DA-U-Net: Densely Connected Convolutional Networks and Decoder with Attention Gate for Retinal Vessel Segmentation

Cong Wu, Yixuan Zou, Jinhao Zhan

Hubei University of Technology, No.28, Nanli Road, Hong-shan District, Wuchang, Wuhan, Hubei, China

Abstract. The segmentation of retinal vessels is greatly significant for doctors to diagnose the fundus diseases. However, existing methods have various problems in the segmentation of the retinal vessels, such as insufficient segmentation of retinal vessels, weak anti-noise interference ability. Aiming to the shortcomings of existed methods, this paper proposes an improved model based on the U-Net networks, which contains densely-connected convolutional network and a novel attention gate (AG) model, referred as Densely-Attention-U-Net (DA-U-Net), to automatically segment the retinal blood vessels. The method can alleviate the vanishing-gradient problem, strengthen feature propagation, substantially reduce the number of parameters, and automatically learn to focus on target structures without additional supervision. By verifying the method on the DRIVE datasets, the segmentation accuracy rate is 96.09%, higher than that of U-Net and R2U-Net.

1. Introduction

Medical image segmentation is a special computer vision area that is very important for many real-life applications. The fundus contains a large number of blood vessels, which is a deep microvascular system that can be directly observed in the human body without damage. It can provide doctors with a wealth of information about eye conditions and general system status. Ophthalmologists can detect early increases in the systemic vascular load caused by hypertension and diabetes, as well as early signs of vascular structural abnormalities such as retinal vein occlusion and retinal artery occlusion, the diseases caused by blood vessels and vascular systems can cause blindness. With the development of technology, it has been widely studied to explore an automatic method to segment retinal vessels, which can help doctors diagnose and analyze the patient's conditions and make early prevention and treatment of the fundus disease, so as to effectively avoid the visual loss caused by the degenerative disease.

In many cases of biomedical applications, a small number of objects is to be found, on the other hand, only small datasets can be acquired, class imbalance is present, and very high recognition quality and robustness is required[1]. Convolutional neural networks (CNN) have already demonstrated their success in image classification, image segmentation and object detection. For almost any computer vision problems, CNN-based approaches outperform other techniques and in many cases even human experts in the corresponding field. In this work we intend to provide a new approach to the medical image segmentation tasks, which is based on well-known and highly-performing U-Net[2] convolutional neural network (CNN) of encoder-decoder style where initial convolutional feature maps are skip-connected[3] to upsampled layers from bottleneck layers. This skip-connection is crucial to segmentation tasks as the initial feature maps maintain low-level features. We refer to the neural network built as DA-U-Net.
With the advent of convolutional neural networks (CNNs), near-radiologist level performance can be achieved in automated medical image analysis tasks. However, this approach leads to excessive and redundant use of computational resources and model parameters. For instance, similar low-level features are repeatedly extracted by all models within the cascade. To address this general problem, we propose a simple and yet effective solution, namely attention gates (AGs), and we replace the convolution layer with a Denseblock at the same time. CNN models with AGs can be trained from scratch in a standard way similar to the training of a FCN model, and AGs automatically learn to focus on target structures without additional supervision. The Denseblock can alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters. We demonstrate the implementation of AG and the Denseblock module in a standard U-Net architecture (DA-U-Net) and evaluate DA-U-Nets on DRIVE datasets, very high segmentation quality has been achieved.

2. Related Work
The retinal vascular network is the only deep blood vessel that can be observed directly with non-invasive means in the human body. The changes of characteristics or morphology in the microvasculature can be leaded by any pathological changes of systemic and hematological characteristics. The segmentation of retinal blood vessels is the key step in retinal image processing and analysis, it has a good research value for the early prevention and diagnosis of systemic and hematological diseases. The characteristics of retinal images are complex, the automatic segmentation of retinal blood vessels is easy to be affected by external conditions and pathological changes, and in the retinal images, the difficulty of segmentation is increased by the reason that contrast ratio of tiny vessels and its background is low, therefore, to improve the segmentation accuracy is an important topic to research.

Convolutional neural networks (CNNs) have become the dominant machine learning approach for visual object. We embrace this observation and introduce the Densely Convolutional Network (DenseNet), which connects each layer to every other layer in a feed-forward fashion. Whereas traditional convolutional networks with L layers have L connections—one between each layer and its subsequent layer—our network has L (L^2+1) direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. Whilst following a simple connectivity rule, DenseNets naturally integrate the properties of identity mappings, deep supervision, and diversified depth. They allow feature reuse throughout the networks and can consequently learn more compact. DenseNets have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters[4]. The Denseblock can be summarised as follows Figure 1:

![Figure 1](image)

**Figure 1.** A 5-layer dense block with a growth rate of k=4. Each layer takes all preceding feature-maps as input.
Attention Gates: AGs are commonly used in natural image analysis, knowledge graphs, and language processing (NLP) for image captioning[5], machine translation[6], and classification[7, 8] tasks. Initial work has explored attention-maps by interpreting gradient of output class scores with respect to the input image. Trainable attention, on the other hand, is enforced by design and categorised as hard- and soft-attention. Hard attention, iterative region proposal and cropping, is often non-differentiable and relies on reinforcement learning for parameter updates, which makes model training more difficult. Recursive hard-attention is used in[9] to detect anomalies in chest X-ray scans. Contrarily, soft attention is probabilistic and utilises standard back-propagation without need for Monte Carlo sampling. For instance, additive soft attention is used in sentence-to-sentence translation[10] and more recently applied to image classification[7,8]. Channel-wise attention is used to highlight important feature dimensions, which was the top-performer in the ILSVRC 2017 image classification challenge. Self-attention techniques have been proposed to remove the dependency on external gating information. For instance, non-local self attention is used in[11] to capture long range dependencies. The Attention Gate can be summarised as follows Figure 2:

Figure 2. Schematic of the proposed additive attention gate (AG). Input features are scaled with attention coefficients ($\alpha$) computed in AG. Spatial regions are selected by analysing both the activations and contextual information provided by the gating signal ($g$) which is collected from a coarser scale. Grid resampling of attention coefficients is done using trilinear interpolation.

3. Approach
In this experiment, we have 40 retinal vessels images, and we use 30 images to train, 5 images to validation, and 5 images to test. Data augmentation is essential to teach the network the desired invariance and robustness properties, when only few training samples are available. In case of microscopical images we primarily need shift and rotation invariance as well as robustness to deformations and gray value variations. Especially random elastic deformations of the training samples seem to be the key concept to train a segmentation network with very few annotated images. Data Augmentation refers to the method of expanding the original data through a series of random transformations to increase the amount of data. The new data set generated by this method that cannot find any two identical images. This helps to suppress overfitting and thus improves the generalization ability of the deep learning model. This paper selects translation, rotation, shearing and scaling for data augmentation.

The presented improved model, which we refer to as DA-U-Net, is depicted on Figure 4. It consists of basic blocks, and each of them follows the encoder-decoder architecture similar to U-Net[2]. We consider two kinds of basic blocks: Denseblock and Attention gate. They both feature connections (shown on the Figure 4), linking layers of the encoder and decoder, which are of very high importance. In this paper, we replace the normal convolution layers as the Bottleneck layers in our network. Although each layer only produces k output feature-maps, it typically has many more inputs. It has been noted in [12, 13] that a 1×1 convolution can be introduced as bottleneck layer before each 3×3 convolution to reduce the number of input feature-maps, and thus to improve computational efficiency. We find this design especially effective for DenseNet and we refer to our network with such a bottleneck layer to the BN-ReLU-Conv (3×3), as shown in the Figure 3. And in the stage of the
decoder, we added Attention gate mechanism, which can automatically learn to focus on target structures without additional supervision.

![Figure 3](image-url)  
**Figure 3.** The DenseBlock in DA-U-Net architecture.

As a loss function, we use \( l(A; B) \):

\[
l(A; B) = -\log d(A; B);
\]

where:

\[
d(A; B) = \frac{2}{i,j} \sum a_{ij}b_{ij} - \sum \frac{a_{ij}^2 + b_{ij}}{a_{ij}}
\]

where \( A = (a_{ij})_{i=1,j=1} \) is a predicted output map, containing probabilities that each pixel belongs to the foreground, and \( B = (b_{ij})_{i=1,j=1} \) is a correct binary output map.

![Figure 4](image-url)  
**Figure 4.** DA-U-Net architecture | U-Net model with DenseBlock and Attention Gate in decoder. Schematic of the AGs is shown in Figure 2. Feature selectivity in AGs is achieved by use of contextual information (gating) extracted in coarser scales.

### 4 Experiment

Experimental data provided by the DRIVE datasets, which is consisted of 40 color retinal images in total, in which 20 samples are used for training and remaining 20 samples are used for testing. The size of each original image is 565×584 pixels[14]. The data set is 900 sets of images generated after...
adopting the data augmentation method. In the training stage, we choose the Adam as the optimization, and the learning rate is 0.0001. The experimental platform is based on Windows 7 64-bit operating system, i7-4790 CPU, 16G memory, GTX 750, using a deep learning framework for TensorFlow.

In order to further prove the effectiveness of the proposed method on retinal vascular segmentation, the comparison with the best found methods for the segmentation of retinal vessels in DRIVE datasets is presented in Table 1. Sensitivity (SE) indicates the percentage of correctly segmented blood vessel pixels as a percentage of true blood vessel pixels, and the specificity (SP) is the percentage of correctly segmented background pixels to the true background pixels. The accuracy (ACC) is the percentage of correctly segmented blood vessels and background pixels in the entire image.

**Table 1.** Segmentation performance of retinal vessel on the DRIVE datasets.

| Methods                  | SE   | SP   | ACC  |
|--------------------------|------|------|------|
| Liskowsk [15]            | 0.7763 | 0.9768 | 0.9495 |
| Qiaoliang Li [16]        | 0.7569 | 0.9816 | 0.9527 |
| U-Net [17]               | 0.7537 | 0.9820 | 0.9531 |
| Residual U-Net [17]      | 0.7726 | 0.9820 | 0.9553 |
| R2U-Net [17]             | 0.7792 | 0.9813 | 0.9556 |
| DA-U-Net                 | 0.7539 | 0.9904 | 0.9609 |

The different prediction between the DA-U-Net and U-Net in the DRIVE can be made based on Figure 5.

![Figure 5](image)

**Figure 5.** The figure shows the different prediction between the DA-U-Net and U-Net. From the Segmented result by U-Net, We can see there are many small blood vessel segmentation effects that are not good, the effect in (d) is obviously better than (c).
5. Conclusion and Discussion

We present the model for the medical image segmentation based on Denseblock and Attention gate used in the well-known U-Net models. And the Segmented result by DA-U-Net is obviously better than the U-Net. We verified the feasibility of the proposed method on the DRIVE datasets, and the accuracy rates were 96.09%. By analyzing and comparing the segmented retinal blood vessel images, the proposed method is more advantageous than other methods. It can provide doctors with a wealth of information about eye conditions to help patients receive timely treatment. Still, our results fail to detect very thin vessels that span only 1 pixel. We expect that additional prior knowledge on the vessel structures such as connectivity may leverage the performance further.

The use of artificial intelligence technology for medical image-assisted diagnosis is a rapidly developing field. With the rapid development of deep learning technology and the continuous opening of medical imaging data, artificial intelligence is gradually maturing in early stage disease screening and doctor-assisted diagnosis.

Acknowledgments

I would like to extend my sincere gratitude to my teacher, Cong Wu, for his instructive advice and useful suggestions on my thesis. I am deeply grateful of his help in the completion of this thesis.

References

[1] Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on, IEEE

[2] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer (2015)

[3] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770 – 778 (2016)

[4] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:2261-2269

[5] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., Zhang, L.: Bottom-up and top-down attention for image captioning and vqa. arXiv preprint arXiv:1707.07998 (2017)

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser,.. Polosukhin, I.: Attention is all you need. In: NIPS. pp. 6000 – 6010 (2017)

[7] Jetley, S., Lord, N.A., Lee, N., Torr, P.: Learn to pay attention. In: International Conference on Learning Representations (2018)

[8] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: IEEE CVPR. pp. 3156 – 3164 (2017)

[9] Ypsilantis, P.P., Montana, G.: Learning what to look in chest X-rays with a recurrent visual attention model. arXiv preprint arXiv:1701.06452 (2017)

[10] Shen, T., Zhou, T., Long, G., Jiang, J., Pan, S., Zhang, C.: Disan: Directional self-attention network for rnn/cnn-free language understanding. arXiv preprint arXiv:1709.04696 (2017)

[11] Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. arXiv preprint arXiv:1711.07971 (2017)

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In CVPR, 2016

[14] Staal, Joes, et al. "Ridge-based vessel segmentation in color images of the retina." IEEE
transactions on medical imaging 23.4 (2004): 501-509

[15] Liskowski, Paweł and Krzysztof Krawiec. "Segmenting Retinal Blood Vessels With Deep Neural Networks." IEEE transactions on medical imaging 35.11 (2016): 2369-2380

[16] Li, Qiaoliang, et al. "A cross-modality learning approach for vessel segmentation in retinal images." IEEE transactions on medical imaging 35.1 (2016):109-118

[17] Alom, Md Zahangir, et al. "Recurrent Residual Convolutional Neural Network based on U-Net(R2UNet) for Medical Image Segmentation." arXiv preprint arXiv:1802.06955 (2018)