Grenze International Journal of Engineering and Technology, June Issue

Retinal Microvessel Segmentation using Variants of U-Net and Domain Adaptation

Swastik Sharma[1] and Sannidhya Rawat[2]
[1-2]Delhi Technological University
Email: swastiksharma_2k19ec192@dtu.ac.in, sannidhyarawat_2k19ec172@dtu.ac.in

***Abstract*--The use of deep learning in medical image segmentation has gained popularity, and one commonly utilised method is U-Net. U-Net is favoured for its ability to accurately identify small objects within images and its adaptable network structure. U-Net has received significant attention in academic literature, with over 2500 citations. Researchers have been continuously enhancing its architecture to improve its performance. This paper presents a proposal that focuses on various U-Net variations for medical image segmentation. It explores their structure, innovative features, and overall efficiency. Our proposal uses multiple variants of U-Net and rotation-based domain adaptation. The paper also reviews the application of domain adaptation on one dataset to get better results, which will be helpful for future research.**

***Keywords*: medical image segmentation; U-NET; Domain adaptation;**

I. Introduction

Precise interpretation of medical images such as CT and MRI scans necessitates extensive expertise and training, as organs and lesions must be segmented meticulously, layer by layer. Manual segmentation can be burdensome for doctors, leading to potential bias when subjective opinions come into play. Analysing complex images often requires the collaboration of multiple doctors for a collective diagnosis, which can be very time-consuming. Moreover, automatic image segmentation poses a significant challenge that remains unresolved for numerous medical utilization due to the wide array of problems such as image processing, compression techniques, and accounting for natural variations in real-world images.
Obstacles encountered in medical image segmentation First, Accurate disease diagnosis is crucial in medical image processing. This involves segmenting images at the pixel or voxel-level by Mohammad et al. [1]. It is often challenging to distinguish boundaries between multiple cells and organs in medical images. Additionally, image data is usually pre-processed before being used to build relevant networks, which require continuous parameter adjustments even after achieving a certain level of accuracy with deep learning models. Secondly, Ways to optimise the utilisation of limited annotation resources and evaluate the accuracy of annotations. Thirdly, Privacy-preserving in the medical domain. Preserving the confidentiality of patient information during the sharing of medical data is a widespread concern.

Grenze ID: 01.GIJET.9.2.419_1

II. Related Work

Deep learning technology has gained significant attention in medical imaging, particularly in detecting and segmenting lesions in medical images. One example is based on domain adaptation using generative adversarial networks (GANs) presented by Ronneberger et al. [2]. The model is trained to generate samples indistinguishable from the target domain while performing well on the source domain. Another example is based on multi-task learning, and the model is trained by considering both the source and target domains concurrently., leveraging the shared information between the domains. Despite the progress made in this field, domain adaptation remains an active area of research with many challenges and opportunities for further exploration.

Numerous algorithms and models have been devised for segmentation of blood vessels in retinal fundoscpic images, as documented in prior studies [3], [4]. The authors concluded that segmenting blood vessels in fundal images involves analysing the pixel intensity and responses from two-dimensional Gabor wavelet transforms at various scales. This approach was effective in extracting blood vessels from the fundoscopic images. In the study by Melinak et al. [5], the researchers used a deep learning framework to segment blood vessels in fundoscpic images. Hoover et. al. [6] created an algorithm that involved utilising both local and global features of blood vessels to segregate the vessel network from the fundoscpoic images. However, it is important to note that this approach relied on certain assumptions. For instance, a Gaussian function was chosen to represent the shape of a blood vessel profile. Additionally, fixed widths and directions were also assumed. When using image processing techniques for segmentation, it is necessary to make specific assumptions which may not necessarily reflect the actual characteristics of the image. To achieve accurate segmentation, it is important to avoid making assumptions and using operations that could alter the original image structure and intensities.

III. Proposed Methodology

To achieve image segmentation in fundoscopic images, we use domain adaptation on four architectures/models based on the U-NET architecture.

A. Domain Adaptation

Domain adaptation improves a model's performance on a target domain by leveraging knowledge from a related source domain. In computer vision, a domain consists of images with similar visual properties. The goal is to tackle the issue of domain shift, which arises due to variations in data distributions between the source and target domains. This disparity can lead to insufficient performance of the models.

Several approaches to domain adaptation include instance reweighting, feature adaptation, and adversarial training. Instance reweighting involves reweighting the source domain instances to reduce their influence on the model's decision-making process.

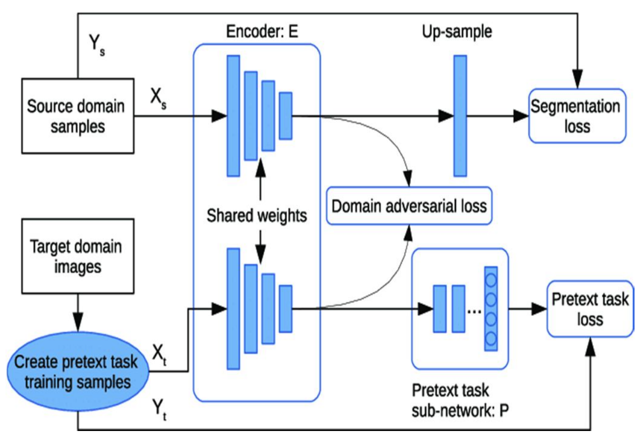

Fig 1: An example of domain adaptation [7].

Feature adaptation aims to acquire a feature representation that remains consistent across different domains by aligning the feature distributions of the source and target domains. Adversarial training involves training a model with an additional adversarial loss that encourages the network to learn discriminative features for the classification task and domain-invariant.

B. Traditional U-NET

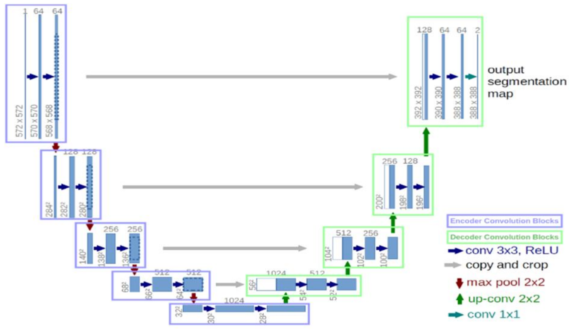

Fig 2: Traditional U-Net with encoder and decoder blocks shown [2]

U-Net is a network architecture that improves upon the Fully Convolutional Network (FCN) in a 2D setting to segment limited training images accurately (Ronneberger et al. [2]). The network employs sequential layers instead of a general shrinkage network and replaces the downsampling operator with a pooling operation and an upsampling operator to enhance output resolution. Unlike traditional CNNs that map feature maps to a fixed-length feature vector, U-Net identifies images at the pixel level, making segmentation at the semantic level easier. It is ideal for medical image segmentation as it can handle large image sizes by cutting them into smaller pieces, which can be fed to the network through overlapping tiling strategies.

C. Residual U-NET

This version of U-net is built on the ResNet architecture as its foundation. The main motivation behind developing ResNet was to overcome the challenges associated with training deep neural networks. According to Zhang et. al. [8] they have observed that neural networks tend to reach a solution faster when they have a greater number of layers. However, increasing the depth of the network beyond a certain point can lead to a decline in performance, as gradients in the weight vector become less effective. To address this issue, ResNet introduced skip connections, which allow the feature maps from one layer to be added to another layer located deeper in the network. This predictive behavior of skip connections helps preserve feature maps more efficiently in deeper networks, ultimately leading to improved performance.

D. Attention U-NET

In image processing networks, it is often advantageous to have the ability to prioritise important objects while disregarding irrelevant areas. The AttU-Net or Attention U-net accomplishes this by incorporating an attention gate. In this version of U-Net we use soft attention using an attention gate which is a component that filters out unnecessary features that are not relevant by giving more weightage to the important features. This helps to focus on the important information for the given task, as described by Oktay et al. [9]. In simpler terms, at every layer in the decoder path, there is a gate that pays attention to all the features from the encoding path. These features need to pass through the gate before they are combined with the features in the decoder path. Using the attention gate multiple times after each layer in the every convolution block greatly enhances the model's segmentation accuracy without excessively increasing the computational requirements.

E. Residual Attention U-NET (RA-UNET)

To improve the processing of edge information in images, RA-UNet employs a self-attention mechanism that is both depthwise and polarised to help a CNN in denoising the image. Extensive experiments have conclusively shown that RA-UNet surpasses current state-of-the-art denoising techniques in both efficiency and performance., Ni et al. [10]. Visualisation results indicate that RA-UNet produces smoother, sharper, denoised images than other methods.

F. Rotation-based Domain Adaptation

In every variant of U-Net, we add the rotation block containing three convolution layers in the decoder convolution blocks, followed by batch normalisation (BN) layers and then a fully connected (FC) layer at the end. The input to these is rotated images from the stare dataset.

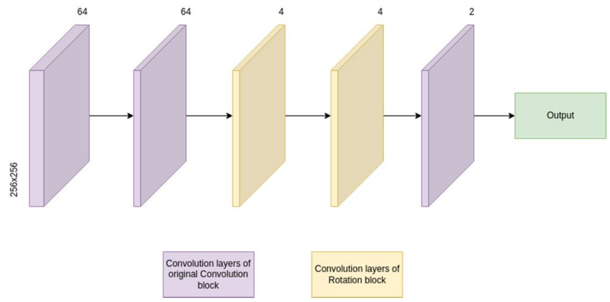

Fig 3: In every decoder block, we have added a rotation block so the model can learn rotation-invariant features.

The reasoning for choosing this method is that our dataset has fundoscopic images of the retina in the two different datasets that are visually very similar.

G. Our Approach

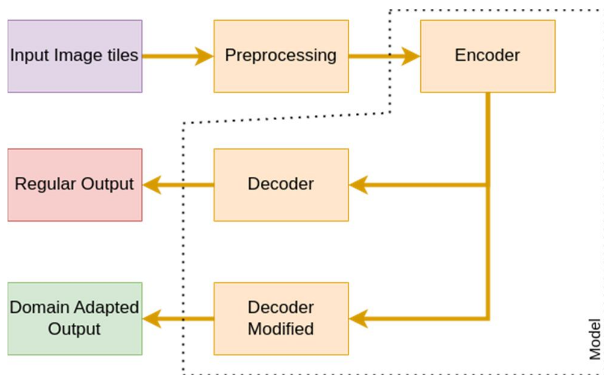

Fig 4: Pre processing of images followed by downsampling to have the results.

We chose the rotation prediction domain adaptation technique published in Ajinkya Tejankar et al. [12], which comes under the feature adaptation category. We start with simple U-Net architecture and add a rotation block after every convolution block. When the model is trained on Drive, the rotation block learns rotation invariant features from the test and vice versa.
Then we follow the same procedure for Attention U-Net with the rotation blocks added after the Attention block and for Residual U-Net where we add the rotation model after the Residual block. In Residual-Attention U-Net, we add the rotation block at the end. Now as the model learns the features from the DRIVE dataset, it also learns the rotation invariant features from the STARE dataset.

IV. EXPERIMENTATION

Retinal vessel segmentation is a method that is utilised to identify and locate blood vessels in fundoscopic images taken of the back of the eye. Our project is the Pytorch implementation of a model using neural networks.

A. Data and its Preprocessing

We perform tests on two datasets of images called DRIVE and STARE datasets. The DRIVE dataset comprises 40 images, whereas the STARE dataset includes 20 images. These datasets were specifically created for researchers to compare different techniques of segmenting blood vessels in retinal fundoscopic images. The DRIVE dataset is a collection of 40 images taken from individuals between the ages of 25 and 90 years. These images were obtained through a diabetic retinopathy screening program carried out in the Netherlands. The STARE (Structured Analysis of the Retina) database was established in 1975 through a collaborative project involving over thirty individuals from various medical, scientific, and engineering backgrounds.
Both datasets consist of manual segmentations of blood vessels performed by two expert annotators. For training and testing our algorithm, we use the segmentations generated by the first annotator as the reference or gold

standard. The segmentations created by the second annotator are then compared to the gold standard to evaluate human performance. The data undergoes preprocessing through two techniques: gamma adjustment and CLAHE (Contrast Limited Adaptive Histogram Equalization).
Figures from 5 to 20 show various results when data is trained on one dataset and tested on another dataset, showing the original image, the predicted image and the ground truth.

B. Drive On Drive

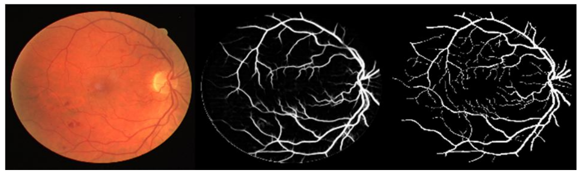

Fig 5: Simple U-NET result, trained and tested with DRIVE dataset

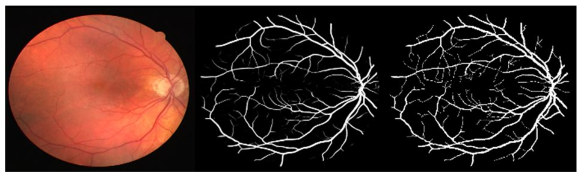

Fig 6: Residual U-NET result, trained and tested with DRIVE dataset

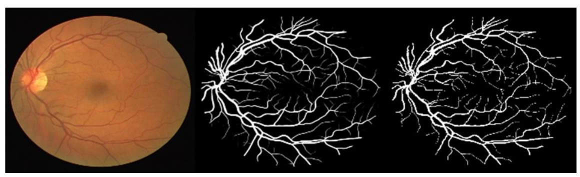

Fig 7: Attention U-NET result, trained and tested with DRIVE dataset

Fig 8: Residual-Attention U-NET result, trained and tested with DRIVE dataset

TABLE I: COMPARISON OF DIFFERENT MODELS

Model Names	Precision	Accuracy
Simple U-NET	0.8643	0.8832
Residual U-NET	0.8825	0.9183
Attention U-NET	0.8977	0.9243
Residual Attention U-NET	0.9103	0.9561

C. Stare on Stare

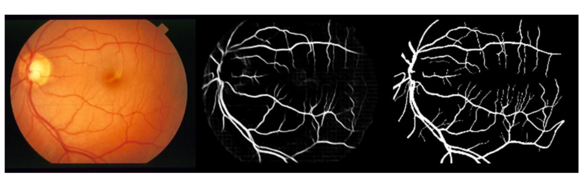

Fig 9: Result of Simple U-NET,when trained and tested with STARE dataset

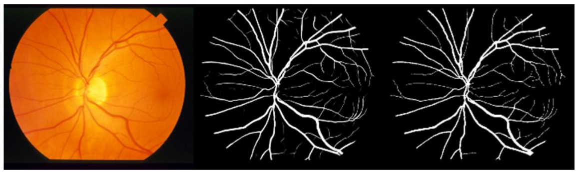

Fig 10: Result of Residual U-NET , when trained and tested with STARE dataset

Fig 11: Result of Attention U-NET, when trained and tested with STARE dataset

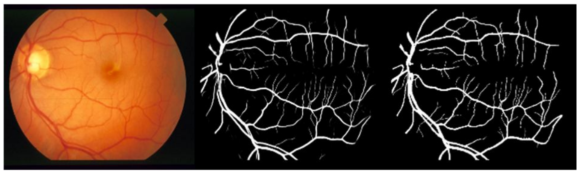

Fig 12: Result of Residual-Attention U-NET, when trained and tested with STARE dataset

TABLE II: COMPARISON OF DIFFERENT MODELS

Model Names	Precision	Accuracy
Traditional U-NET	0.8727	0.9044
Residual U-NET	0.8932	0.9109
Attention U-NET	0.9116	0.9422
Residual Attention U-NET	0.9244	0.9625

D. DRIVE on STARE before Domain Adaptation

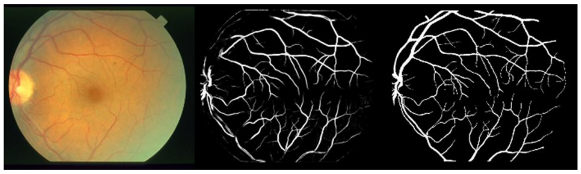

Fig 13: Attention U-NET result before domain adaptation, trained with DRIVE and tested with STARE

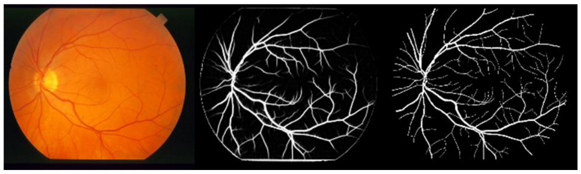

Fig 14: Residual U-NET result before domain adaptation, trained with DRIVE and tested with STARE

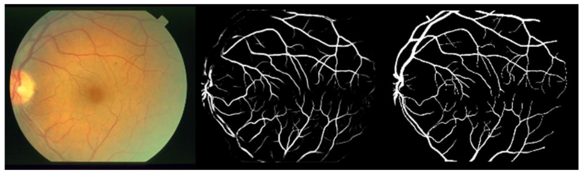

Fig 15: Residual-Attention U-NET result before domain adaptation, trained with DRIVE and tested with STARE

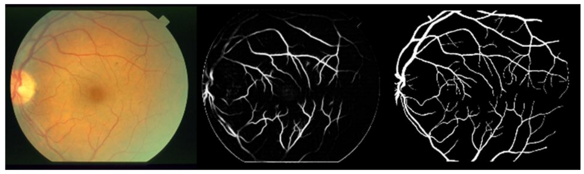

Fig 16: Simple U-NET result before domain adaptation, trained with DRIVE and tested with STARE

TABLE III: COMPARISON OF DIFFERENT MODELS

Model Names	Precision	Accuracy
Traditional U-NET	0.7498	0.7621
Residual U-NET	0.7621	0.7905
Attention U-NET	0.8102	0.8433
Residual Attention U-NET	0.8323	0.8632

E. Drive on Stare after Domain Adaptation

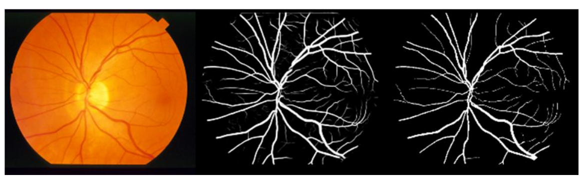

Fig 17: Attention U-NET result after domain adaptation, trained with DRIVE and tested with STARE

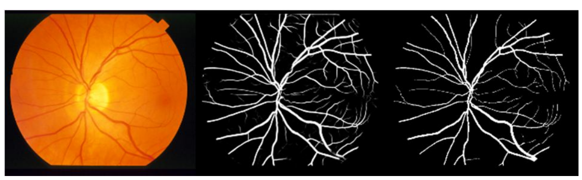

Fig 18: Residual U-NET result after domain adaptation, trained with DRIVE and tested with STARE

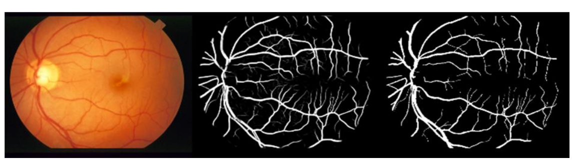

Fig 19: Residual-Attention U-NET result after domain adaptation, trained with DRIVE and tested with STARE

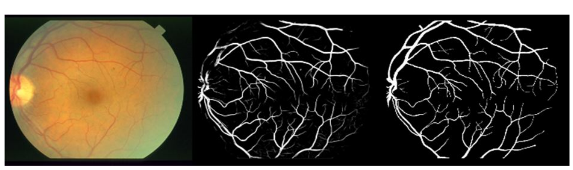

Fig 20: Simple U-NET result after domain adaptation, trained on DRIVE dataset and tested on STARE dataset

TABLE IV: COMPARISON OF DIFFERENT MODELS

Model Names	Precision	Accuracy
Simple U-NET	0.8569	0.8532
Residual U-NET	0.8673	0.8764
Attention U-NET	0.8833	0.9121
Residual Attention U-NET	0.9021	0.9467

VI. CONCLUSION

We tested different UNET models (Simple, Residual, Attention, and Residual Attention) on two retina vessel segmentation datasets (DRIVE and STARE). Based on our findings, we accuracies are in the following order:

Residual Attention Unet > Attention Unet > Residual Unet > Simple Unet.

We observed that the UNET models performed well on the datasets they were trained on, but their accuracy significantly decreased when tested on unseen datasets. To address this issue, we employed domain adaptation.

REFERENCES

[1] Hesamian, Mohammad Hesam, et al. "Deep learning techniques for medical image segmentation: achievements and challenges." Journal of digital imaging 32 (2019): 582-596.

[2] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[3] Vengalil, Sunil Kumar, et al. "Customizing CNNs for blood vessel segmentation from fundus images." 2016 international conference on signal processing and communications (SPCOM). IEEE, 2016.

[4] Ricci, Elisa, and Renzo Perfetti. "Retinal blood vessel segmentation using line operators and support vector classification." IEEE transactions on medical imaging 26.10 (2007): 1357-1365.

[5] Melinscak, Martina, Pavle Prentasic, and Sven Loncaric. "Retinal Vessel Segmentation using Deep Neural Networks." VISAPP (1). 2015.

[6] Hoover, A. D., Valentina Kouznetsova, and Michael Goldbaum. "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response." IEEE Transactions on Medical imaging 19.3 (2000): 203-210.

[7] Xu, Jiaolong, Liang Xiao, and Antonio M. López. "Self-supervised domain adaptation for computer vision tasks." IEEE Access 7 (2019): 156694-156706.

[8] Zhang, Zhengxin, Qingjie Liu, and Yunhong Wang. "Road extraction by deep residual u-net." IEEE Geoscience and Remote Sensing Letters 15.5 (2018): 749-753.

[9] Oktay, Ozan, et al. "Attention u-net: Learning where to look for the pancreas." arXiv preprint arXiv:1804.03999 (2018).

[10] Ni, Zhen-Liang, et al. "Raunet: Residual attention u-net for semantic segmentation of cataract surgical instruments." Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part II. Cham: Springer International Publishing, 2019.

[11] Tejankar, Ajinkya, and Hamed Pirsiavash. "A simple baseline for domain adaptation using rotation prediction." arXiv preprint arXiv:1912.11903 (2019).