Shinakth - Envisioning Reliable Diagnosis of Diabetic Retinopathy using U-Net Architecture

Peerzada Anzar Azmat[1], Avni Garg[2] and Malik Najeeb Ul Habib[3]

[1]Department of CSE, New Horizon College of Engineering
[2]Department of CSE, New Horizon College of Engineering
[3]Department of CSE, New Horizon College of Engineering

Email ids: [1]peerzada.anzar@ieee.org, [2]avnigarg@ieee.org, [3]maliknajeeb@ieee.org

***Abstract*—Deep neural networks outperform traditional methods for semantic and instance segmentation of biomedical data. This method, however, is computationally expensive. The computational cost can be reduced by simplifying the network after training or selecting the appropriate architecture, which provides segmentation with less accuracy but much faster. We investigated the accuracy and performance of various UNet architectures for the problem of image segmentation in this study. The DRIVE dataset was analysed using images from a diabetic retinopathy screening programme. Retinal vessel segmentation and delineation of morphological attributes of retinal blood vessels, such as length, width, tortuosity, branching patterns, and angles, are used in the diagnosis, screening, treatment, and evaluation of a variety of cardiovascular and ophthalmologic diseases, including diabetes, hypertension, arteriosclerosis, and chorodial neovascularization. Among the six different models of UNet being compared, R2-UNet delivered the most promising results in terms of accuracy.**

***Index Terms*—biomedical segmentation, semantic segmentation, neural network performance, UNet, ResUNet , Attention U-Net, RA-UNet, BT-Unet, R2U-Net, retinopathy**

I. Introduction

Deep learning is a powerful technology that can be used to help clinicians better understand problems in their own environment. Deep learning is changing how we see the world. It has made the most significant contribution to biomedical image segmentation by automating the process of delineation in medical imaging. To complete such a task, the models must be trained on a massive amount of annotated data that highlights the region of interest with a binary mask. However, efficient annotation generation for such massive data sets necessitates the use of expert biomedical analysts and extensive manual labour. The U-Net architecture, which was first published in 2015, was a game changer in the field of deep learning.With the U-Net architecture, image segmentation of sizes 512X512 can be computed in a short amount of time using a modern GPU. Because of its phenomenal success, this architecture has experienced numerous variants and modifications.

In present work, we compare the accuracy and performance of six neural networks (UNet [1], ResUNet [2], Attention U-Net [3], RA-UNet [4], BT-Unet [5], and R2U-Net [6]) for image segmentation.

For analysis, we chose DRIVE: Digital Retinal Images for Vessel Extraction which was established to allow comparative studies on blood vessel segmentation in retinal images. The DRIVE database photographs were obtained from a diabetic retinopathy screening programme in the Netherlands. The screening population included 400 diabetic subjects aged 25 to 90. The primary reason for this selection is that various cardiovascular and ophthalmologic diseases like chorodial neovascularization, diabetes, arteriosclerosis, and hypertension are diagnosed, screened for, treated for, and evaluated using retinal vessel segmentation and delineation of morphological characteristics of retinal blood vessels like tortuosity, length, width, branching patterns, and angles. [7]. Let us describe the dataset briefly.

Blood vessel morphology in retinal fundus images is an important indicator of diseases such as glaucoma, hypertension, and diabetic retinopathy. The accuracy of retinal blood vessel segmentation affects the quality of retinal image analysis, which is used in modern ophthalmology diagnosis methods. In addition to research into the connection between hypertensive retinopathy, vessel tortuosity, and vessel diameter measurement in relation to hypertension diagnosis, and computer-assisted laser surgery, automatic detection and analysis of the vasculature can aid in the implementation of diabetic retinopathy screening programmes. Automatic retinal map generation and branch point extraction have been employed for temporal or multimodal image registration as well as retinal image mosaic synthesis.

II. Methods

A. U-Net

The idea behind U-net was that if we fed the image to an encoder that keeps decreasing the spatial size of the feature block, the network would generalise to store only the important features and discard less useful data after enough training. Finally, the encoder output, followed by a decoder, will generate the desired output mask. The issue was that the decoder layers were not receiving enough context from the encoder output to generate the segmentation mask.

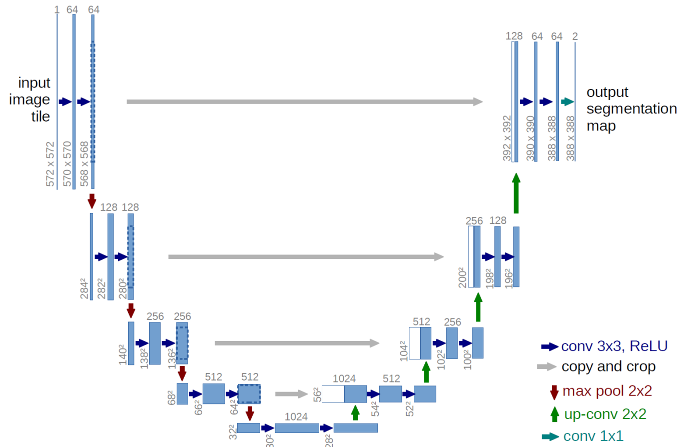

Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). A multi-channel feature map is represented by each blue box. The number of channels is indicated on the box's top. The x-y size is provided at the box's lower left edge. White boxes represent feature maps that have been copied. The arrows represent the various operations.

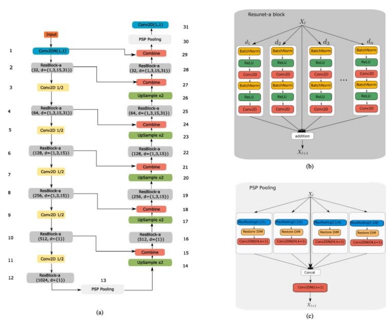

Fig. 2. A diagram of the ResUNet-a d6 network. (a) The architecture's encoder is located on the left (downward) branch. The decoder is located on the right (upward) branch. The final convolutional layer contains as many channels as distinct classes. (b) A ResUNet-a network building block. Every unit in the residual block has the same number of filters as the others. Here different dilation rates are denoted by d d,, n 1. (c) Pooling layer for pyramid scene parsing. Pooling occurs in 1/1, 1/ 2, 1/4, and 1/8 of the original image.

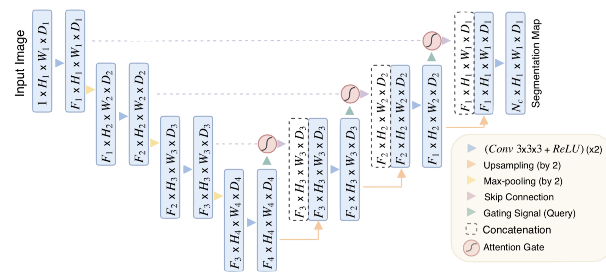

Fig. 3. The Attention U-Net segmentation model is depicted as a block diagram. In the network's encoding section, the input image is progressively filtered and downsampled by a factor of two at each scale. The number of classes is denoted by N_c. The features propagated through the skip connections are filtered by attention gates (AGs). In AGs, feature selectivity is achieved through the use of contextual information (gating) extracted at coarser scales.

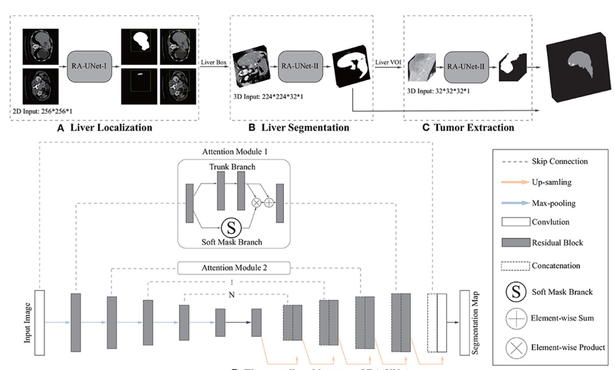

Fig. 4. An overview of the proposed liver and tumour segmentation pipeline. (a) For coarse localization of a liver region within a boundary box, a simple version of 2D RA-UNet (RA-UNet-I) is used. (b) The 3D RA-UNet (RA-UNet-II) is intended for extracting attention-aware features of liver VOIs within the liver boundary box in a hierarchical manner. (c) RA-UNet-II is in charge of accurate tumour extraction within the liver VOIs. (d) RA-overall UNet's architecture.

B. ResUNet

ResUnet is an intriguing concept that combines the performance gain of Residual networks with the U-Net. It is a powerful network with an unusually large number of parameters.

C. Attention U-Net

This network took the idea of an attention mechanism from NLP and applied it to skip connections. It gave the skip connections an additional idea of which region to focus on when segmenting the given object. Because of the attention in the skip connections, this works great even with very small objects. This one is a little more difficult to implement from scratch, but the idea is quite clever and simple.

D. RA-UNet

The RA-UNet network capitalises on the U- Net's, residual learning's, and attention residual mechanism's strengths. To begin, the use of attention modules causes adaptive changes in attention-aware features. Second, residual blocks are stacked into the architecture, allowing it to go deep and solve the gradient vanishing problem. Finally, the U-Net is used to collect multi-scale attention data and integrate low-level and high-level features.

E. BT-Unet

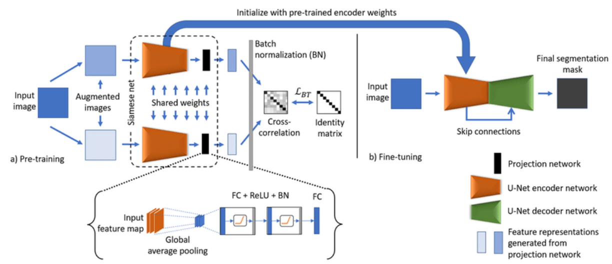

Fig. 5. The BT-Unet framework a) A pre-trained U-Net encoder network, and b) A fine-tuning U-Net model with pre-trained encoder weights.

The BT-Unet framework employs a redundancy reduction-based Barlow Twins strategy for unsupervised pre-training of the U-Net model's encoder network with feature representations of the data, followed by fine-tuning of the U-Net model for downstream biomedical image segmentation tasks with limited annotated data samples.

F. R2U-Net

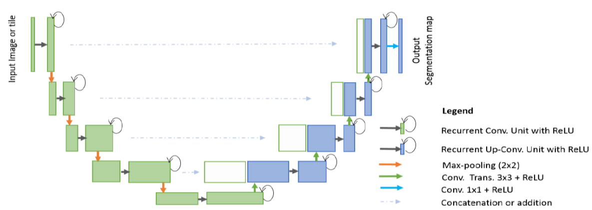

Fig. 6. RU-Net architecture with convolutional encoding and decoding units based on recurrent convolutional layers (RCL). RCL and residual units are used in the R2U-Net architecture.

Recurrent Convolutional Neural Networks and Recurrent Residual Convolutional Neural Networks were proposed as extensions to the U-Net architecture in this architecture.

III. Experimental Results

A. Model Training

As shown in Fig. 7, with respect to the time taken to train the models, it is observed that the BT-Unet network took the least time whereas training RA-Unet was most time consuming.

B. Comparison Metrics

From Fig. 8, Fig. 9 and Table 1 it can be observed that the accuracy of R2-UNet among all other neural networks compared and studied under this study was better for the data set used.

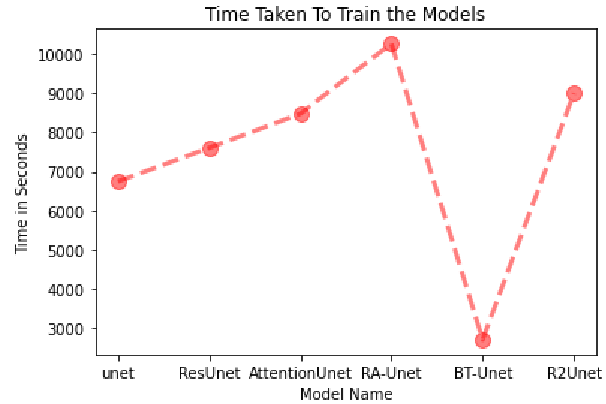

Fig. 7. Time taken to train the models.

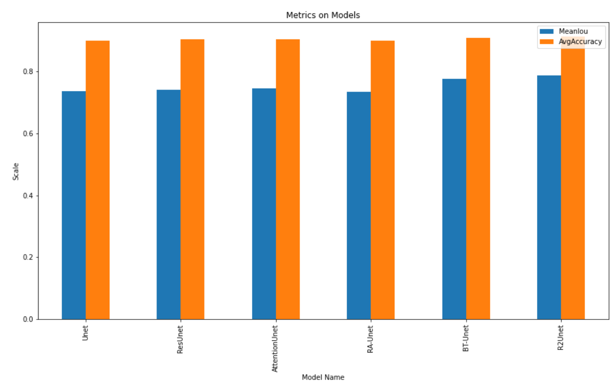

Fig. 8. Time taken to train the models.

TABLE I
COMPARISON OF U-NET ARCHITECTURES ON PERFORMANCE

UNet Architecture	**Performance**		
	Time Taken	***Mean IoU***	***Avg. Accuracy***
UNet	45x150 seconds 150 epochs	0.736	0.9
RES-UNet	50.7x150 seconds 150 epochs	0.741	0.903
Attention UNet	56.5x150 seconds 150 epochs	0.745	0.905
RA-UNet	68.4x150 seconds 150 epochs	0.735	0.9
BT-UNet	210+5x100x5 seconds 100 epochs	0.775	0.909
R2-UNet	50x3x60 seconds 100 epochs	0.787	0.9133

IV. Future Work

In this work we have evaluated various U-Net Architectures, and we observed that such architectures are promising and effective for the Retinopathy Image Segmentation. Hence we would like to propose a novel architecture based on U-Net to achieve the same in our future work.

V. Conclusion

In this study, we compared the precision and performance of UNet, ResUNet, Attention U-Net, RA-UNet, BT-Unet, and R2U-Net. The compariosn was performed on the data obtained from a diabetic retinopathy screening program. The dataset was obtained from a population of 400 diabetic subjects between 25-90 years of age where some images showed no

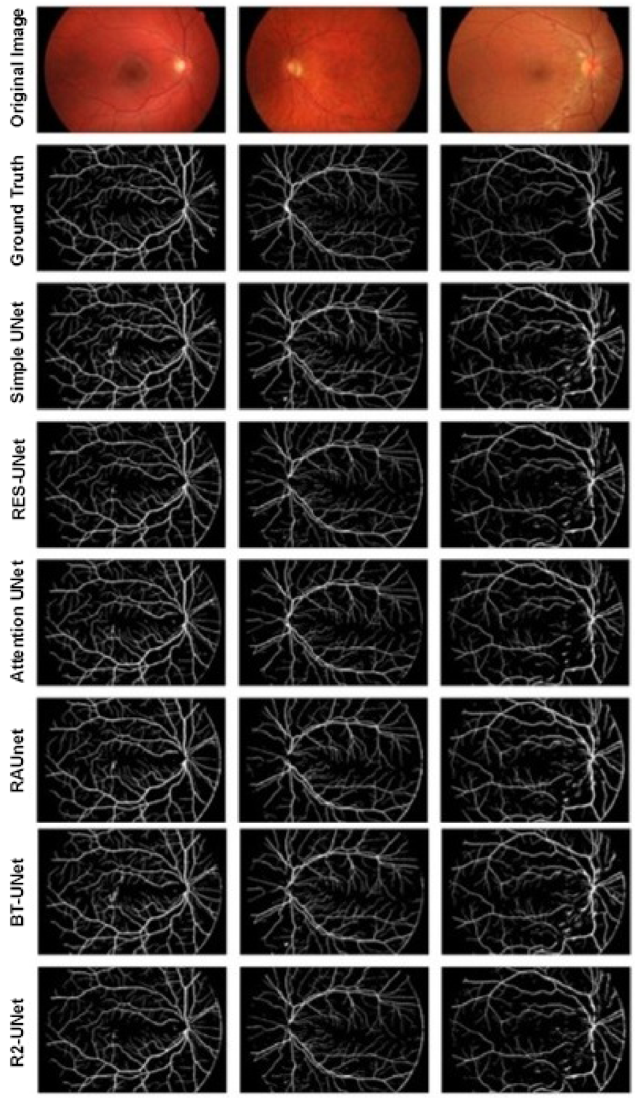

Fig. 9. Original image, ground truth, and semantic segmentation performed with six neural networks.

sign of diabetic retinopathy while few showed signs of mild early diabetic retinopathy. It was observed that R2-UNet model showed the most promising results.

ACKNOWLEDGEMENT

We thank Ms. Shanmugapriya for his guidance and support we received during the course of this project.

REFERENCES

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation

[2] Rui Li, Shunyi Zheng, Chenxi Duan, Jianlin Su, and Ce Zhang: Multi-stage Attention ResU-Net for Semantic Segmentation of Fine-Resolution Remote Sensing Images

[3] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla,Bernhard Kainz, Ben Glocker, and Daniel Rueckert: Attention U-Net: Learning Where to Look for the Pancreas

[4] Qiangguo Jin, Zhaopeng Meng, Changming Sun, Leyi Wei, and Ran Su: RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans

[5] Narinder Singh Punn and Sonali Agarwal: BT-Unet: A self-supervised learning framework for biomedical image segmentation using Barlow Twins with U-Net models

[6] Md Zahangir Alom, Student Member, IEEE, Mahmudul Hasan, Chris Yakopcic, Member, IEEE,Tarek M. Taha, Member, IEEE, and Vijayan K. Asari, Senior Member, IEEE: Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation

[7] DRIVE: Digital Retinal Images for Vessel Extraction (https://drive.grand-challenge.org/)