ResUnet is a new architecture that combines the advantages of deep residual learning and Unet.[12] This structure simplifies training and optimizes extrac-

tion. In the medical industry, ResUNet can measure spinal parameters[13] and extract the path of cell segmentation[14],satellite imagery of medical organisms. According to the research results[13], the results obtained using the ResUNet structure are not much different from the doctor’s diagnosis results and have high accuracy. In addition, RestUNet can also be used in clinical medicine to automatically segment zebrafish blood vessels (very similar to humans). It solves the problem that it is difficult to identify the segmented overlapping area due to uneven fluorescence distribution.[15]

For semantic segmentation of high-resolution aerial images, a new deep learning architecture ResUNet-a can be used. This is based on the encoder/decoder paradigm, where the standard convolution is replaced with a ResNet unit, which contains multiple parallel atrous convolutions. After some experiments prove that this method can improve semantic segmentation performance.[16]

In medical image segmentation, a structure called ResUNet++ is also used. It is a semantic segmentation neural network using residual blocks, extrusion and excitation blocks, spatial pyramid pool (ASPP), and attention blocks. It is mainly used for the segmentation of intestinal polyps and is suitable for a relatively small number of images.[17]

Stacked U-nets are stacked based on U-nets and combine the image characteristics of different resolutions while maintaining the resolution. S-Unets solves the traditional model structure that limits images’ effectiveness in secondary tasks such as pixel-level positioning or classification.[18] After a series of processing, the S-Unets structure can remove the mesh artifacts that often plague the expanded network structure’s output.[19] S-Unets structure combined with multiple outputs, reducing thresholds, searching for the shortest path, and other optimization techniques can improve accuracy and are usually used to extract road satellite images.[20]

U2-Net is a deep learning framework for photoreceptor layer segmentation in pathological OCT scanning. This structure, combined with cognitive uncertainty graphs, highlighting potential areas of pathology or segmentation errors, can speed up manual labeling.[21]

Generative Adversarial Network (GAN) is an effective data anonymization method and provides an additional form of data augmentation. GAN can extract and generate labeled synthetic retinal images in the medical field and can synthesize brain CT through brain nuclear magnetic images.[22] At the same time, GAN can be used to segment lungs in chest X images. GAN is composed of generating G network and discriminating D network. The generator is used to generate data from random noise. The discriminator can distinguish the real data and the data generated by the generator.[23]

DeepLab v3 is a convolution-based image segmentation technology, which can be used to diagnose melanoma cases. This technique uses convolution, where the filter is applied to each image interval’s pixel matrix rather than to all adjacent pixels. In this way, the generated feature map can maintain spatial resolution without consuming a lot of calculation time. DeepLab uses multiple convolutions, followed by a pyramid pooling method, where four parallel atrous convolutions are applied to the input, and all have different atrous rates.[24]

The fully convolutional densely connected convolutional network is established in [25], the feature maps are accumulated after the dense block and transition up process which plays the role of upsampling. The upsampling method reduces the volume of the feature maps and leads to a deep network with relatively small quantity of parameters. The model is applied to solve the semantic segmentation problem on urban areas, it performs well even without data pre-processing or pretraining.

In [26], the researchers provide a solution called efficient neural network to deal with high resolution image segmentation in data-driven applications on mobile and high-end GPU platforms. The input size are greatly reduced after initial blocks, thus helping to achieve good performance on real-time pixel wise segmentation. Due to the reduction of floating point operation, the computing resources required on long-time processing are largely cut down.

Another novel architecture used in cityscape and scene image segmentation is the pyramid scene parsing network mentioned in [27]. It utilizes a 4-level pyramid pooling module to extract context information given by the feature maps generated by a ResNet model. The output of the pyramid pooling module is the concatenation of multiple levels of extracted features as a global pyramid feature and are fed directly into convolutional classifier module afterwards.

A flexible architecture to decrease the information loss in very deep neural networks due to downsampling is introduced in [28]. The multi-path refinement network uses adaptive convolutional blocks with residual convolutional units that apply identity mapping, before entering a chained residual pooling block in the form of fusion of multiple resolution. The network aims at tackling the problem of overly compressed resolutions in image and is able to produce high resolution image segmentation predictions.

The architecture provided in [29] includes 2 parts, the label refinement network and the gated feedback refinement network. The first part adopts a novel way called coarse labeling grained prediction by combining features from previous layers, followed by refining towards better results at several stages. The second part extracts the ambiguity of information by applying gate units to control the information flow. The model works well on segmentation and labeling problems in a more generic context.

For medical image segmentation, an improved model of UNet++ is proposed in [30]. For each decoder layer, the UNet 3+ performs the incorporation of features maps on 5 scales to make a full scale aggregation of small or same scale feature maps from encoders and larger feature maps from decoders. It also applies deep supervision on each decoder stage to learn about representations. The model produces fine results on the refinement of labeling out the boundaries of organs and can avoid the over segmentation problems when given an image without organs.

### 3 Background

In this section, five medical image segmentation models used for the experiment are explained briefly with the help of the description, simplified network architectures and equations.

#### 3.1 Unet

Unet model is a breakthrough in the medical image, it can realize end-to-end training on medical images.

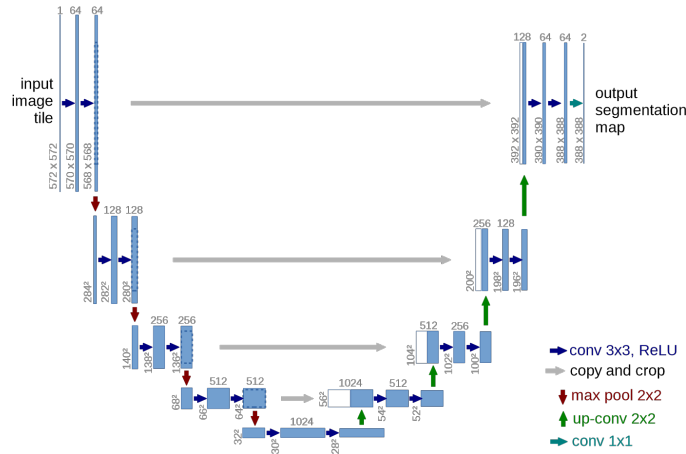


Figure 1: The network structure of Unet

The network structure of U-Net is shown in the figure 1. It has a contracting path on the left side and an expansive path on the right side. [6] The left side can be regarded as an encoder, and the right side can be regarded as a decoder. The encoder has four sub-modules, each of which contains two convolutional layers. After each sub-module, there is a down-sampling layer implemented by max pool. The decoder contains four sub-modules, and the resolution is sequentially increased by upsampling until it is consistent with the resolution of the input image. As the convolution uses the valid mode, the actual output is smaller than the input image. The network also uses a skip connection to connect the upsampling result with the output of the sub-module with the same resolution in the encoder as the input of the next sub-module in the decoder.

#### 3.2 Unet++

Medical image segmentation usually needs a higher accuracy than normal image segmentation because marginal segmentation errors in medical images can lead to strong uncertainties. Besides, inaccurate segmentation may also

lead to unsure changes in machine diagnosis.[7] The discovery of Unet++ helps recover more detailed information in medical images.

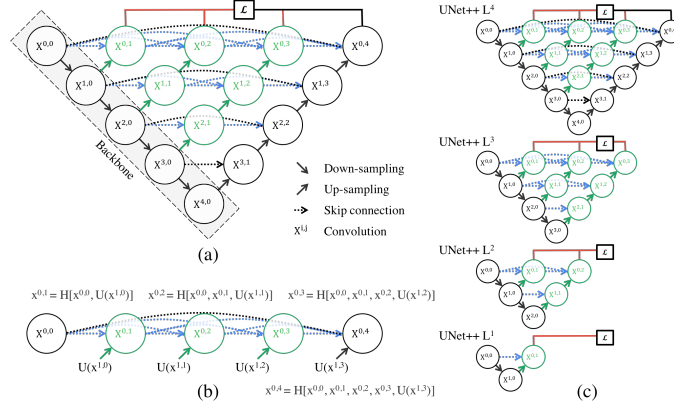


Figure 2: The detail of Unet++

Figure 2 shows some detail of the structure of the Unet++ network. The subfigure (a) describes that UNet++ improves the encoder and the decoder by adding nested dense convolutional blocks. So that UNet++ can bridge the semantic gap between the feature maps. Lines in subfigure (b) illustrates the first skip pathway in Unet++. Cause the Unet++ model can operate in the accurate mode in which all segmentation branches' outputs are averaged, and the fast mode which only calculates the final result by selecting on a branch, deep supervision has been used in Unet++. And the subfigure(c) shows the pruning procedure when the Unet++ model is trained in deep supervision.[7, 8]

### 3.3 SegNet

Figure 3 shows the structure of SegNet. The encoder network consists of 13 convolutional layers, which correspond to the first 13 convolutional layers in the VGG16 network [9], and every encoder network has its own corresponding decoder layer.[10] The encoder network alternately uses convolutional layers and max pooling layers. On the contrary, the decoder uses de convolutional layers and upsampling layers. And the softmax layer is used for pixel classification.



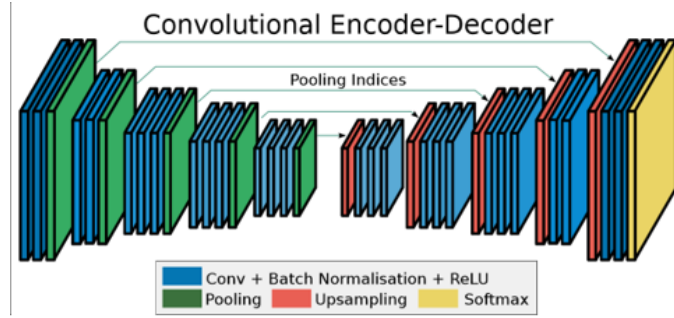


Figure 3: The model of SegNet

The max pooling layer can maintain translation invariance, so every pixel in the output feature map contains much of the input image's spatial information. Because of translation invariance, the decoder network has strong robustness while loses information of boundary details, which is a great disadvantage. So, a method that saves information of boundary details makes a great contribution to this network. However, it is impossible to save all encoder feature maps (after downsampling). Therefore, the model only stores the maximum pooling index, which is the location of the largest feature value in each pooling window used for each encoder feature map.

### 3.4 Seg-Unet

Figure 4 shows the structure of Seg-Unet. The Seg-Unet model combines U-Net and Seg-Net, for Seg-Net has auto-encoder architecture, and Unet is fit for medical image segmentation. This network is a newly developed one that not many researchers have studied.

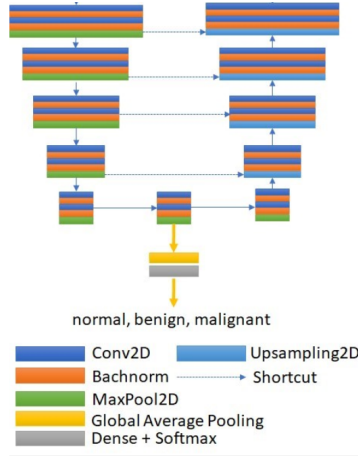


Figure 4: The model of Seg-UNet

### 3.5 Res-UNet

Res-UNet combines the advantages of U-Net architecture and Residual neural network. Generally the deeper the neural network, the better the performance of processing data, but usually there is no training algorithm to support it. Res-UNet balances the complexity and accuracy of network training to achieve better results.[15]

#### 3.5.1 U-Net

In order to obtain a good segmentation effect, use low-level information while retaining high-level semantic information. There are some good ways to solve this problem in U-net. There are many ways to solve this problem in U-net, one of which is that it can expand the data. In addition, U-net's own structure also helps to solve the training problem. U-net's own unique structure can not only supplement low-level language details to high-level semantic features, but also promote back propagation during training.[12]

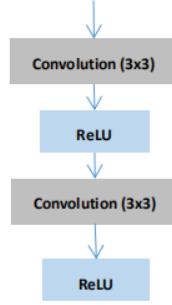


Figure 5: The architecture of U-Net

### 3.5.2 Residual Network

The idea of ResNet is to assume that we involve a network layer and there is an optimal network layer, so often the deep network we design has many network layers as redundant layers. Then we hope that these redundant layers can complete the identity mapping to ensure that the input and output passing through the identity layer are exactly the same. Which layer is the identity layer, this will be judged by yourself during network training.

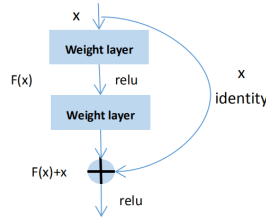


Figure 6: Residual Network: a building block

It can be seen from Figure 6 that  $X$  is the input of this layer of residual block, also called  $F(x)$  is the residual,  $x$  is the input value,  $F(X)$  is the output after the first layer of linear change and activation, the figure shows In the residual network,  $F(x)$  is added to the input value  $X$  of this layer before activation after the linear change of the second layer, and then output after activation. Add  $X$  before the activation of the output value of the second layer. This path is called a shortcut connection.

### 3.5.3 Res-Unet

All pre-activated residual units are used to build deep Res-Unet. Combining U-Net and Residual network will simplify the training of the network, build a

neural network with fewer parameters, and achieve better results in semantic segmentation. The specific network structure is shown in Figure 7.[12]

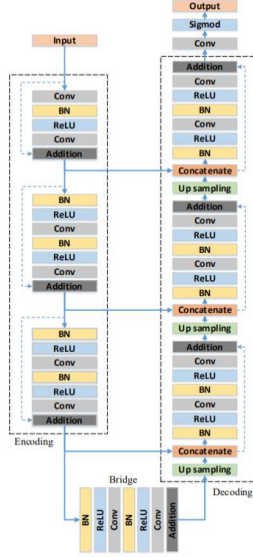


Figure 7: The architecture of ResUNet

## 4 Experiment and Results

In this section, the nuclei dataset used for the experiments, the evaluation measures, and lastly, the results are discussed in detail. It is important to note that the experiment was conducted under the Windows environment. The models were implemented in Tensorflow 1.15 and trained on the Nvidia GTX-1080Ti.

### 4.1 Dataset

The dataset is obtained from Kaggle 2018 Data Science Bowl[31], which consists of 735 cell’s nuclei images. The images were acquired under various conditions and vary in the cell type, magnification, and imaging modality. Each nuclei image is annotated with a series of masks marked by experts, and each image is represented by an associate ImageID. Files belonging to an image are contained in a folder with the ImageID, and within each folder, there are two subfolders:

1. “images” contains the image file
2. “masks” contains the segmented masks of each nucleus.

In the pre-processing step, both training and testing data were down-sampled to keep the images light and manageable. The testing data’s original sizes were recorded to up-sample the predicted masks and create correct run-length encodings later on. An example of nuclei image and mask are shown in Figure 8. Among the dataset, 670 are used as training data, and 65 are chosen as testing data to measure the networks’ performance.

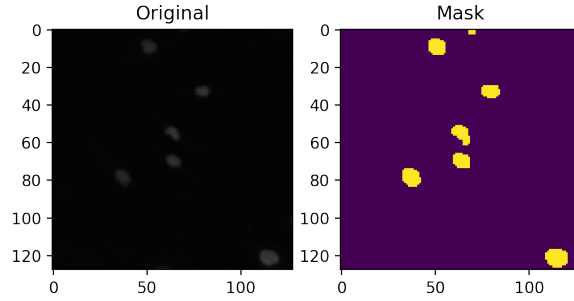


Figure 8: Example of nuclei image and mask.

## 4.2 Evaluation Measures

Generally speaking, semantic segmentation task performance can be measured based on two parameters: the Dice coefficient and the Jaccard coefficient. The Dice and the Jaccard coefficient is calculated by:

$$Dice(A, B) = \frac{2 \|A \cap B\|}{\|A\| + \|B\|} \quad (1)$$

$$Jaccard/IoU(A, B) = \frac{\|A \cap B\|}{\|A \cup B\|} \quad (2)$$

In functions (1) and (2), A and B are two segmentation masks for a given class,  $\|A\|$  is the norm of A, and  $\cap$ ,  $\cup$  are the intersection and union operators. Both the Dice and Jaccard indices are bounded between 0 (when there is no overlap) and 1 (when A and B match perfectly).

The Jaccard index is also known as Intersection over Union (IoU), and because of its intuitive and straightforward expression, it is widely used in computer vision applications. The IoU is used when calculating mean average precision. It is a number from 0 to 1 that specifies the overlap between the predicted and ground truth-bound boxes. An IoU of 0 means that there is no overlap between the boxes, and an IoU of 1 means that the union of the boxes is the

same as their overlap, indicating that they are completely overlapping. IoU is an important accuracy measure to track when gathering human annotations. The Dice coefficient also called the overlap index, is the most used metric in validating medical volume segmentations. It has been widely used as a statistical validation metric to evaluate reproducibility.

Suprosanna et al.[32] proposed a new metric called “clDice” that is able to evaluate tubular and linear structure segmentation based on intersecting skeletons with masks. Their finding is that training on clDice leads to segmentation with more accurate connectivity information, higher graph similarity, and often superior volumetric scores.

$$\text{Combined loss}(\alpha = 0.5) = (\alpha * \text{Dice coefficient}) + (\alpha * \text{clDice coefficient}) \quad (3)$$

In the experiment, we adopt IoU as metrics and the combined loss as training loss. The combined loss consists of both Dice and clDice coefficient, the function is shown in (3) where we take half of the result from both the Dice and the clDice coefficient separately to form the final loss.

### 4.3 Hyperparameter Setup

Training of models is carried out using the weight optimizer Adam(Adaptive Moment Estimation). The learning step for all architectures is 0.0001. The batch size is set to 16, and the total epochs for all architectures are 60, while the validation split rate is 10%. The early stopper is adopted with the patience of 15 steps to obtain the best-trained model. Meanwhile, the Reduce Learning Rate On Plateau function is used to avoid training struck on plateau. The function is set to monitor validation loss to adjust the learning step accordingly. The patience for the function is set to 5 steps, and the factor equals 10%.

### 4.4 Results

Figure 9 represents the loss curve of Unet, Unet++, and Res-Unet on the training dataset. Notice that the epoch for each model is different as we adopt the early stopper at the training stage. The orange line denotes the loss on the training dataset, while the blue line represents the loss on the validation dataset. As shown in Figure 9, the three models’ validation losses decrease as the number of epochs increases, and all of the three models are considered well fitted as there is no sign of overfitting or underfitting. However, the loss curve of SegNet and Seg-Unet is not that smooth as the three models mentioned above. According to Figure 10, overfitting happened on the SegNet model as training loss is lower than validation loss, and the difference between training and validation losses seems to continue to expand with epoch. Besides, the validation loss curve is not smooth enough as there are many obvious changes on the curve. There is no sign of overfitting found on the Seg-Unet, but both training and validation losses are too high, which suggests that the model requires more epoch to be trained.

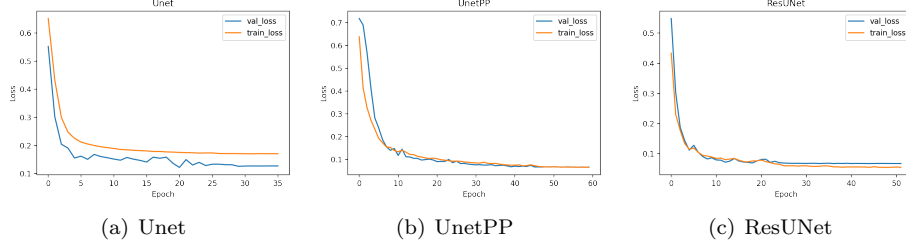


Figure 9: Loss Curves of Unet, Unet++ and Res-Unet

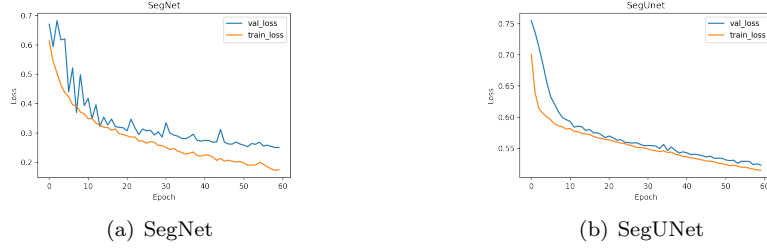


Figure 10: Loss Curves of SegNet and Seg-Unet

Figure 11 shows the segmentation results of different models on the testing dataset. The Unet, Unet++, and Res-Unet show the correct segmentation for objects on the image border, while SegNet and SegUnet did not achieve a satisfactory result. For Seg-Unet, all the segmentation results are larger than the ground truth, indicating that the model is underfitting. While for SegNet, most of the segmentation results are in the wrong shape and connected with each other. One of the reasons that could cause this phenomenon is because SegNet does not have a U architecture to down-sample and up-sample the image like Unet.

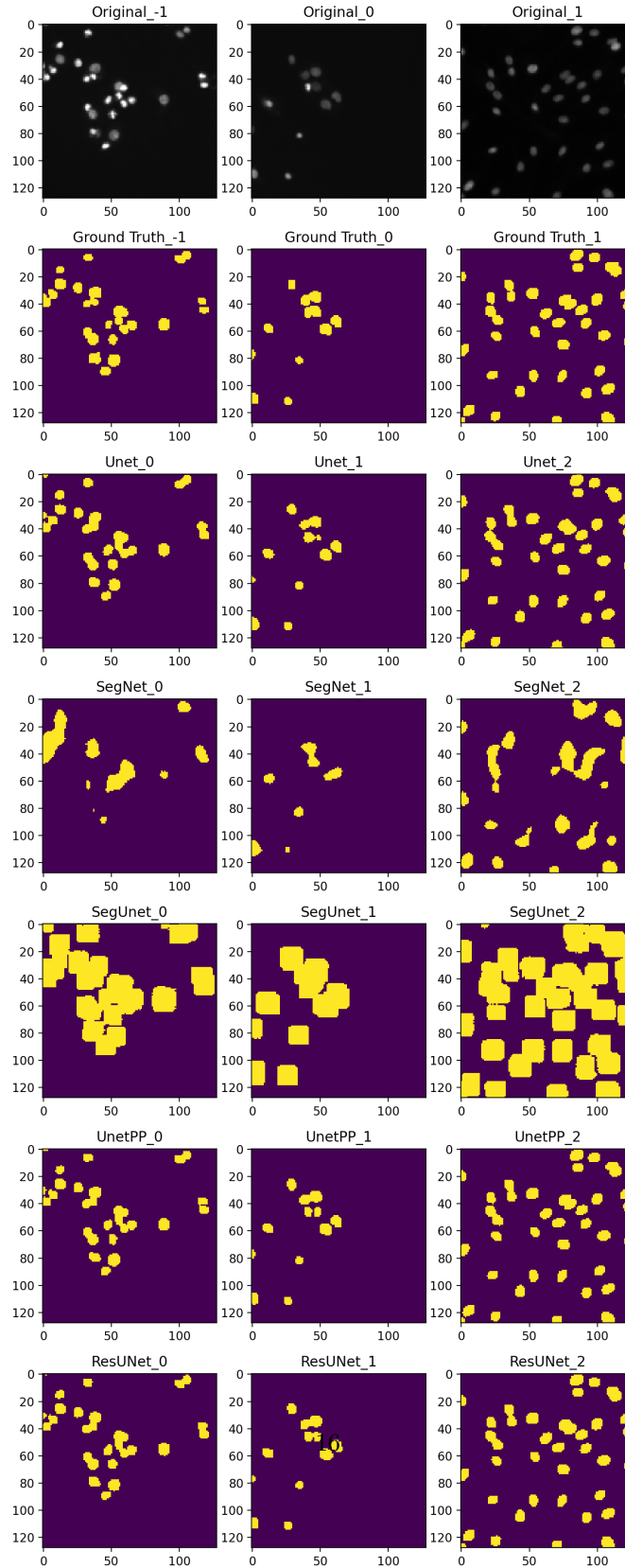


Figure 11: Original image, ground truth, and semantic segmentation performed with different models.



Table 1 compares the performance of the five different models on the testing dataset. As we can see from the table, Unet++ obtains the highest dice score(0.929) among all the models, followed by Res-Unet with a dice score of 0.9. The Unet has a relatively good score, while the performance of SegNet and Seg-Unet is not satisfactory, especially the SegUnet, which obtained the lowest score of 0.529.

	Unet	Unet++	Res-Unet	SegNet	Seg-Unet
Dice Coefficient(mean)	<b>0.864</b>	<b>0.929</b>	<b>0.92</b>	<b>0.734</b>	<b>0.529</b>

Table 1: Comparison of the neural network accuracy for the testing dataset

## 5 Conclusion

In the current study, we compared five different semantic segmentation models that include Unet, Unet++, Res-Unet, SegNet, and Seg-Unet. The comparison was performed on the Kaggle nuclei image dataset. The result indicates that Unet++ has the best quality of segmentation in comparison with other networks. Further studies are connected with changing encoder net of different models. Instead of using common encoders, it would be better to use ResNet and Inception for feature extraction.

## References

- [1] Dzung L Pham, Chenyang Xu, and Jerry L Prince. Current methods in medical image segmentation. *Annual review of biomedical engineering*, 2(1):315–337, 2000.
- [2] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and José García Rodríguez. A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857, 2017.
- [3] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of digital imaging*, 32(4):582–596, 2019.
- [4] Qiangguo Jin, Zhaopeng Meng, Tuan D Pham, Qi Chen, Leyi Wei, and Ran Su. Dunet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178:149–162, 2019.
- [5] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

- [6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [7] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165, 2018.
- [8] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multi-scale features in image segmentation, 2020.
- [9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [10] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.
- [11] Nhu-Tai Do, Sang-Don Joo, Hyung-Jeong Yang, Sung Taek Jung, and Soo-Hyung Kim. Knee bone tumor segmentation from radiographs using segunet with dice loss. In *Proc. of International Workshop on Frontiers in Computer Vision (IWFCV 2019)*, 2019.
- [12] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.
- [13] Chi-Hung Weng, Chih-Li Wang, Yu-Jui Huang, Yu-Cheng Yeh, Chen-Ju Fu, Chao-Yuan Yeh, and Tsung-Ting Tsai. Artificial intelligence for automatic measurement of sagittal vertical axis using resunet framework. *Journal of clinical medicine*, 8(11):1826, 2019.
- [14] Nikolas Lessmann, Bram van Ginneken, Pim A de Jong, and Ivana Išgum. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Medical image analysis*, 53:142–155, 2019.
- [15] Kun Zhang, Hongbin Zhang, Huiyu Zhou, Danny Crookes, Ling Li, Yeqin Shao, and Dong Liu. Zebrafish embryo vessel segmentation using a novel dual resunet model. *Computational intelligence and neuroscience*, 2019, 2019.
- [16] Foivos I Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, 2020.
- [17] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 225–2255. IEEE, 2019.

- [18] Sohil Shah, Pallabi Ghosh, Larry S Davis, and Tom Goldstein. Stacked u-nets: a no-frills approach to natural image segmentation. *arXiv preprint arXiv:1804.10343*, 2018.
- [19] Arthita Ghosh, Max Ehrlich, Sohil Shah, Larry S Davis, and Rama Chellappa. Stacked u-nets for ground material segmentation in remote sensing imagery. In *CVPR Workshops*, pages 257–261, 2018.
- [20] Tao Sun, Zehui Chen, Wenxiang Yang, and Yin Wang. Stacked u-nets with multi-output for road extraction. In *CVPR Workshops*, pages 202–206, 2018.
- [21] José Ignacio Orlando, Philipp Seeböck, Hrvoje Bogunović, Sophie Klimscha, Christoph Grechenig, Sebastian Waldstein, Bianca S Gerendas, and Ursula Schmidt-Erfurth. U2-net: A bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological oct scans. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1441–1445. IEEE, 2019.
- [22] Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine P Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *International workshop on simulation and synthesis in medical imaging*, pages 1–11. Springer, 2018.
- [23] Faizan Munawar, Shoaib Azmat, Talha Iqbal, Christer Grönlund, and Hazrat Ali. Segmentation of lungs in chest x-ray image using generative adversarial networks. *IEEE Access*, 8:153535–153545, 2020.
- [24] Yujie Wang, Simon Sun, Jahow Yu, Dr Yu, et al. Skin lesion segmentation using atrous convolution via deeplab v3. *arXiv preprint arXiv:1807.08891*, 2018.
- [25] Simon Jégou, Michal Drozdal, David Vazquez, Adriana Romero, and Yoshua Bengio. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 11–19, 2017.
- [26] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*, 2016.
- [27] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

- [28] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1925–1934, 2017.
- [29] Md Amirul Islam, Mrigank Rochan, Shujon Naha, Neil DB Bruce, and Yang Wang. Gated feedback refinement network for coarse-to-fine dense semantic image labeling. *arXiv preprint arXiv:1806.11266*, 2018.
- [30] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020.
- [31] Inc. Kaggle. 2018 data science bowl, 2018.
- [32] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Andrey Zhylka, Ivan Ezhov, Alexander Unger, Josien PW Pluim, Giles Tetteh, and Bjoern H Menze. cldice—a topology-preserving loss function for tubular structure segmentation. *arXiv preprint arXiv:2003.07311*, 2020.