CSE 703: Dense U-Net for Biomedical Image Segmentation

Jun Zhuang **Zhenggang Xue** **Yuhao Du**

junzhuan@buffalo.edu zhenggan@buffalo.edu yuhaodu@buffalo.edu

Abstract

White matter hyper intensities (WMH) is one of main consequences of small vessel disease. Automated WMH segmentation techniques will play an important role in future clinical research and practice. Technically, U-Net can yield more precise segmentations. However, it sometimes losses more resolution as network goes deep. In addition, it usually depends on data augmentation or larger numbers of filters. To solve this problem, A new architecture, Dense U-Net is proposed to reinforce feature use, reduce the numbers of parameters as well as retain sufficient receptive fields without losing resolution. The experiment proves that our network achieves good performance and has significant improvement against the naive U-Net.

1. Introduction

Due to limited size of available dataset, traditional convolutional network has not been payed much attention by scientists even if it existed for a long time. Thanks to the breakthrough by Krizhevsky et al. [1], convolutional network achieved a great success since 2012. However, this network cannot deal with the input with arbitrary scales. Fully convolutional network (FCN) releases this constrains by replacing the the fully connected layer to the convolutional layer [2]. Especially in the field of biomedical image segmentation, a kind of FCN, so-called "U-Net" [3], is wildly used by many scientists in recent two years. Compared to FCN and SegNet [4], indeed, U-Net has more elegant architecture which can yield more precise segmentations while using fewer training dataset.

Recent work shows that convolutional networks are getting deeper and deeper. Nevertheless, U-Net will lose more resolution as network goes deeper. In order to gain more precise performance, U-Net may sometimes depend on data augmentation or larger numbers of filters. One of solution to the lost of resolution is using dilated convolution [5]. Dilated convolution can retain sufficient receptive fields without losing resolution even if network goes deeper.

Enlighten by DenseNet [6] and dilated convolution, this project proposes an advanced architecture, Dense U-Net, which implements dense connection and dilated convolution, which can reinforce feature reuse, reduce the numbers of parameters and retain lager receptive fields. Based on the dataset of WMH Segmentation Challenge (Grand Challenge) at MICCAI 2017, this model achieved dice coefficient 0.9358 in training set and 0.8329 in validation set without data augmentation. Compared to the naive U-Net model, moreover, the model save running time by 26% and also reduce the size of weight by 83%.

2. Dataset and Methods

2-1. Dataset

As one of main consequences of small vessel disease, white matter hyper intensities (WMH) can be well-recognized on brain MRI [7]. High quality of WMH segmentation can assist diagnosis in clinical practice. Thus, automated WMH segmentation techniques will play more and more important role in future clinical research and practice.

This dataset was acquired from three different hospital in Netherlands and Singapore. For each patient, T1-weighted image, multi-slice FLAIR image and manual label were provided.
In this project, the dataset was split by 4:1 in patient level. 80% of data were used as training set while 20% were used a testing set. Especially in training set, 20% of them were used as validation set.

This project used the brain MRI dataset for biomedical image segmentation.

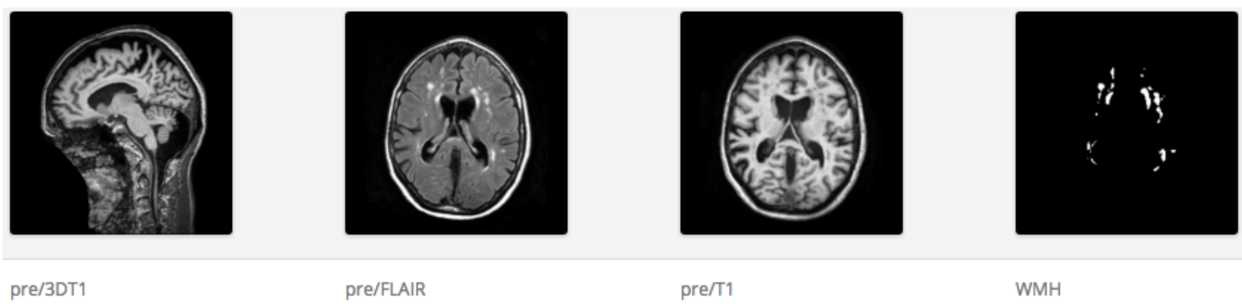

Figure 1: Left most image is the 3D T1 image with mask (remove the face); T1-weighted and FLAIR are used in our project as the two different channels; Right most image is the manual label.

2-2. Methods

This study proposes an advanced architecture, Dense U-Net, for WMH segmentation. The network has two main innovations, dense connection and dilated convolution. Specifically, this network has 10 layers. The 5th layer implements dilated convolution, whose dilated rate is equal to 2. In the rest layers from one to nine, batch normalization [8] and pre-activation, LeakReLU [9], are applied before first convolutional layer. As mentioned in the technical report of DenseNet [10], batch normalization can provide a unique scale and bias to previous input while pre-activation can reduce the error significantly. In addition, those layers have two convolutional layers with the same numbers of filters. Kernel size is 3x3. The number of filters from layer one to four is [16, 32, 64, 128]. Symmetrically, The number of filters from layer six to nine is [128, 64, 32, 16]. The 10th layer is a output layer with sigmoid activation function. From layer one to four, 2D MaxPooling layer is used as downsampling. From layer six to nine, deconvolution is implemented to up-sampling with 2x2 kernel size and strides. Its number of filters reduces 50% compared to the number of filters in previous convolutional layer. Both convolution and deconvolution use "same" padding and "He Normal" initializer [11]. Compared to naive U-Net, most importantly, our model use dense connection in layer seven to nine. For example, both layer one and two are concatenated back to layer nine after deconvolution. Other layers have the same pattern correspondingly.

This study used the model mentioned above to do the experiment and gained expected result. More details will be presented in next part.

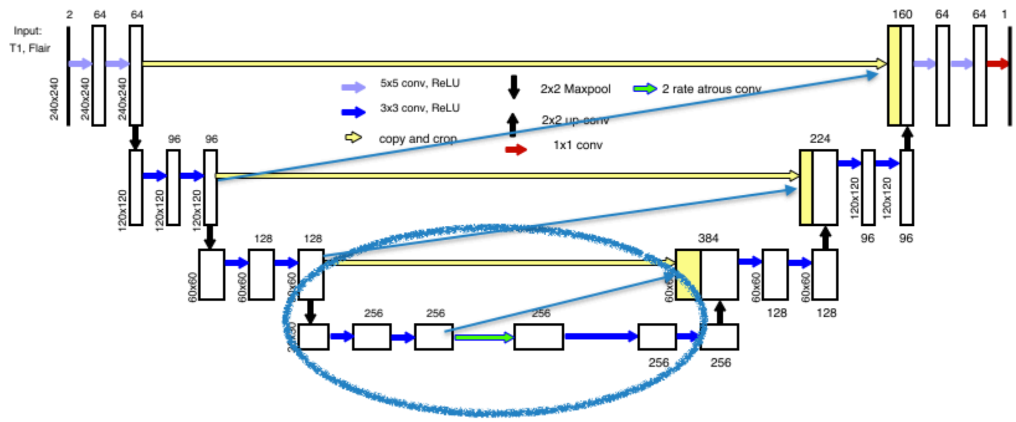

Figure 2: The architecture of Dense U-Net.

3. Experiment and Discussion

3-1. Experiment

3-1-1. Data Preprocessing

As mentioned before, this dataset comes from three different hospitals. Each of them contains 20 patients. At the beginning, each T1-weighted image and FLAIR image are combined into one two-channel image. Each patient's images will be removed 1/8 on both edges. Then, dataset is randomly split into training set and testing set on patient's level. After that, images are padded with zero or cropped for resizing to 240 by 240. Lastly, different set of data are stacked into different arrays correspondingly.

3-1-2. Training

This experiment is conducted in the platform of Amazon web service with the instance, p2.xlarge. This instance uses 4 virtual CPUs, 61 GB memory. This experiment uses NVIDIA Tesla K80 for computation.

Compared with stochastic gradient descent (SGD), Nadam [12] and Adam [13], this network finally is trained by Adam. Learning rate is set to be 1e-4 with step decay in the future epochs. Optimal batch size is 32. Using other sizes will decrease the performance. Here chooses dice coefficient to evaluate the performance. Thus, the loss function is negative dice coefficient. The dropout [14] was added in layer four and five but got worse result. Finally this regularization is not considered in our model. After 100 epochs' training, this model got the best result. However, different epochs are used in different kinds of experiment in our project.

3-1-3. Experiment Result

A. Naive U-Net vs Dense U-Net

Figure 3 shows that our model gained better performance both on training set and validation set. Especially, our model has more stable error curve. Table 1 shows that our model has better performance while use less running time and small size of parameter.

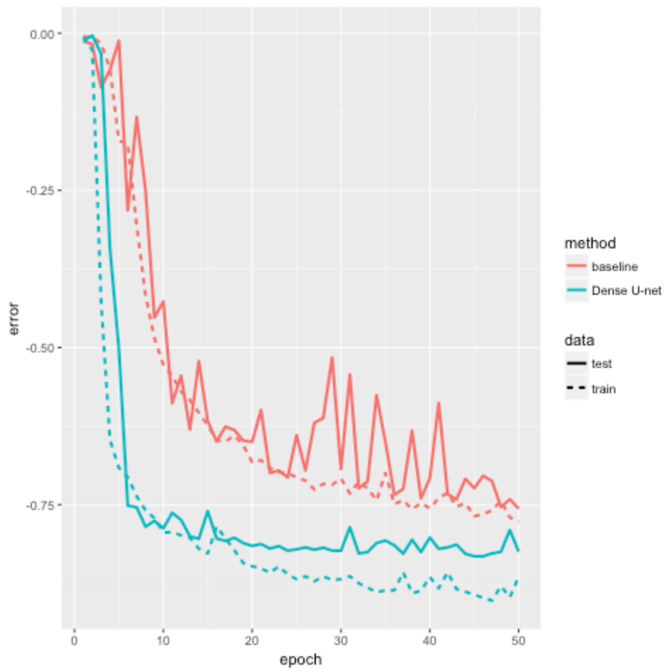

Figure 3: The error rate curve of Naive U-Net and Dense U-Net.

Model	Testing Dice Coef.	Avg. RunTime / Epoch	Weight
Baseline U-Net	0.71	92s	93.3 Mb
Dense U-Net	0.79	70s	16 Mb

Table 1: Dense U-Net has fast running time and small size of parameters.

B. Prediction

Figure 4 shows that our our model yields precise result and has high precision and recall.

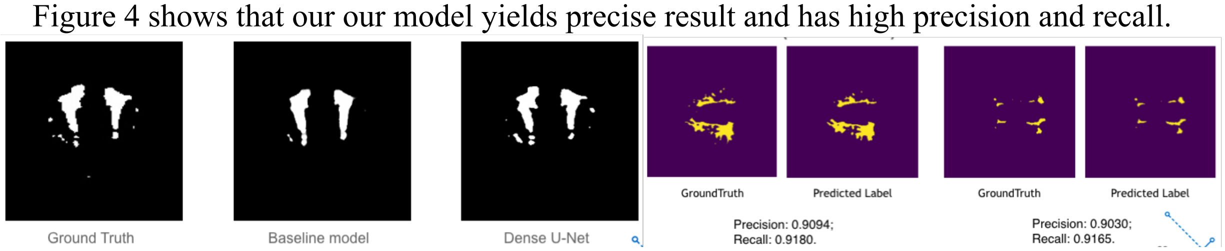

Figure 4: Predicted label of our model.

C. Dice coefficient compared to other models

Team	Approach	Dice (Testing)
Our Model	Dense U-Net	0.79
sysu_media	U-Net	0.8
clan	Gated Recurrent Units	0.78
nlp_logix	CNN	0.77
nic-vicorob	10 layers CNN	0.77
k2	U-Net	0.77
lrde	Deep FCN	0.73
misp	3D deep CNN with 18 layers	0.72
ipmi-bern	2 FCN with encoder and decoder	0.69
nih_cidi	U-Net	0.68
scan	Pooling-free FCN with dense skip	0.63

Table 2: Our model ranks 2nd compared to other models. Note that our preprocessing didn't use data augmentation.

3-2. Discussion

Based on the result shown above, our model does have better performance than the baseline model. Note that our model only use half number of filters. This reduction sacrifices accuracy of dice coefficient but reduce the running time and the number of parameters, which makes the model much more lightly. In addition, the predicted label images show that our model can output high quality prediction with high precision and recall.

4. Conclusions

This study proposes a new architecture, Dense U-Net, for WMH segmentation. Dense connection and dilated convolution make three main contributions: 1. Reinforce feature reuse; 2. Reduce the numbers of parameters and running time; 3. Retain sufficient receptive fields without losing resolution. The experiment shows that our model gains better performance against naive U-Net model as well as yields high quality prediction. However, this model has two shortcomings. Firstly, dense concatenation will occupy extra memory. What's more, this network is still too large and can be further compressed in order to achieve real time segmentation. This works are left to be open for further improvement.

Acknowledgments: Thanks Professor Gao for instruction and weekly helpful discussion.

5. References

[1] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012.

[2] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.

[3] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.

[4] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation”. In *arXiv:1511.00561*, 2015.

[5] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” In *ICLR*, 2016.

[6] G. Huang, Z. Liu, K. Q. Weinberger, “Densely connected convolutional networks”. In *CVPR*, 2017.

[7] J. M. Wardlaw et al. “Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration”. In *Lancet Neurology*, 2013.

[8] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

[9] A. L. Maas, A. Y. Hannun, and A. Y. Ng. “Rectifier nonlinearities improve neural network acoustic models”. In *ICML*, 2013.

[10] G. Pleiss, D. Chen, G. Huang, T. Li, L. van der Maaten, and K. Weinberger. “Memory-efficient implementation of densenets”. In *arXiv:1707.06990*, 2017.

[11] K. He, X. Zhang, S. Ren, and J. Sun. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”. In *ICCV*, 2015.

[12] C. Ye et al. “On the Importance of Consistency in Training Deep Neural Networks”. In *arXiv:1412.6980*, 2017.

[13] D. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In *arXiv:1412.6980*, 2014.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, Vol. 15, pp. 1929–1958*, 2014.