UNDERSTANDING U-NET ARCHITECTURE

MINGXI LI , YIZHOU WANG, HAORAN YUAN, RUI YU, RAMESH KONATALA

INTRODUCTION

Identifying the cell nuclei is the starting point for most analyses in biomedical domain which helps to free biologists to focus on solutions.

U-Net is one of the most commonly used architecture which focuses on biomedical semantic segmentation. It has been proved that such a network can be trained end-to-end from very few images and out performs the prior best method.

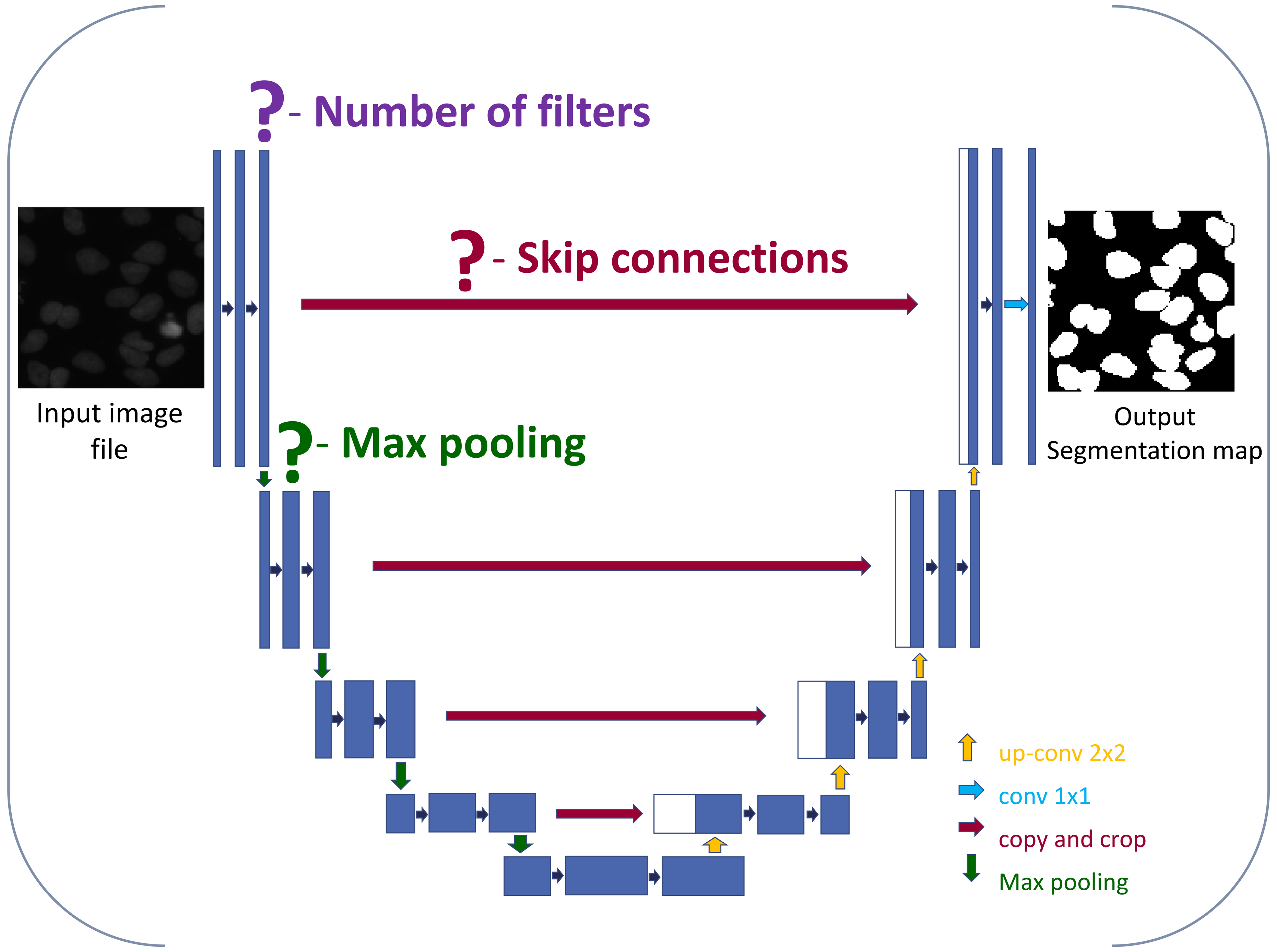
However the reason why U-Net has these properties is not clear. Our assignment aims to answer the question why U-Net needs less images to give the same performance as other networks do, and how different settings of parameters affects its performance. Eventually we find the best U-Net architecture based on our data set.

RESEARCH QUESTION

What is the contribution of the following Components from U-NET to the performance in segmentation tasks.

- The number of filters of each layer
- The skip connections
- The pooling layers
- The depth of the network

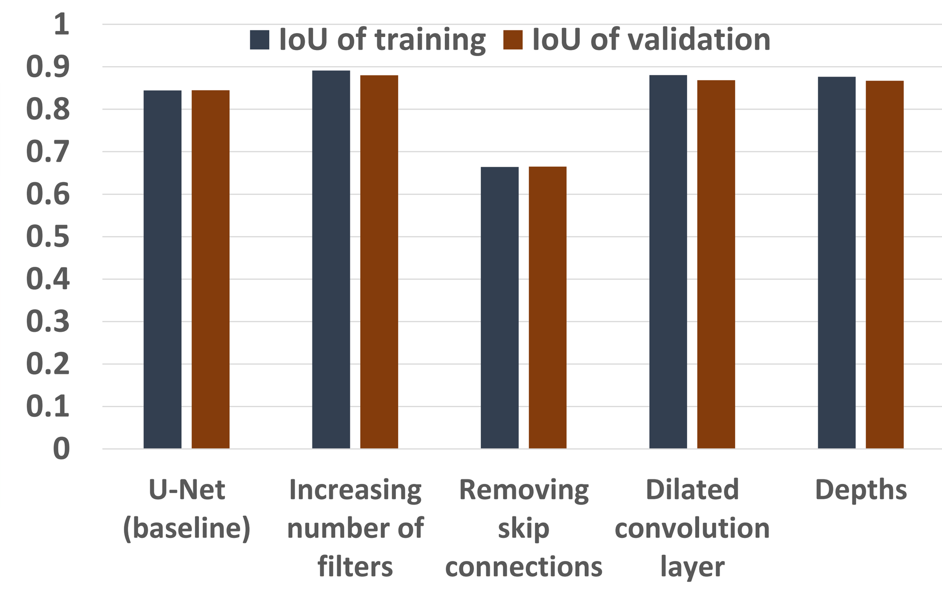
EXPERIMENT SETTING



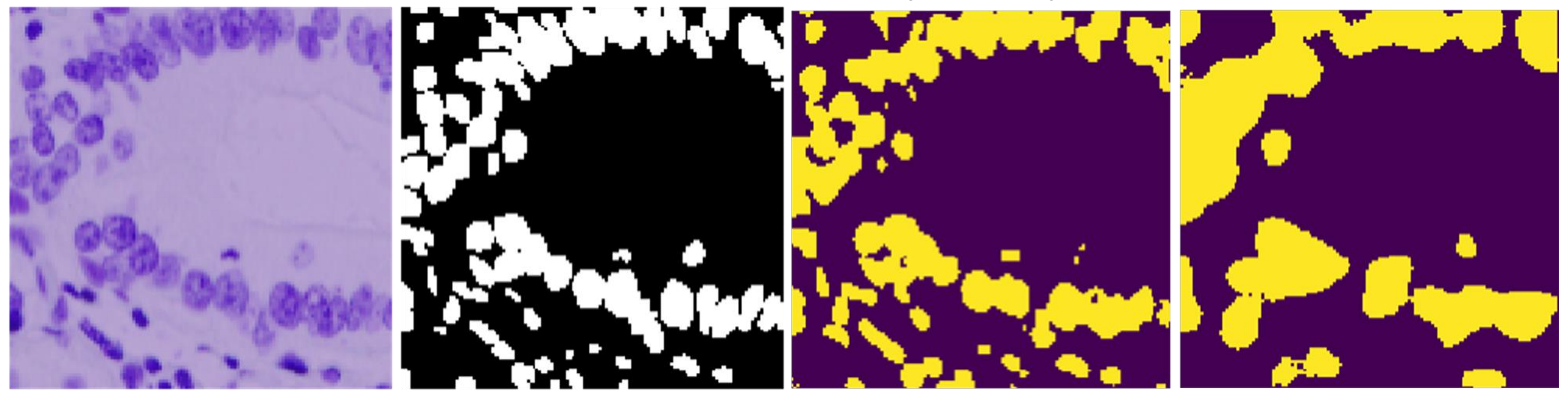
QUANTITATIVE RESULTS

Evaluation metric : $IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$

	IoU of training	IoU of validation
U-Net (baseline)	0.8446	0.8449
Increasing number of filters	0.8914	0.8799
Removing skip connections	0.6642	0.665
Dilated convolution layer	0.8805	0.8685
Depths	0.8765	0.8671



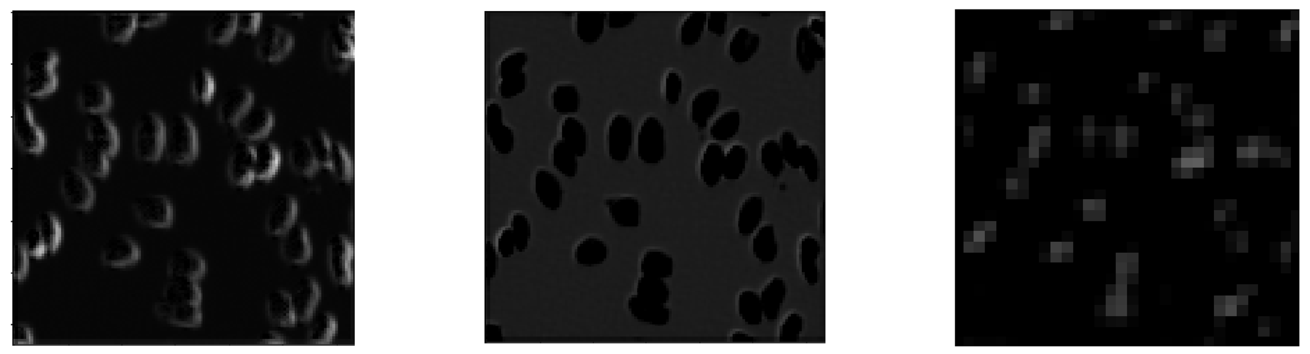
QUALITATIVE RESULTS



DISCUSSION

1. Increasing the number of filters can extract more non-linear features from the images

To understand what convolutional layers extract during training outputs of certain layers with different number of filters are visualized



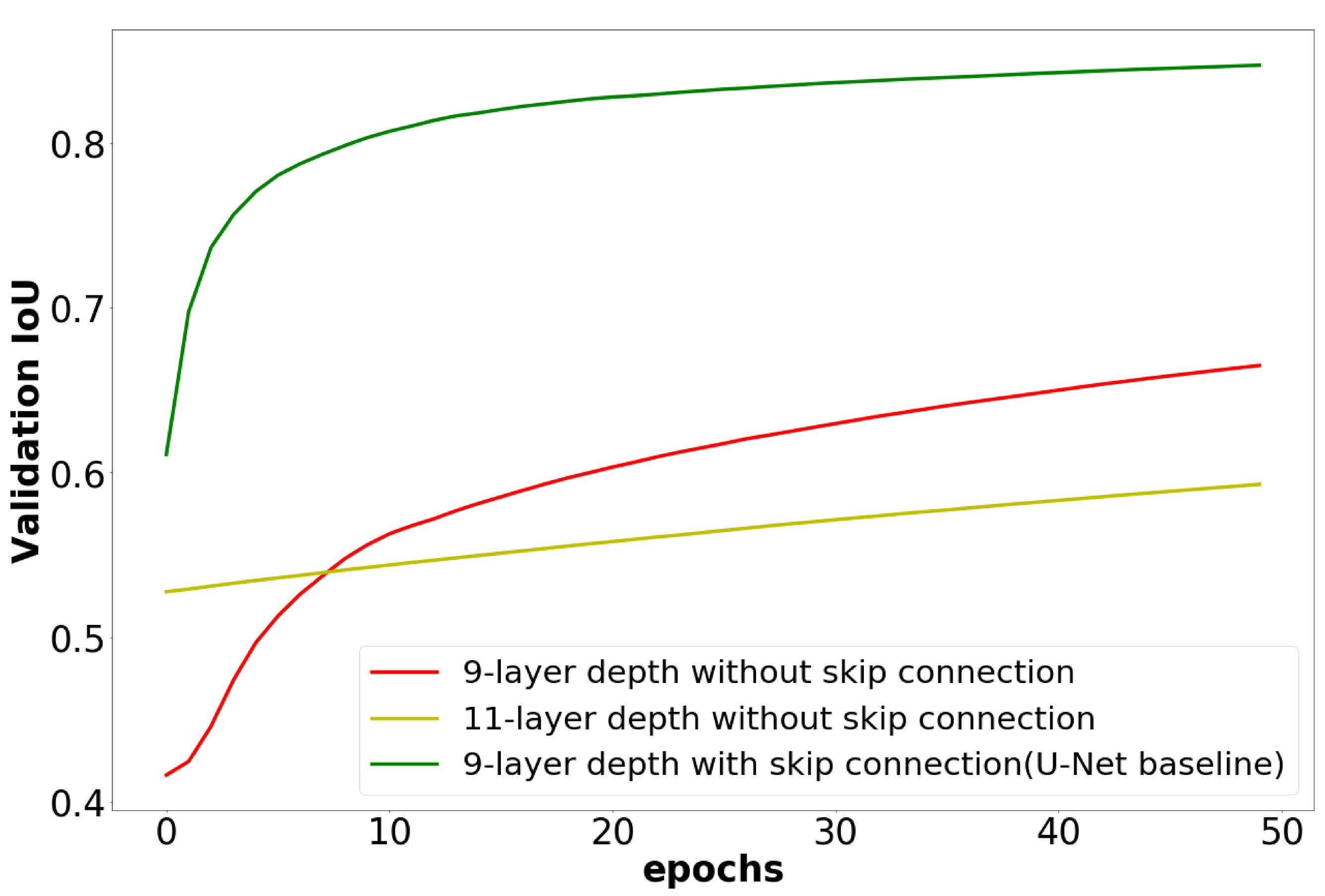
8 filters | 16 filters
Output of lower level convolutional layers

16 filters
Output of higher layer

2. Skip connections prove to be very important as they allow the network to propagate context information to higher resolution layers

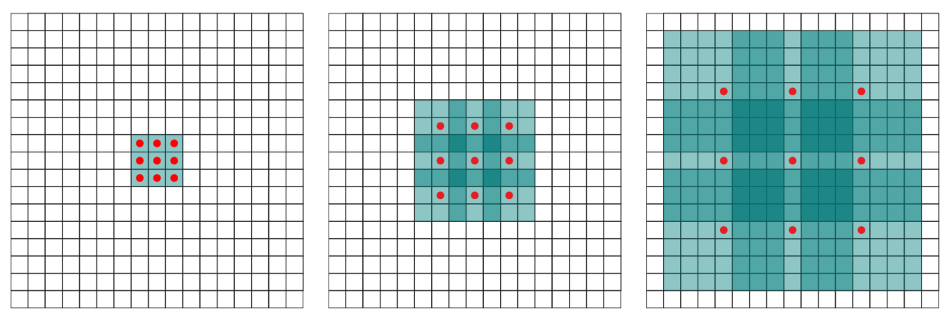
- Deeper layer model is failing to converge but we think that it can achieve convergence if we increase no. of epochs or by training model on larger dataset
- When we remove the skip paths from previous layers, even increasing the depths of the whole network will not boost the performance

Effect of depth/skip connections of architecture



3. By replacing the pooling layer with dilated convolutional layer, the performance of the model was improved

Dilated convolution can increase the receptive field without the information loss (no down sampling up-sampling process) unlike pooling layer.



Red dots are the inputs to a filter which is 3x3 in this example, and green area is the receptive field captured by each of these inputs.

CONCLUSIONS

- The skip connections contributes most to the good performance of U-Net because it allows the network to combine precise localization and high label non-linear features.
- Increasing the number of filters and depth of layer can also improve the performance.
- Introducing dilated convolutional layer will to some extent avoid information loss between down-sampling and up-sampling.

FUTURE WORK

- Introduce contour information in our model, like assigning more loss weights to the edge of nucleus.
- Implement active learning which can achieve similar performance with fewer data.

REFERENCES

- Olaf Ronneberger, Philipp Fischer, Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv:1505.04597
- Fisher Yu, Vladlen Koltun. "Multi-scale Context Aggregation By Dilated Convolutions." arXiv:1511.07122
- Matthew D Zeiler, Rob Fergus. "Visualizing and Understanding Convolutional Networks." arXiv:1311.2901