CS482/682 Final Project Report Group 28

Retinal Vessel Image Segmentation using Attention Mechanism

Lin Cheng/lcheng29, Kefan Song/ksong13, Tianyi Ye/tye5, Yi Zhang/yzhan496

1 Introduction

Background Retinal vessel segmentation studies the variations of retinal vessels' morphological features that could indicate possible vision related diseases. Automatic segmentation and analysis of the vascular features allow for faster diagnosis of such diseases. U-Net based architectures are commonly used in solving such problems, but they lack ability in detecting very fine structures. Attention is a technique that mimics cognitive attention, which enhances some parts of the input data while diminishing other parts. Thus, the objective of this project is to improve U-Net's ability to detect tiny retinal vessels by implementing attention mechanisms.

Related Work Li *et al.*[1] proposed a residual U-Net architecture for the task, which includes several residual blocks with strong dropouts and batch normalization operation to increase the model capability and tackle the overfitting problem. Guo *et al.*[2] designed a spatial attention U-Net(SA-UNet) which infers the attention map along the spatial dimension and multiplies the attention map by the input feature map for adaptive feature refinement.

2 Methods

Dataset The DRIVE dataset[3] that this project utilizes includes 40 TIF images obtained from a Dutch diabetic retinopathy screening program. These color fundus images, each captured using 8 bits per color plane at 768 × 584 pixels, are equally split into training and testing sets by the data source site. We cropped each image into a 512 × 512 snippet as model input and used 25% of the training set as the validation set. In addition, a mask image is available for every retinal image, indicating the region of interest.

Setup, Training, Evaluation Using U-Net[4] as our baseline, we introduced four types of attention mechanisms (attention block, spatial, channel, spatial-channel) on top of it (See Appendix for detailed layout of architectures). Both attention block (ABU) and spatial attention (SAU) utilize the inter-spatial relationship of features and generate a spatial attention map by extracting a feature descriptor from each input channel and then applying a convolution layer on the descriptor. The feature descriptor of the attention block was extracted using convolution, while that of the spatial attention was extracted using additional pooling layers. Channel attention (CAU) utilizes the inter-channel relationship of features and generates a channel attention map by extracting a feature descriptor across all channels and then applying a convolution layer on it. Spatial-channel attention (SCAU) is the combination of spatial and channel attention, and its attention map is the product of the maps from the two mechanisms. These attention maps adjust the weight of the feature maps extracted by the U-Net encoder to focus on certain features while suppressing others. Parameters were tuned to be optimal using the validation set. All five networks mentioned above were trained with the optimal batch size of 3, and ran for five trials. SGD was used as the optimizer, with a learning rate of 0.01 and momentum of 0.9. Binary Cross Entropy was used as the loss function for all models. The models were trained for 500 epochs with early-stopping implemented. The model performance is then evaluated using the DICE score, precision, recall, accuracy and ROC AUC score. The value of each performance metric is calculated by taking the mean of five trials.

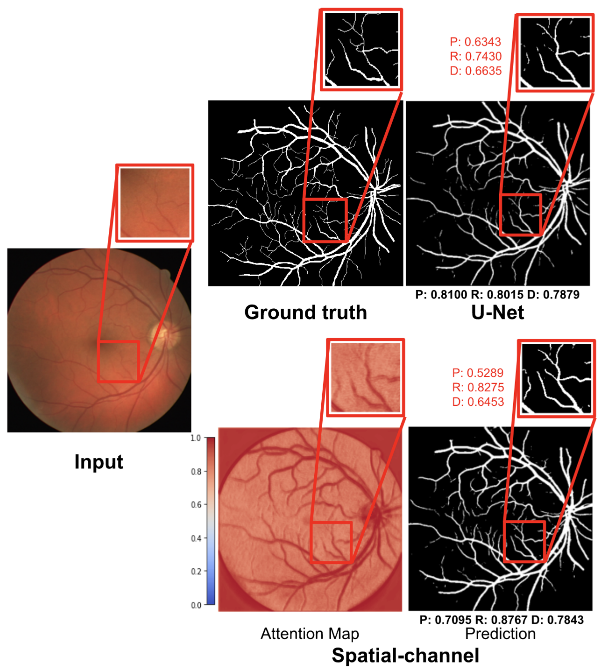

Figure 1: Sample Test Case: Predictions of U-Net and SCAU and Their Corresponding Performance (See Appendix for Full Image)

3 Results

	U-Net	ABU	SAU	CAU	SCAU
DICE	0.801	0.793	0.801	0.797	**0.804**
AUC	0.977	0.976	0.978	0.977	**0.978**
ACC	0.956	0.956	**0.957**	0.952	0.950
PRC	**0.822**	0.811	0.808	0.769	0.748
Recall	0.778	0.791	0.809	0.830	**0.854**

Table 1: Model Performance Comparison

The Dice, precision (PRC), recall, accuracy (ACC) and ROC AUC scores of the five models are shown in Table 1. As we can see, the SCAU has the highest dice score, recall and ROC AUC scores. Compared to the U-Net baseline, SCAU increases recall significantly while sacrificing PRC, and also achieved better predictions on small vessels. This is demonstrated using a randomly selected test case, as shown in Figure 1. In addition, a region containing only small vessels is extracted from the predicted mask and evaluated using all five metrics. A corresponding visualization of the spatial-channel attention map for the extracted region is also displayed to show the focus of the attention mechanisms.

4 Discussion

Overall, the Dice/AUC/ACC score for all models are similar, which is reasonable since they all can finely segment large vessels that take up the majority of the pixels. Since the ability to segment small vessels, which take on much lesser pixels, can hardly be interpreted using these metrics, we resort to PRC and recall for further analysis. Table 1 shows that when moving from left to right, the PRC decreases and recall increases. This is because getting the precise boundary of the foreground is a common issue in segmentation, and such issue is magnified by attention mechanisms, which focus more on the small vessels. This led to thickening of small vessels, which decreases false negative and increases false positive, thus the lower PRC and higher recall without significant change in ACC.

Judging from these results, whether the attention mechanism enhances retinal vessel segmentation depends on the task objective. Attention-integrated U-Nets are not suitable for tasks that require precise vessel diameter, but for tasks that prioritize the general shape of vessels over their quantitative properties, implementing the attention mechanism can provide a lot more information compared to the regular U-Net approach. In the future, as model capacity increases with additional attention mechanisms, residual logic can be included to achieve better performance. Further data augmentation methods and dropout layers may be necessary to alleviate the potential problem of overfitting. Alternative loss functions may also be implemented to achieve better precision while maintaining the recall score.

5 Reference

[1] Li, D., Dharmawan, D. A., Ng, B. P., Rahardja, S. (2019, September). Residual u-net for retinal vessel segmentation. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 1425-1429). IEEE.

[2] Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., Fan, C. (2021, January). Sa-unet: Spatial attention u-net for retinal vessel segmentation. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 1236-1242). IEEE.

[3]Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. IEEE transactions on medical imaging, 23(4), 501-509.

[4]Ronneberger, O., Fischer, P., Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

6 Appendix

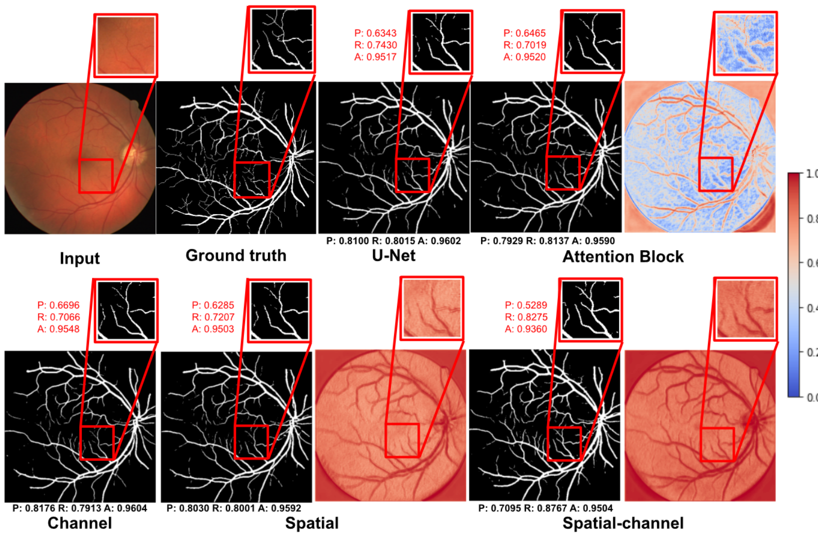

Figure 2: Sample Test Case: Predictions of Five Models and Corresponding Performance

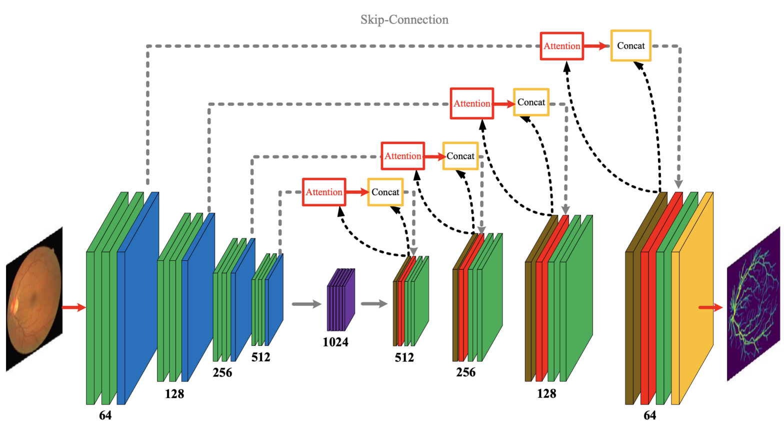

Figure 3: Attention Model Architecture

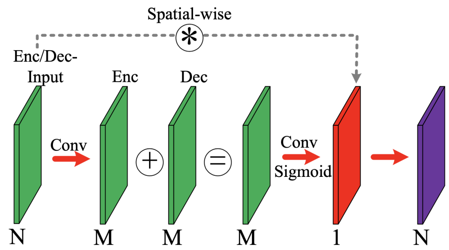

Figure 4: Attention Block

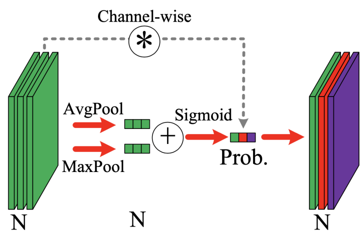

Figure 5: Channel Attention

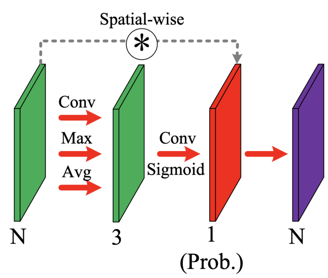

Figure 6: Spatial Attention

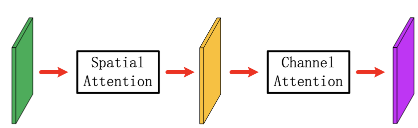

Figure 7: Spatial-Channel Attention