COMP9444 Project Summary

Digital Retinal Images for Vessel Segmentation

Qingru Liu: z5449255

Jeremy Lu: z5536953

Ash Peng: z5473493

Anthony Chen: z5581750

Peilin Tan: z5530348

July 2024

1 Introduction

Our project is to develop an efficient semantic segmentation algorithm to achieve retinal blood vessel segmentation, which is a key part of the diagnosis and treatment of various eye diseases. To address this, we implemented a high-precision and efficient method based on deep learning to automatically complete the task of retinal blood vessel segmentation. First we selected the model for baseline experiments, and then chose U-Net as our modified architecture, introducing the ASPP module in DeepLabv3 and Attention Block to improve the feature extraction ability of the Endoer part of the network, thereby improving the accuracy and efficiency of automatic retinal blood vessel segmentation.

2 Related Work

In our research, we conducted surveys of the models of semantic segmentation. There are various types of models were designed for this task. FCN is a pioneering model that replaced the fully connected layers with convolutional layers[5], and DeepLabv3 [1] is another popular used model that introduces atrous convolutions and spatial pyramid pooling to capture multi-scale information to make it more effective for semantic segmentation tasks. And the U-Net model, designed for biomedical image segmentation, with an encoder-decoder structure with skip connections, which has proven to be very effective for image segmentation.[2] In the field of medical image semantic segmentation, U-Net is one of the most widely used models because of its ability to handle different object sizes and the high accuracy. Many variant models based on U-Net have been proposed. ResU-Net introduces residual blocks into the U-Net architecture, making the network deeper and significantly improving segmentation accuracy.[4] Attention U-Net introduces an attention mechanism, improve the ability of the model to focus on relevant features in the image, improving the segmentation effect of small and complex structures. [6] These U-Net variants have shown excellent performance in medical image segmentation and have become our source of inspiration.

3 Methods

Method Overview: In this project, we used a U-Net-based model as the foundation and improved model performance by integrating custom Feature Fuse modules and Residual Block modules.
Motivation: After conducting a series of experiments with baseline models, we found that the U-Net architecture performs better in our task. Therefore, we decided to make improvements on the U-Net architecture. Simultaneously, to enhance the feature extraction capability of the encoder, we incorporated the ASPP module from DeepLabv3. This combination aims to fully leverage the effectiveness of U-Net in image segmentation tasks while enhancing the capture of image details through the ASPP module.
Modified Model implementation:

- **Introduce Feature Fuse Module:** We modified and introduce this module to enhance the feature extraction in model's encoder. The Feature Fusion module is similar to the ASPP module in DeepLabV3, extracting multi-scale features through parallel convolutions. However, our implementation focuses more on fusing features of different scales using addition operations instead of concatenation. Additionally, batch normalization layers further standardize the output, enhancing the model's stability. This module contains a 1×1 and a 3×3 convolution layer, a 3×3 dilated convolution layer. Additionally, dilated convolution can increase the receptive field without increasing computational complexity.

- **Modified Residual Block:**This module replaces the original convolution block in the upsampling part, adds a convolution and BN in the Residual Block, replaces the ReLU with Leaky ReLU, and adds a dropout layer before the final normalization to prevent overfitting.

- **Introduce Attention Block**: We also introduce this block refer to Attention UNet to focus more on key feature areas in the image, by weighting important parts of the image (such as lesion areas or specific structures), thereby ignoring irrelevant background information.

- **Use Pre-trained Model and Fine-tuning:**We used a pre-trained ResNet-18 model as backbone, and the layers extracted from the pre-trained model were used to extract features from the input images. The code segments the layers of ResNet into different blocks (from layer0 to layer4), each of which is enhanced for feature extraction through the custom module FeatureFuse.

4 Experimental Setup

4.1 Datasets

Due to the limited size of publicly available retinal blood vessel segmentation medical image datasets, we merged three distinct datasets, **DRIVE**, **STARE**, and **HRF**, to form our experimental dataset comprising a **total of 85** pairs of images and masks.

4.2 Evaluation Metrics

Intersection over Union(IoU): In semantic segmentation, IoU involves calculating the overlap between the predicted segmentation area and the true segmentation area.
Dice coefficient: The Dice coefficient is calculated by doubling the overlap area between the predicted segmentation and the ground truth and dividing by the total number of pixels in both segmentations.
Accuracy: Accuracy refers to the proportion of pixels in a given image that are correctly classified in the context of semantic segmentation.

4.3 Other Experiment Setup

We implemented an aggregation U-Net with the python library PyTorch and conducted experiments on Google Colab. We divided the **85** original datasets to **80%** for training, **10%** for validation, and **10%** for Testing. While training, We use AdamW as optimizer, BCEWithLogitsLoss as loss function. The expected training epochs is maximum of **100** epochs, and we added a mechanism that the training may stop earlier if there is no improvement on validation.

5 Results

5.1 Key Experimental Results

Data Augmentation Method	DICE	IoU	ACC
Rotate	**0.7605**	**0.6142**	**0.9513**
Add Noise	0.7487	0.5991	0.9499
Gaussian Blur	0.7193	0.563	0.9411

Table 1: Data Augmentation Performance

Model	DICE	IoU	ACC
UNet	0.7437	0.5949	0.9511
DeepLabv3	0.4424	0.2854	0.7848
Attention UNet	0.7543	0.6081	0.9539
Aggregation UNet	**0.7784**	**0.6383**	**0.9619**
Aggregation UNet with Attention	0.7720	0.6300	0.9571

Table 2: Model Performance

5.2 Main Findings and Analysis

In the data augmentation comparison experiments, rotation performs the best and hence can be used as a further experimental setup. In the comparison of our improved model with baseline we find that Aggregation UNet finally performs the best and improves the DICE by 3.47% from 74.37 to 77.84 compared to the UNet model, and the misclassification of noise is also significantly reduced. Aggregation UNet shows substantial improvements in segmentation accuracy, especially in terms of IoU and noise immunity. These improvements are crucial for applications in medical diagnostics, especially noise immunity.

5.3 Evaluation Against Baseline Models

In comparison with the baseline model, the improved Aggregation UNet has a 3.47% dice improvement compared to UNet, which shows that the introduced Feature Fuse module effectively strengthens the feature extraction of the encoder, especially for blood vessels, and achieves noise resistance for the original image. The improvement is greater than that of DeepLabv3, indicating that complex models perform worse on tasks with very limited data sets. Meanwhile, the skip connection structure of UNet is crucial for blood vessel segmentation, a task that requires a large amount of spatial information. And compared to Attention UNet, the improved network also achieved a 2.41% dice improvement.

5.4 Real-World Model Deployment and Applications

Although our model can initially segment blood vessels, its accuracy needs further improvement; it remains inadequate for detailed medical diagnosis tasks, such as quantifying vessel curvature and diameter.
For deployment, the model can be exported using Libtorch to a C++ compatible, which provides higher performance than Python. This facilitates faster real-time inference. Additionally, employing deep supervision and network pruning significantly reduces network complexity, enhancing speed and suitability for real-time medical applications.

6 Conclusions

6.1 Overview

Due to the limited size of the original dataset, we employed data augmentation techniques to increase the dataset, which helps improve the model's generalization ability. We chose U-Net as the baseline model and compared its performance with five different models (such as DeepLab) to determine which model is more suitable for the task.
We made the following modifications to U-Net: preserved skip connections, modified residual block, feature fusion, implemented residual-attention blocks. With these improvements, our model achieved a significant enhancement in IoU metric, with an increase of 4.34%.

6.2 Future Work

Limitation of Datasets: This issue may still affect the generalization ability of the model. Due to the very limited public medical image datasets, we can further introduce unsupervised learning, such as the GAN adversarial generation mechanism[3], to further improve the segmentation performance of the model.
Model pruning: By introducing the deep supervision mechanism, the model is pruned to further reduce the number of model parameters, thereby improving the reasoning speed and achieving real-time segmentation[7].

References

[1] Liang-Chieh Chen et al. "Rethinking Atrous Convolution for Semantic Image Segmentation". In: *CoRR* abs/1706.05587 (2017). arXiv: 1706.05587. URL: http://arxiv.org/abs/1706.05587.

[2] Liang-Chieh Chen et al. "Rethinking atrous convolution for semantic image segmentation". In: *arXiv preprint arXiv:1706.05587* (2017).

[3] Shijie Cheng et al. "SUGAN: A Stable U-Net Based Generative Adversarial Network". In: *Sensors* 23.17 (2023). ISSN: 1424-8220. DOI: 10.3390/s23177338. URL: https://www.mdpi.com/1424-8220/23/17/7338.

[4] Anita Khanna et al. "A deep Residual U-Net convolutional neural network for automated lung segmentation in computed tomography images". In: *Biocybernetics and Biomedical Engineering* 40.3 (2020), pp. 1314–1327. ISSN: 0208-5216. DOI: https://doi.org/10.1016/j.bbe.2020.07.007. URL: https://www.sciencedirect.com/science/article/pii/S0208521620300887.

[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* (2015), pp. 3431–3440.

[6] Ozan Oktay et al. "Attention u-net: Learning where to look for the pancreas". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2018, pp. 605–613.

[7] Sunil Vadera and Salem Ameen. *Methods for Pruning Deep Neural Networks*. 2021. arXiv: 2011.00241 [cs.LG]. URL: https://arxiv.org/abs/2011.00241.