frontiers

Simultaneous Segmentation of Multiple Structures in Fundal Images using Multi-tasking Deep Neural Networks

Sunil Kumar Vengalil *, **Bharath K, and Neelam Sinha**

International Institute of Information Technology Bangalore, Bangalore, India

Correspondence*:
Sunil Kumar Vengalil
sunilkumar.vengalil@iiitb.ac.in

ABSTRACT

Fundal imaging is the most commonly used non-invasive technique for early detection of many retinal diseases like Diabetic Retinopathy(DR). An initial step in automatic processing of fundal images for detecting diseases is to identify and segment the normal landmarks: Optic Disc, Blood Vessels and Macula. In addition to these, various abnormalities like Exudates that help in pathological analysis are also visible in fundal images. Segmenting structures like Blood Vessel, pose multiple challenges, such as being very fine grained structures that need to be captured at original resolution. Further, they are spread across the entire retina with varying patterns and density. Similarly, Exudates appear as white patches of irregular shape that occur at multiple locations. They can be confused with Optic Disc, if features like brightness or color are used for segmentation. Segmentation algorithms solely based on image processing involve multiple parameters and thresholds that need to be tuned. Another approach is to use machine learning models with hand-crafted features as input to segment the image. The challenge in this approach is to identify the right features and then devising algorithms to extract these features. End-to-end deep neural networks take raw images (with minimal preprocessing such as resizing and normalization) as input and then learn a set of in the intermediate layers and perform the segmentation in the last layer. These networks tend to have longer training and prediction time because of complex architectures that involve millions of parameters which also necessitates huge numbers of training images (2000-10000). For structures like Blood Vessels and Exudates that are spread across the entire image, one approach used to increase the training data is to generate multiple patches from a single training image, thus increasing the total number of training samples. Patch-based time can not be applied to structures like Optic Disc and Fovea that appear only once per image. Also the prediction time is larger as segmenting a full image involves segmenting multiple patches in the image. Most of the existing research focuses on segmenting these structures independently aiming to achieve high performance metrics. In this work, we propose a multi-tasking deep learning architecture for segmenting Optic Disc, Blood Vessels, Macula and Exudates simultaneously. Both training and prediction is performed using the whole image. The objective is to improve the prediction results on Blood Vessels and Exudates, which are relatively more challenging, while utilizing segmentation of Optic Disc and Macula as auxiliary tasks. Our experimental results on publicly available datasets show that simultaneous segmentation of all these structures results in significant improvement in the performance. The proposed approach makes predictions of all the four structures, for the whole image, in a single forward pass. We use modified U-Net architecture with only convolutional and de-convolutional layers with comparatively less number of parameters leading to faster (by a factor of 12) prediction

time, compared to the current AUC-wise best performing Exudate segmentation deep learning model, ERUnet. The proposed approach has been evaluated on publicly available datasets DRIVE, HRF, CHASE_DB and IDRiD datasets. Comparison of results with the state-of-the-art have also been presented. Among the segmented structures maximal improvement of 15% is obtained on Exudates when trained using multi-tasking loss function.

Keywords: Fundal Image Segmentation, Deep Learning, Multi-task learning, Exudates Segmentation, Blood Vessels, Macula, Optic Disc

1 INTRODUCTION

Fundal imaging, capturing images of retina using specialized cameras, is the most widely used non-invasive technique for screening of retinal diseases. These images are used to identify common eye diseases like Diabetic Retinopathy (DR) Hu et al. (2015) and Glaucoma which are the most common cause for blindness and could be indicators of many other cardiovascular diseases. Blood Vessels (BV), Optic Disc (OD) and Macula are the normal landmarks visible in a healthy fundal image. Certain features of BV like tortuosity are widely used for early detection of various cardio-vascular diseases Krestanova et al. (2020). However, manual identification and demarcation of fine structures like BV require domain expertise, besides being prone to manual errors. Features extracted for OD like, Cup-to-Disc ratio can be used for detection of Glaucoma. Hence automatic detection of major landmarks in fundal image has become an active research area Kou et al. (2020); Guo et al. (2020); Nur and Tjandrasa (2018); Jiang et al. (2018); Joshua and et al. (2020); Dash et al. (2020).

Figure 1 shows a fundal image with normal landmarks like BV, OD and Macula marked. The OD is the point of exit of the optic nerves that carry information from the eye to the brain. It is also the point where all the BV enter the eye. Since there are no photo sensors (rods and cones) present in the OD, it corresponds to a blind spot in the retina. Macula is a small region with a lot of cone cells packed together and hence this region is responsible for sharp vision Wikipedia (2022). The center point of Macula is called Fovea. BV that carry blood to the eye are spread across the entire region of the retina and vary in thickness and density.

Figures 2a and 2b show a sample fundal image with Exudates and its corresponding ground truth. Exudates are the fluid (pus) leaking out of Blood Vessels. They are indicative of an advanced stage of Diabetic Retinopathy. Exudates appear as unstructured, scattered, bright patches in the fundal image.

Segmentation of Blood Vessels is the most challenging task among the four structures considered as they are spread across the entire image with varying patterns and density. Further, they vary in thickness being thicker and denser near OD and become fine-grained towards the end of the branch. Down-sampling the image as required by some approaches like neural networks will result in loss of such fine-grained vessels in the segmented image. Exudates, which are visible as white patches of irregular shape in fundal image, can also be spread across the entire image. Exudates can further be confused with OD, especially those that appear very close to OD, if features like brightness, color and position are used for segmentation. On the other hand, Fovea and Optic Disc are present only once per fundal image at a fixed location. Further, their shapes are also relatively predictable compared to Exudates and Blood Vessels. Because of these characteristics, it is relatively easy to predict OD and Fovea.

One of the approaches to segment these structures is to use image processing algorithms such as thresholding, Edge and shape detection algorithms, morphological operations etc. One of the major drawbacks of such algorithms is the use of multiple parameters that need to be tuned for different types of

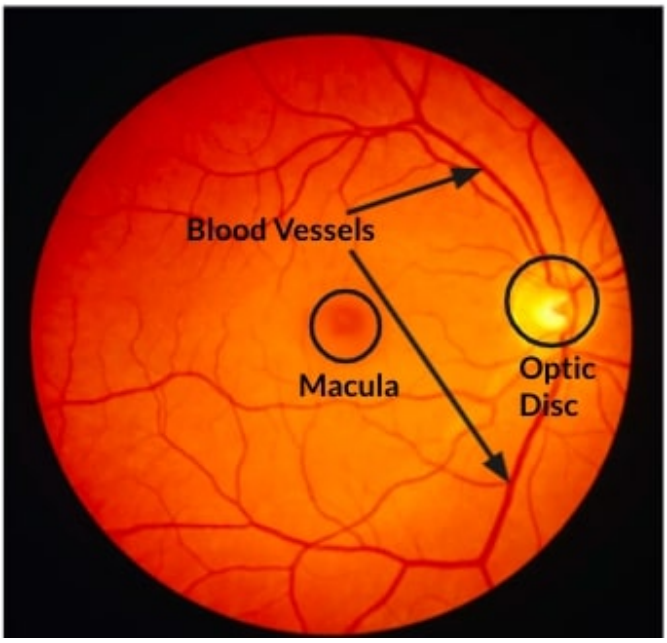

Figure 1. Sample fundal image showing normal landmarks BV, OD and Macula

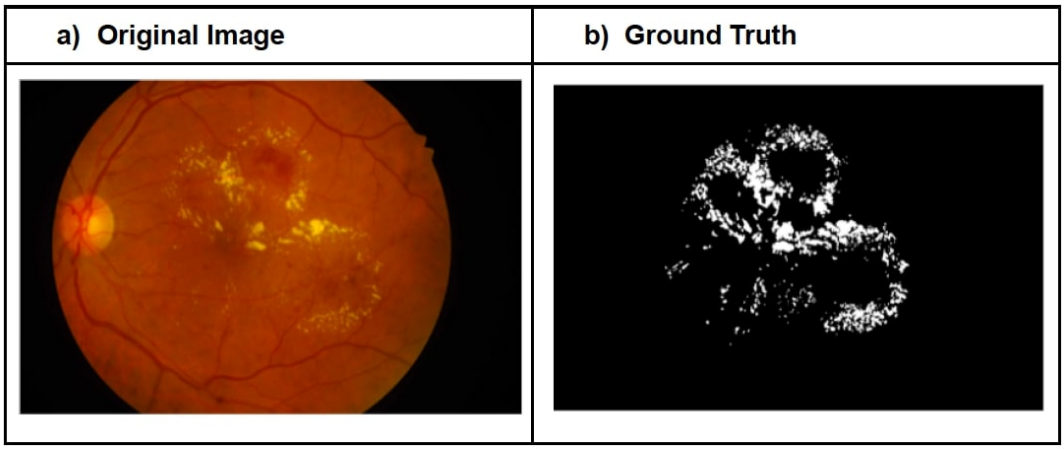

Figure 2. Sample fundal image with Exudates and corresponding ground truth image for Exudates

images with varying acquisition artifacts and pathologies. This can be resolved by a data-driven approach where one first identifies some generic features, such as edges, color, brightness, location etc, that are crucial for the segmentation task and then extract these features using image processing algorithms and finally using a machine learning model, such as clustering, Conditional Random Field etc. in order to segment the image using the features. The challenge in this approach is to identify the right features that can be easily extracted, which will work across tasks and across images (that vary in resolution, pathology and acquisition artifacts) and then devising algorithms to extract these features. End-to-end deep neural networks that take raw images (with minimal preprocessing such as resizing and normalization) as input and then learn a set of generic features in the initial layers, more task specific features towards the final layers and finally perform the segmentation task in the last layer, is becoming popular and shown to out-perform both the techniques mentioned earlier.

Since the break-through success of deep learning in solving tasks in domains like computer vision for classification Krizhevsky and et al. (2012) Simonyan and Zisserman (2014) the framework has been successfully extended to more complex tasks such as semantic segmentation Chen and et al. (2017); Yu et al. (2018); Wang et al. (2018). The primary reason for the success of these deep neural networks are that the features are learned from data in the initial layers of the network and the segmentation task is performed on these learned features in the final layer.

Recently, several deep learning architectures that were successful in segmentation tasks (Chen and et al. (2017)) in natural images were tried for segmenting Blood Vessels in retinal images. Many deep learning architectures have been utilized for segmenting challenging structures, such as BV Vengalil et al. (2016) Zhuang (2018) Jiang et al. (2018) Park and et al. (2020) and Exudates Kou et al. (2020); Guo et al. (2020); Nur and Tjandrasa (2018) in fundal images. The results were significantly better than using conventional image processing techniques. However, One of the main challenges in using deep neural networks for segmentation tasks is that the reduction in resolution of featuremap as one goes deeper will result in loss of finer details like edges, which are crucial for segmentation tasks. Another major issue in using deep learning architecture for medical images is limited availability of annotated training data. Deep learning models for segmentation need a large number of training images as the models have a huge number of parameters (typically in the range of 60-100 million Long et al. (2015). The number of images required to train a model increases with the number of training parameters in the model. Our model has fewer (31M) parameters compared to other deep learning models as we have less number of layers and also there is no dense layer (i.e the encoder network is fully convolutional and the decoder network is fully de-convolutional). Hence we are able to get good performance, without overfitting, even with relatively less number of training images. Many approaches, like taking multiple training patches from a single image Vengalil et al. (2016) and transfer learning Vengalil et al. (2016), have been successfully explored. These methods also tend to have longer prediction times because of complex neural network architectures and patch-based training and prediction.

Most of the existing research focuses on segmenting different structures independently. However, in almost all the practical use cases the end goal is to detect the early development of anomalies, like 1) Exudates that are indicators of pathologies like diabetic retinopathy, 2) abnormal cup-to-disc ratio in OD which is indicator of diseases like Glaucoma and 3) anomalies in tortuosity of BV which can be indicative of cardiovascular diseases. When an ophthalmologist analyses a fundal image, it is customary to look for all these anomalies. So for any automatic diagnostic system, that can reduce the manual intervention of an ophthalmologist, it is necessary that all these structures be segmented as a first step. Further, note that even if the use case is just to segment only one single anomaly like Exudates alone, the multi-tasking model does not add any additional overhead as opposed to a model that just predicts Exudates alone. Rather, the segmentation of Exudates has shown a significant improvement by using a multi-tasking model due to the correlation between other structures and Exudates. The remaining outputs can just be ignored. In general, segmenting multiple structures using separate models suffers from the following issues.

1. It does not take care of the correlation between structures like, for example, a) BV and OD - Blood vessels are thicker and denser near and inside OD. b) OD and Macula - The line connecting the centroid of OD and Macula approximately lies along a diameter of the fundal image. c) Exudates and BV- Development of small exudates starts near the blood vessels as exudates are pus leaking out of blood vessels. d). OD and Exudates -the possibility of confusing OD Vs Exudates in fundal images with poor quality can be avoided if OD and Exudates are segmented together by a single model.

2. Segmenting each structure separately using different models will increase the training and prediction time as a separate DL model needs to be trained for each structure.

In this work, we propose a multi-tasking deep learning architecture for simultaneous segmentation of BV, OD, Macula and Exudates. Our results show that a single network that predicts multiple structures performs better compared to detecting each structure independently using different networks, as the single network can make use of the correlation between the multiple tasks. This correlation is evident from Figure 1, where it is clear that BV is thicker and denser near and inside the OD. So the task of the OD segmentation can help the segmentation task for BV and vice versa. Similarly, the relative position of OD and Macula can help to improve the segmentation performance of each of these structures. We perform experiments and report results on four popular datasets DRIVE , HRF Budai et al. (2013), CHASE_DB and IDRiD Porwal and et al. (2018).

We use publicly available and well evaluated datasets, namely DRIVE, HRF and CHASE_DB, for our studies on BV segmentation. Since the number of images in these datasets are relatively less (40, 45 and 28 respectively for the above mentioned datasets), we use data augmentation techniques such as horizontal flip, vertical flip, rotations, elastic transformation, grid distortion and optic distortion to increase the number of training images by a factor 4-6. For Macula segmentation, we used the IDRiD localization dataset, which contains 413 training and 103 testing images. The IDRiD segmentation dataset for exudates segmentation contains a total of 81 images out of which 54 images are used for training and 27 for testing. The major contribution of this work is to propose a multi-tasking model for simultaneous segmentation of BV, OD, Macula and Exudates. The proposed multi-tasking model resulted in an improvement of F1 score by 15% for Exudates besides being faster (by a factor of 12).

2 RELATED WORK

2.1 Blood Vessel Segmentation

Existing techniques for fundal image segmentation mainly fall under two categories 1) using traditional image processing techniques and 2) using deep learning techniques. Examples of image processing based techniques include filter based: Zhang and et al. (2010), Yavuz and Köse (2011) and Aslan and et al. (2018) and morphological operations based Hassan and et al. (2015) and Singh and et al. (2014). Image processing based techniques have the advantage that domain knowledge can be easily incorporated through hand crafted features, however they are not easily generalizable across diverse datasets. Further, these algorithms are based on many customized parameters that may vary from dataset to dataset. Generalization becomes challenging since there could be hardware differences, change in acquisition conditions, different pathologies etc.

Several deep learning architectures, that were successful in segmentation tasks (Chen and et al. (2017)) in natural images, were tried for segmenting Blood Vessels in retinal images and the results were significantly better than using conventional image processing techniques. Vengalil et al. (2016) used a popular segmentation model deeplab (Chen and et al. (2017)) which was pre-trained on natural images for semantic segmentation, to segment Blood Vessels at pixel level. Jiang et al. (2018) proposes a pre-trained fully convolutional network for segmenting BV and reports accuracy of cross-dataset test on four different datasets. In M-GAN, proposed by Park and et al. (2020), a multi-kernel pooling block added between stacked convolutional layers supports scale-invariance which is a highly desirable feature for BV segmentation.

One of the main challenges in using deep neural networks for segmentation tasks is that the reduction in resolution of featuremap as one goes deeper will result in loss of finer details like edges, which are crucial for segmentation tasks. In order to circumvent this, the U-Net(Ronneberger et al. (2015)) model was introduced specifically for medical image segmentation which has multiple skip connections. In their recent work, Joshua et.al. Joshua and et al. (2020) used a modified version of U-NET for segmenting BV in retinal images and reported state-of-the-art accuracies. Laddernet, introduced by Zhuang et.al. in 2018 Zhuang (2018), is a sequence of multiple U-Nets cascaded together.

2.2 Exudates Segmentation

Like BV segmentation, reported studies on Exudates segmentation were also performed both using traditional image processing techniques and deep learning techniques . Existing works on Exudates segmentation using deep learning approaches include Perdomo et al. (2017); Tan et al. (2017); Feng et al. (2017); Zheng et al. (2018). The work reported in Perdomo et al. (2017); Kou et al. (2020) used a convolutional neural network, LeNet LeCun et al. (1989), to classify patches extracted from fundal images into classes of "with Exudates" or "without Exudates". They extract patches of size 48×48 and hence the network does not provide a pixel-wise segmentation. The work reported in Kou et al. (2020) proposed a deep learning approach, called Enhanced Residual U-Net (ERU-Net). Their proposed model had three U-paths each with three upsampling paths and one downsampling path. This structure enhances the feature fusion capability of the networks capturing more details of fundal image. The proposed model also made use of residual blocks.

3 PROPOSED METHOD

3.1 CNN architecture

A modified version of the UNET architecture proposed by Ronneberger et al. (2015), shown in Fig 3, is used for segmenting various structures. The modifications are:

1. Our proposed network retains the original dimensions of the input image, whereas the original U-Net mentioned in Ronneberger et al. (2015) reduced the original input size from 572×572 to 388×388.
2. We use de-convolutional layers with a stride of 2 for up-sampling as opposed to upsampling layers used in original U-Net architecture. Our method has the advantage that the network also learns the interpolation weights using a deconvolutional layer.
3. We add batch normalization after each convolutional layer in order to stabilize the training process as well as for faster training.

The encoder and decoder consists of 4 stages each. Each stage of the encoder comprises two convolution layers, each followed by batch normalization and ReLU activation function. A max-pooling layer with a stride of 2 is added at the end of each stage of the encoder which down samples the image by a factor of two. Each decoder layer up-samples the image by a factor of 2 using a deconvolution layer followed by a convolution layer. The sigmoid function is used in the final output channel with two filters instead of the softmax activation function because the BV and OD features are not mutually exclusive. These features share common connections in the fundal image, and hence, their simultaneous segmentation also yields the best results.

For multi-tasking experiments, the auxiliary task chosen was Optic Disc segmentation. We also evaluate models with different dimensions of latent representation and different number of channels in the bottleneck

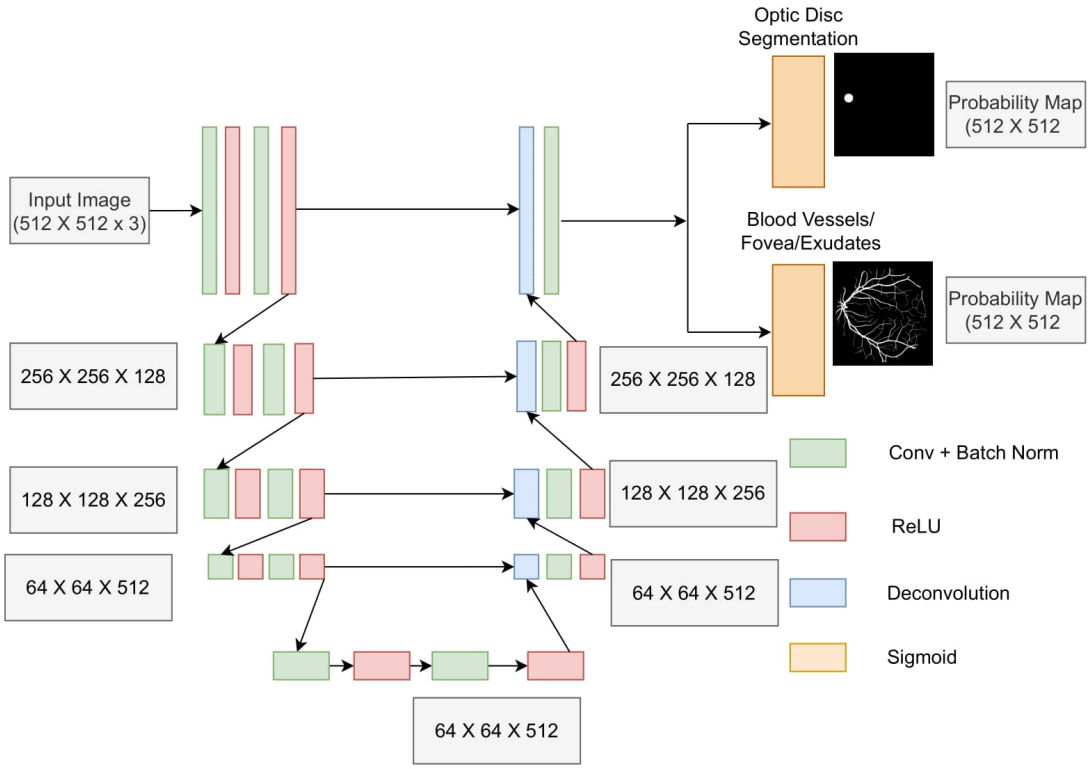

Figure 3. Architecture of the proposed multi-tasking U-Net model

layer and report the results for all combinations. For latent representation dimension, we experimented with different values 16×16, 32×32, 64×64 and 128×128. For the number of channels, experiments are carried out with 384, 512, 768, 1024 and 1280.

3.2 Dataset

We use the DRIVE, HRF Budai et al. (2013), CHASE_DB and IDRiD Porwal and et al. (2018), datasets. The DRIVE dataset contains 20 training images and 20 testing images of resolution 565×584 pixels. The dataset also provides ground truth images for BV segmentation annotated by a human expert. As the DRIVE and CHASE_DB dataset does not have Optic Disc annotations, we annotated the Optic Disc in each image ourselves. HRF dataset has 15 high resolution fundal images along with ground truth annotation for BV segmentation.

IDRiD, Porwal and et al. (2018), localization dataset which contains 413 training images and 103 test images along with Fovea ground truth, was used for Macula localization. For Exudates segmentation, we used the IDRiD segmentation dataset which contains a total of 81 images. OD ground truth is available as part of the dataset but BV ground truths are not available. For training Exudates in multi-tasking mode, we first predicted BV segmentation on these images using the model trained on images taken from HRF, DRIVE and CHASE_DB datasets. These predictions were used as BV ground truth for multi-tasking training.

Data-augmentation: horizontal and vertical flipping, grid and elastic distortion provided by the library Albumentations, Buslaev and et al. (2020), was used to increase the number of training samples by a factor of 4.

No pre-processing, other than resizing the images to 512×512, was performed on the original images.

3.3 Experiments performed

We use full images, as opposed to image patches, for training the network as a full image will have more context and hence can be more effective for predicting structures like Optic Disc and Macula. We perform multiple experiments, for individual and simultaneous prediction of Blood Vessels, Optic Disc, Macula and Exudates.

To start with, we perform BV segmentations on the individual datasets with the proposed U-Net architecture to achieve the best possible results. To further improve the results obtained on the Blood Vessels, we utilize simultaneous segmentation of Blood Vessels and Optic Disc. Apart from these experiments, we also modify the U-Net architecture to find the best suitable parameters for both Blood Vessels and Optic Disc.

Another experiment we carried out is to train a model for BV segmentation using training set containing images from HRF, DRIVE and CHASE datasets and we evaluate the performance of this trained model on 1) on a held out validation dataset which comprises images from HRF, DRIVE and CHASE_DB 2) test dataset containing images from IDRiD dataset. On the IDRiD dataset, we evaluate the model qualitatively as no ground truth was available for this dataset. Further in order to measure the generalizability of the trained model we also report across the dataset results for all tasks.

The obtained BV results on the IDRiD dataset is further utilized to improve the Optic Disc, Macula and Exudates segmentation results.

3.4 Loss function and training

A sigmoid activation function is used at the output layer. The model is trained with dice loss and with a combination of dice loss and binary cross entropy loss. In most of our experiments, we noticed that the combination of two losses gave a better F1-score.

For predicting the structures separately, four separate networks with the same architecture are trained independently one for each of the structures: Blood Vessel, Optic Disc, Macula and Exudates.

In the multi-tasking model, we train three separate networks each with two output channels, for predicting the following three combinations:

1. BV and OD
2. Macula and OD
3. Exudates and OD
4. BV and Macula
5. BV and Exudates
6. Exudates, BV and OD
7. Macula, BV and OD

The network is trained for 60 epochs in all the cases.

4 RESULTS AND DISCUSSION

We use multiple metrics, Accuracy, Dice score, ROC-AUC and Jaccard Index (Intersection over Union) for evaluating the model performance. Definition and mathematical expression of each of these metrics is given below:

True Positive (TP): Number of positive pixels in the image which the model also correctly predicted as positive.

True Negative (TN): Number of negative pixels which the model correctly predicted as negative.

False positive (FP): Number of negative pixels in the image which the model predicted incorrectly as Positive.

False Negative (FN): Number of pixels which are actually positive but the model predicted as negative.

Accuracy: Accuracy is the ratio of the total number of correctly predicted positive and negative pixels (sum of True Positives and True Negatives) to the total number of pixels in the image.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Dice Score: Dice score is the ratio of two times overlap (intersection of predicted and actual positive pixels) to that of sum of number of positively labelled pixels in ground truth and the number of pixels which the model predicted as positive.

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{2}$$

Jaccard Index (JI): Also known as Intersection over Union (IoU), is the ratio of intersection of predicted and ground truth pixels (which is same as TP) to the union of prediction and ground truth.

$$Jaccard = \frac{TP}{TP + FP + FN} \tag{3}$$

True Positive Rate (TPR): Also known as Sensitivity or Recall, is the ratio of number of True Positives to the total number of positive samples.

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

False Positive Rate (FPR): It is the ratio of the number of False Positives to the total number of negative pixels.

$$FPR = \frac{FP}{TN + FP} \tag{5}$$

ROC-AUC: Receiver Operating Characteristic (ROC) is a plot of True Positive Rate vs False Positive Rate computed at various thresholds. Area under ROC curve is a measure of a model's ability to discriminate between positive and negative samples and this metric independent of the threshold.

In addition to above metrics, we also compare our approach with other state-of-the-art approaches with inclusion of prediction time.

For comparison of multi-tasking segmentation with segmentation of individual structures, experiments on individual prediction were performed. In this case, the network had one output channel corresponding to the segmentation map. The network outputs a binary image, of the same resolution as the input, which indicates

Table 1. Comparison of results with and without multi-tasking. For all the tasks except OD, multi-tasking was done with OD as an auxiliary task. For OD, multi-tasking was done in combination with Blood Vessels.

		Individual		**Multi-tasking**	
	Dataset	***Dice(%)***	***JI(%)***	***Dice(%)***	***JI(%)***
Blood Vessel	DRIVE	77.00	62.78	80.31	67.35
	HRF	78.11	64.29	81.66	69.04
	CHASE_DB	74.34	59.21	80.45	67.32
Optic Disc	DRIVE	76.24	66.15	78.63	69.85
	IDRiD	85.73	64.65	94.51	89.98
Macula	IDRiD	68.13	60.16	70.52	61.77
Exudates	IDRiD	50.34	34.56	61.37	46.33

Table 2. Results of various multi-tasking experiments run on segmentation of Exudates. We got the best results when the model is trained with multi-tasking loss for a combination of Exudates, OD and BV.

Experiments Run	**Dice(%)**	**AUC**
Exudates Alone	50.34	0.8402
Exudates and OD	53.32	0.8078
Exudates and BV	61.37	0.8365
Exudates, OD and BV	**65.00**	**0.9993**

pixel by pixel segmentation. For multi-tasking models, additional channels for predicting additional structures in combination with another structure are added at the output layer.

Table 1 compares segmentation performance of various structures when the model is trained with and without multi-tasking. The table provides a comparison of results when the model is trained in multi-tasking mode with two different tasks Vs the results with an individual task. In all these cases, OD is added as an auxiliary task along with other main tasks like BV, Macula and Exudates. The row for OD shows comparison of segmentation performance of OD when trained with OD alone Vs results of multi-tasking loss with BV. As evident from the table, Dice score for Exudates segmentation resulted in an improvement of 10% when the model was trained in combination with Optic Disc. A similar trend is observed in BV and OD segmentation with an improved Dice score of about 4% (for HRF) and 6% (for CHASE_DB). For Macula, the segmentation results decreased by 2% upon addition of Optic Disc as an additional task during training. This happens because the Macula and Optic Disc are mutually exclusive as they appear at different locations in the image. In all our simultaneous segmentation experiments we use sigmoid as output activation function which is used for multi-label prediction which is not true in this case. However, when trained with Blood Vessels, Macula and OD together we observed a 3% increase in the score for Macula.

For Exudates we carried out multiple experiments which are tabulated in Table 2. As given in the table, when multi-tasking was done with three tasks (OD, Exudates and BV segmentation) the score for Exudates segmentation increased by 15% to a value of 65%. Also AUC improved from 0.8402, when trained individually, to 0.9993 when trained in combination with BV and OD. Figure 4 shows the ROC-AUC curve for all four Exudates segmentation experiments.

Figure 5 shows sample BV segmentation results with and without multi-tasking for HRF, DRIVE and CHASE_DB. As evident from the figure, the increase in Dice score is due to less number of false negatives

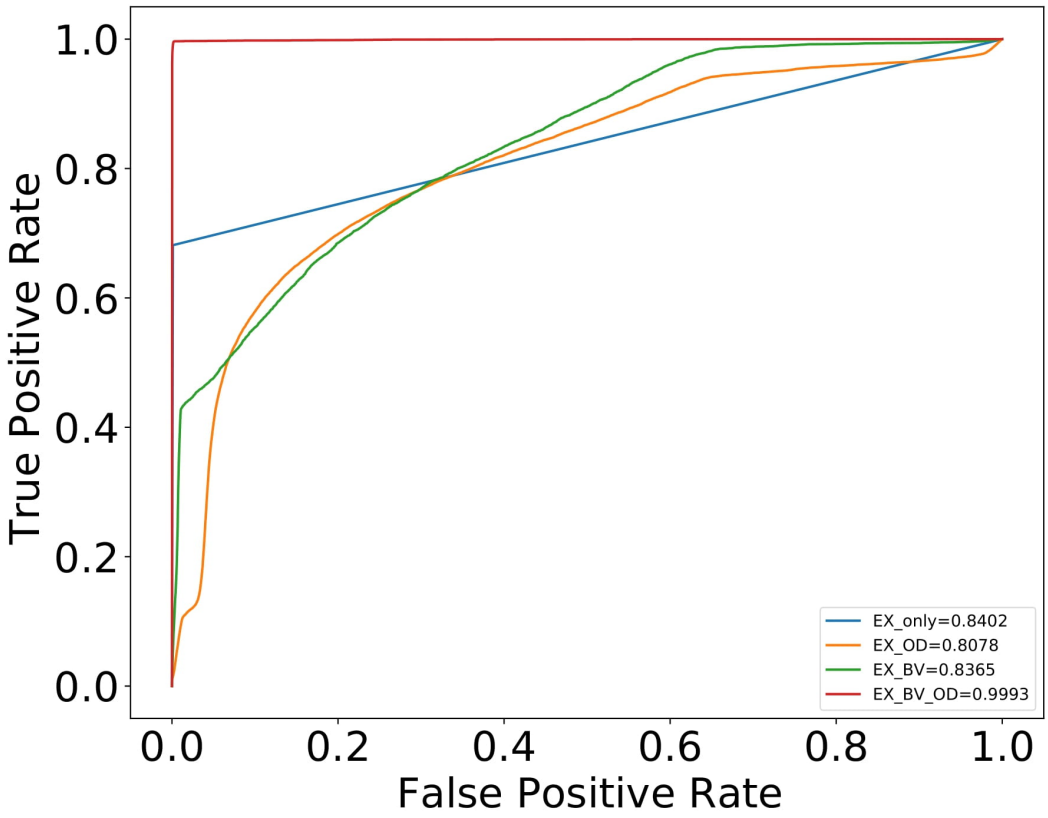

Figure 4. ROC-AUC curve for different experiments run on Exudates segmentation.

(red pixels) in the prediction (resulting in larger precision). Similarly Figures 6, 7 and 8 shows comparison of segmentation results for OD, Macula and Exudates respectively.

The improved performance with multi-tasking is a consequence of direct correlation between the two predicted structures. When trained together, the network is able to learn new hidden layer features that can contribute to the prediction of both structures.

When trained individually, the OD can easily be confused as Exudates as both appear as white patches. In simultaneous segmentation, the network learns to discriminate Optic Disc and Exudates, using some other features like shape for example, which improves the segmentation results for both.

Figure 9 shows the plot of validation Dice score Vs feature map dimension and Fig 10 shows Dice score as a function of number of channels in the bottleneck layer. As evident from the graphs, we got the best results when the feature map dimension is 64×64 and the number of channels is 1024.

Figure 11 shows the results of BV segmentation on the IDRiD dataset using the model trained with images from DRIVE, HRF and CHASE_DB datasets. These segmentation results were used as BV ground truth of IDRiD images while training for Exudates in multi-tasking mode.

Figure 12 shows the Receiver Operating Characteristics(ROC) curve of the best results we got for all three structures BV, OD and Macula. Figure also shows the ROC curve for the pathological indicator Exudates. BV segmentation results are on DRIVE test images. OD, Macula and Exudates results are on the IDRiD dataset. As evident from the figure, the ROC curve for Exudates, which is an indicator for early

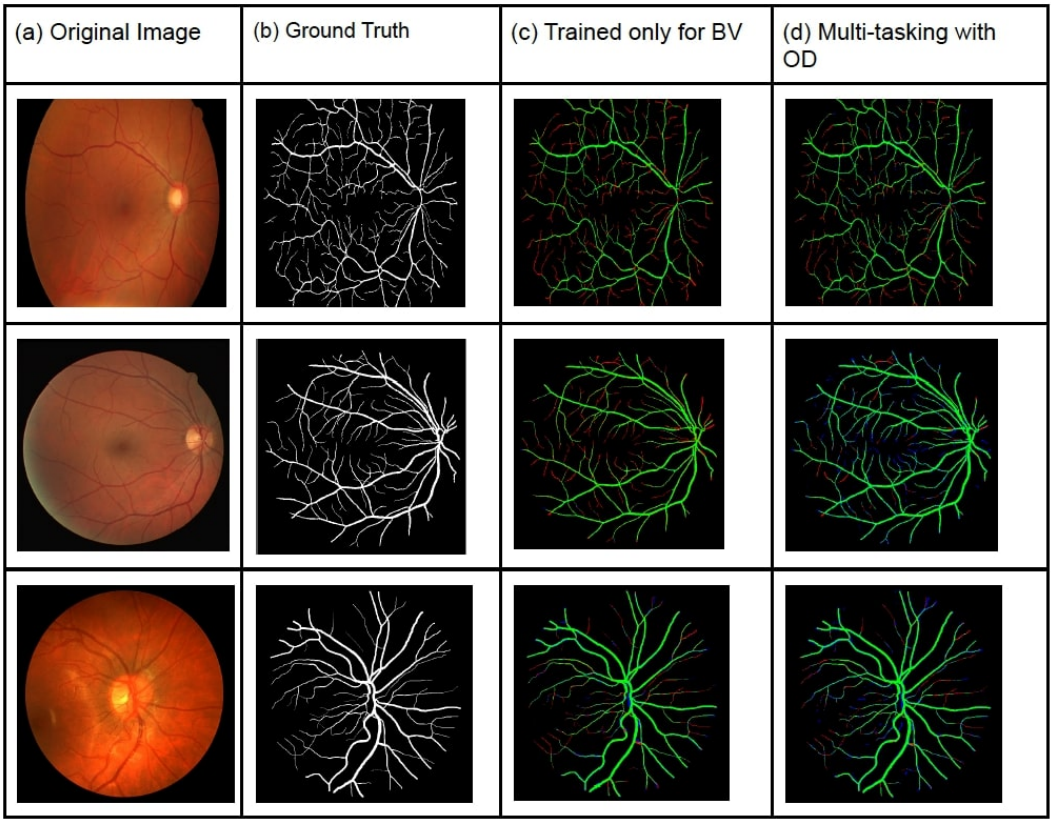

Figure 5. A comparison of Blood Vessels segmentation results for HRF, DRIVE and CHASE_DB (from top to bottom row) datasets with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false negatives. An increase in F1-score score of 4.67%, 3.31%, 5.83% is observed in HRF, DRIVE and CHASE_DB respectively.

Figure 6. A comparison of OD segmentation results for the IDRiD dataset with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false negatives. An increase in F1-score score of 9% is observed.

detection of many retinal diseases like DR, is encouraging. It is worth noting that we got these improved results as a consequence of multi-tasking with BV and OD.

Table 3 shows comparison of our Exudates segmentation results with state-of-the-art techniques. AUC of our segmentation result is better than state-of-the-art DL methods Kou et al. (2020) and more significantly our prediction time is significantly better than the state-of-the-art DL approaches. This is because, we are doing whole image segmentation in single forward-pass, where as many DL state-of-the-art approaches approaches used patch-wise training and hence prediction of a single image takes longer as the network

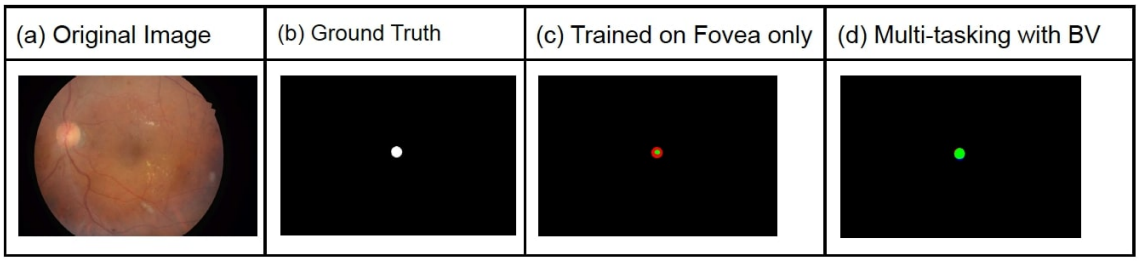

Figure 7. A comparison of Macula segmentation results for the IDRiD dataset with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false negatives. An increase in F1-score score of 58% is observed on this image.

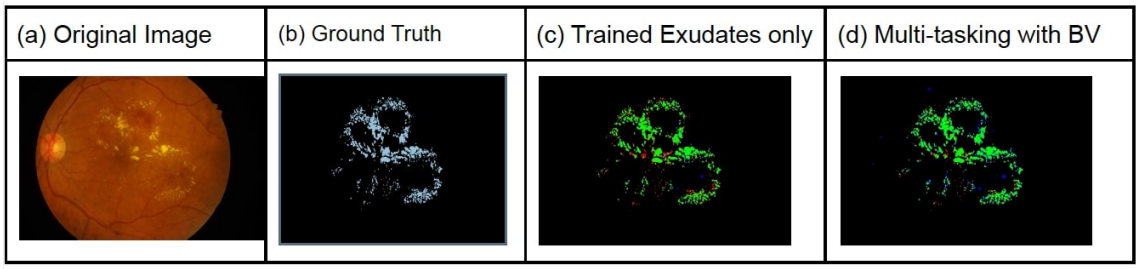

Figure 8. A comparison of Exudates segmentation results for the IDRiD dataset with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false negatives. An increase in F1-score score of 15% is observed.

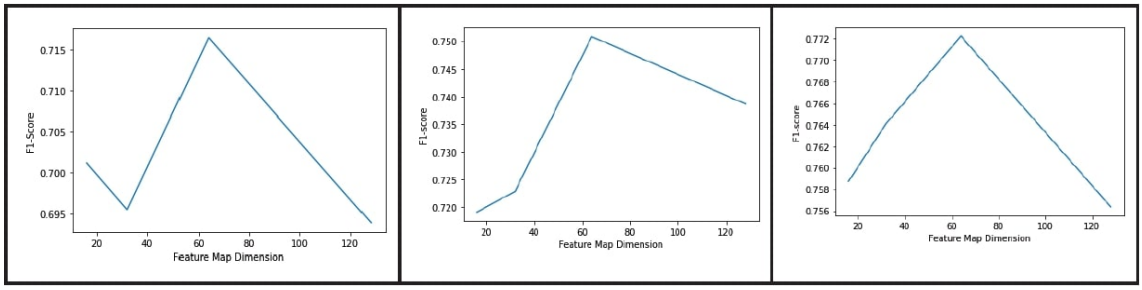

Figure 9. Plot of Dice score Vs Dimension of feature map in bottleneck layer for Blood vessel segmentation on DRIVE, CHASE_DB and HRF (from left to right) datasets.

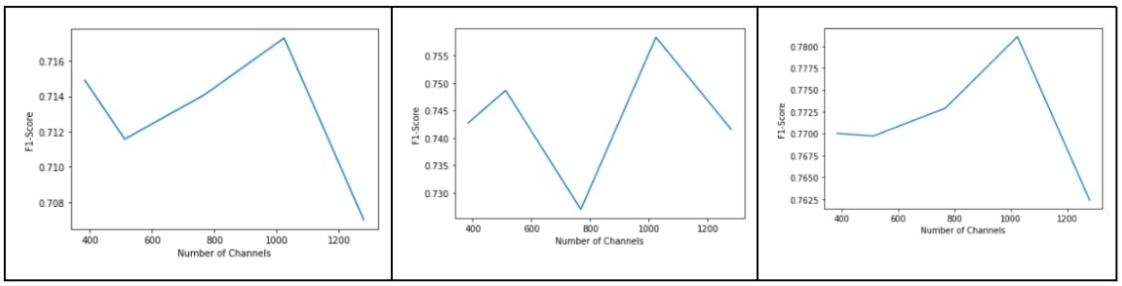

Figure 10. Plot of Dice score Vs number of channels in bottleneck layer for Blood vessel segmentation on DRIVE, CHASE_DB and HRF (from left to right) datasets.

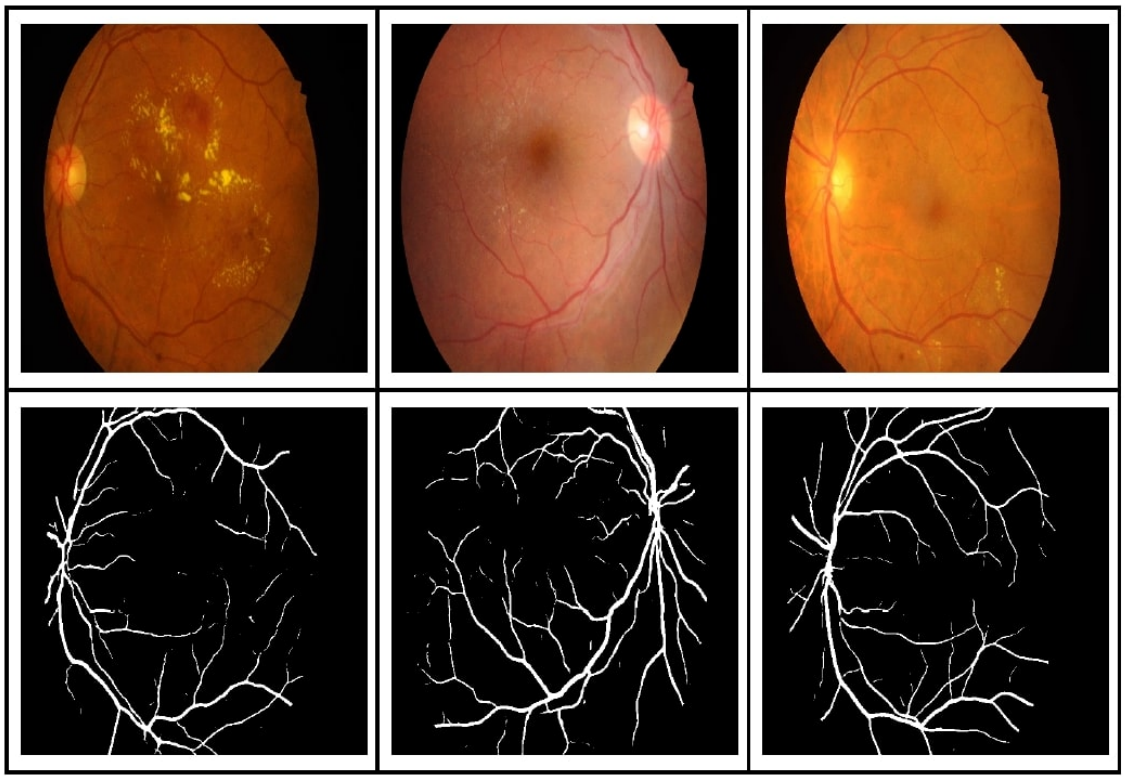

Figure 11. BV segmentation results on the IDRiD dataset using the model trained with images selected from DRIVE, HRF and CHASE_DB.

needs to predict on each patch and finally all the predictions needs to be combined together to get the overall all segmentation results for the image. We believe that the AUC can further be improved by adding more data-augmentation techniques and also by including more images in the original training set. Our improvement in AUC is attributed to faster learning of more generalized features as a consequence of multi-tasking. Another major advantage of our approach is that we get segmentation of two additional different structures, BV and OD, along with Exudates segmentation. Nur and Tjandrasa (2018) got an accuracy of 99.33% by removing the OD first and then obtaining the salient regions using intensity thresholding. The challenge with this method is that the threshold can vary from dataset to dataset and further the accuracy also depends on the OD removal step.

Figure 13 shows comparison of Exudates segmentation results of the proposed approach with the approach mentioned in Guo et al. (2020) for some sample images. Even though we got better accuracy (99.42) and AUC (0.9993) when averaged over the entire test images, our model failed to capture very small Exudates, as shown in the second row of Figure 13. This is because such finer Exudates get removed when the image is resized from original size of 4288×2848 to 768×512.

Table 4 shows the comparison of our BV segmentation results with the state-of-the-art techniques. As evident from the table accuracy (95.89%) of our approach on the DRIVE dataset is better than the currently reported state-of-the-art and AUC is close to the state-of-the-art. Dice scores on all three dataset are lower than patch based state-of-the-art techniques. However, since we train on whole images, both training and prediction time is about 20 times faster than patch based deep learning approaches.

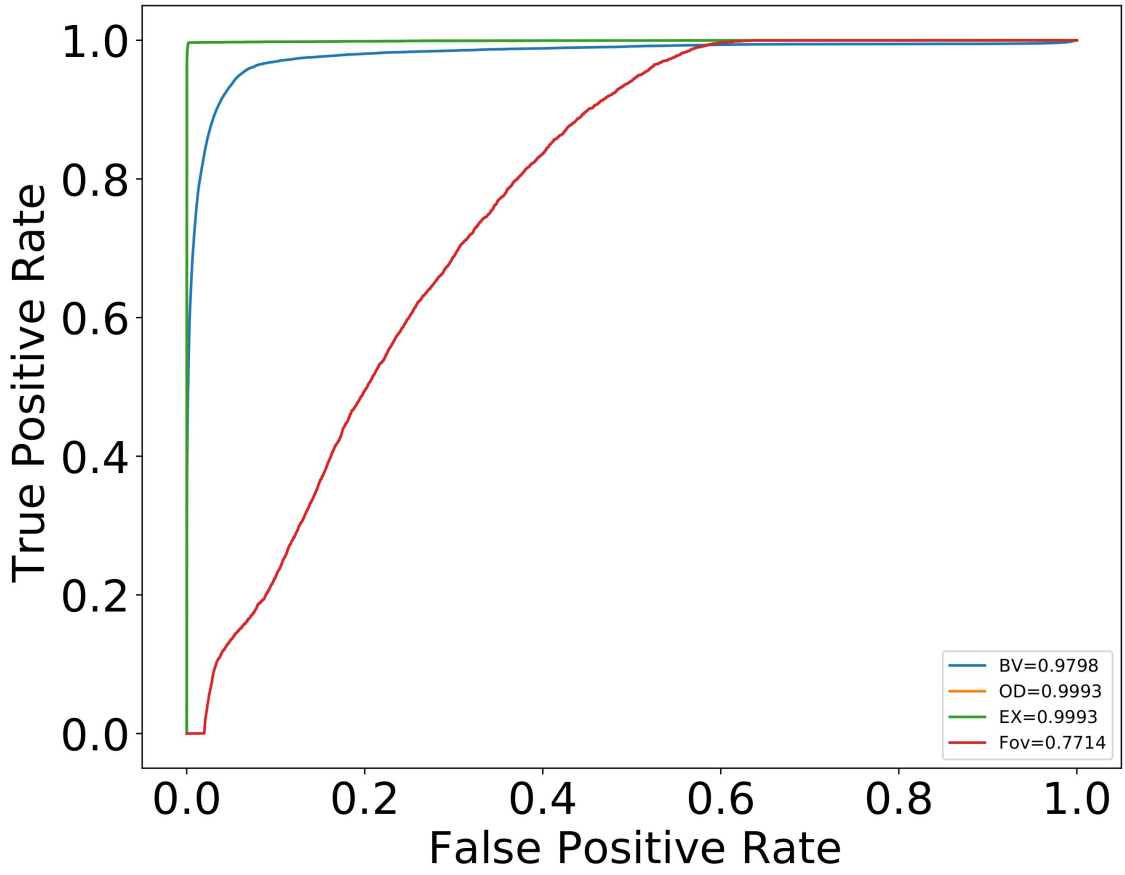

Figure 12. ROC curve for all three structures BV, OD and Macula and the pathological indicator Exudates. BV results are on the DRIVE test dataset. OD, Macula (FOV) and Exudates (EX) are on the IDRiD dataset.

Table 3. Comparison of Exudates segmentation results of our approach with other state-of-the-art approaches on the IDRiD dataset. As we are predicting on whole images, our prediction time is much faster (about 12 times) compared to other state-of-the-art techniques on the IDRiD dataset. FCN:Fully Convolutional Networks, ACC:Accuracy, SUP:Supervised, UNSUP: Unsupervised

Author/ Year	**Approach**			**Performance Metrics**		
	Method	**SUP/ UNSUP**	**Patch-wise/ Whole Image**	**AUC**	**ACC(%)**	**Prediction time(sec)**
Kaur and Kaur (2022)	UNet with InceptionV3	SUP		-	99.83	-
Hamad et al. (2020)	FCM Clustering	UNSUP	Patch-wise (256×256)	-	99.2	30
Kou et al. (2020)	ER-UNet	SUP	Patch-wise	0.9801	98.00	37.3
Guo et al. (2020)	Deeplab-V2 with Bin loss	SUP	Patch-wise (51×51)	0.9162	99	-
Nur and Tjandrasa (2018)	Saliency based	UNSUP	Patch-wise (32×32)	-	99.33	-
Our approach	Modified U-NET Multi-tasking	SUP	Whole Image	**0.9993**	**99.42**	**2-3**

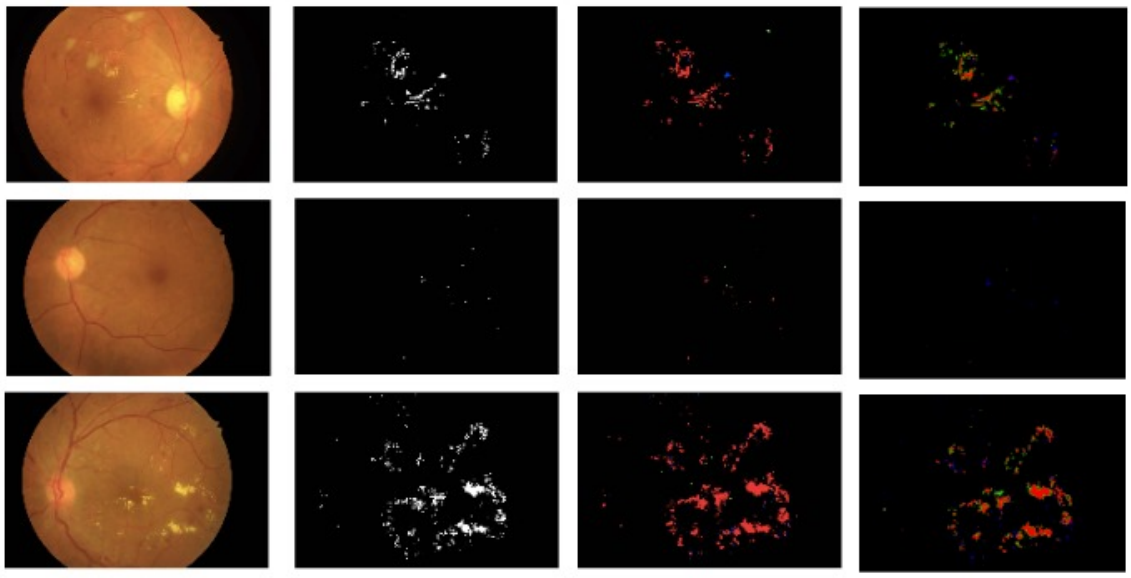

Figure 13. Comparison of Exudates segmentation result of the proposed approach with Guo et al. (2020). Each column from left shows the original image, ground truth segmentation results of Guo et al. (2020) and segmentation result of proposed approach. Color code used is kept the same as in Guo et al. (2020) for ease of comparison. Red: True Positive, Green: False Positive, Blue: False Negative.

5 CONCLUSION

In this work, we illustrate the efficacy of modified multi-tasking U-Net for segmenting Blood Vessels, Optic Disc, Macula and Exudates in fundal images. The proposed approach resulted in a peak increase of 15% in Dice score for Exudates segmentation compared to the individual segmentation result with the same architecture. Using the proposed approach, we are able to get a state-of-the-art accuracy of 95.89% on DRIVE test images which is 0.7% greater than one of the recently reported results. With our proposed method, the prediction times on the images are significantly lesser (12 times) compared to most other deep learning methods. In addition to this increased speed in prediction, we could improve the AUC and Accuracy for Exudates from 0.9801 to 0.9993 and 99.33% to 99.42% on the IDRiD test dataset.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Sunil Kumar Vengalil and Bharath K contributed in conceptualization, investigation and to develop the methodology. Bharath K carried out the implementation and ran all the experiments mentioned in this work. Both Bharath K and Sunil Kumar Vengalil contributed equally in writing and formatting the manuscript. Neelam Sinha supervised the work and was involved in writing-reviewing-editing. Neelam Sinha validated the work and provided the resources.

Table 4. Comparison of BV segmentation results of our approach with other state-of-the-art approaches.ACC: Accuracy

Author/ Year	**Approach**	Patch-wise/ Whole Image	**Dataset**	**Performance Metrics**		
				Dice	**AUC**	**ACC**
Xu et al. (2021)	Residual Attention with ASPP and Deep Supervision	Patch-wise (64×64)	DRIVE	-	0.97	95.9
Liu (2021)	Hand-crafted features with MLP	Whole Image	DRIVE	-	-	95.82
Jiang et al. (2018)	FCNN Transfer learning	Patch-wise (50×50)	DRIVE HRF CHASE_DB	-	0.98 0.97 0.98	-
Joshua and et al. (2020)	U-Net	Whole Image	DRIVE HRF CHASE_DB	87.62 85.11 85.69	-	-
Park and et al. (2020)	M-GAN	Path-wise (48×48)	DRIVE HRF CHASE_DB	83.24 79.92 81.10	-	-
Sun et al. (2021)	Data Augmentation	Whole Image	DRIVE CHASE_DB	82.09 75.65	-	-
Adapa et al. (2020)	Zernike Moment	Whole Image	DRIVE	-	-	94.5
Dash et al. (2020)	**Preprocessing**: CLACHE Gabor Hessian **Segmentation**: k-means **Post Processing**: Morphological cleaning	Whole Image	DRIVE CHASE_DB	-	-	95.2 95
Our approach	Multi-tasking using U-NET	Whole Image	DRIVE HRF CHASE_DB	80.31 81.66 80.45	0.98	95.89

FUNDING

This work was funded by Machine Intelligence and Robotics Center (MINRO) Project GoK, IIITB. It is supported by Karnataka Innovation & Technology Society, Dept. of IT, BT and S&T, Govt. of Karnataka vide GO No. ITD 76 ADM 2017, Bengaluru; Dated 28.02.2018

DATA AVAILABILITY STATEMENT

1. Digital Retinal Images for Vessel Extraction Dataset (DRIVE) dataset is available at `http://www.isi.uu.nl/Research/Databases/DRIVE/`
2. High-Resolution Fundus (HRF) Image Database `https://www5.cs.fau.de/research/data/fundus-images/`
3. CHASE_DB1 Retinal Image Database `https://www.idiap.ch/software/bob/docs/bob/bob.db.chasedb1/master/index.html`

4. Indian Diabetic Retinopathy Image Dataset (IDRiD) `https://idrid.grand-challenge.org/`

REFERENCES

Adapa, D., Joseph Raj, A. N., Alisetti, S. N., Zhuang, Z., and Naik, G. (2020). A supervised blood vessel segmentation technique for digital fundus images using zernike moment based features. *Plos one* 15, e0229831

Aslan, M. F. and et al., C. (2018). Segmentation of retinal blood vessel using gabor filter and extreme learning machines. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (IEEE), 1–5

Budai, A., Bock, R., Maier, A., Hornegger, J., and Michelson, G. (2013). Robust vessel segmentation in fundus images. *International journal of biomedical imaging* 2013

Buslaev, A. and et al., I. (2020). Albumentations: fast and flexible image augmentations. *Information* 11, 125

Chen, L.-C. and et al., P. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* 40, 834–848

Dash, J., Parida, P., and Bhoi, N. (2020). Retinal blood vessel extraction from fundus images using enhancement filtering and clustering. *ELCVIA: electronic letters on computer vision and image analysis* 19, 0038–52

Feng, Z., Yang, J., Yao, L., Qiao, Y., Yu, Q., and Xu, X. (2017). Deep retinal image segmentation: a fcn-based architecture with short and long skip connections for retinal image segmentation. In *International conference on neural information processing* (Springer), 713–722

Guo, S., Wang, K., Kang, H., Liu, T., Gao, Y., and Li, T. (2020). Bin loss for hard exudates segmentation in fundus images. *Neurocomputing* 392, 314–324

Hamad, H., Dwickat, T., Tegolo, D., and Valenti, C. (2020). Exudates as landmarks identified through fcm clustering in retinal images. *Applied Sciences* 11, 142

Hassan, G. and et al., E.-B. (2015). Retinal blood vessel segmentation approach based on mathematical morphology. *Procedia Computer Science* 65, 612–622

Hu, F. B., Satija, A., and Manson, J. E. (2015). Curbing the diabetes pandemic: the need for global policy solutions. *Jama* 313, 2319–2320

Jiang, Z., Zhang, H., Wang, Y., and Ko, S.-B. (2018). Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Computerized Medical Imaging and Graphics* 68, 1–15

Joshua, A. O. and et al., N. (2020). Blood vessel segmentation from fundus images using modified u-net convolutional neural network. *Journal of Image and Graphics* 8, 21–25

Kaur, J. and Kaur, P. (2022). Uniconv: An enhanced u-net based inceptionv3 convolutional model for dr semantic segmentation in retinal fundus images. *Concurrency and Computation: Practice and Experience* , e7138

Kou, C., Li, W., Yu, Z., and Yuan, L. (2020). An enhanced residual u-net for microaneurysms and exudates segmentation in fundus images. *IEEE Access* 8, 185514–185525

Krestanova, A., Kubicek, J., and Penhaker, M. (2020). Recent techniques and trends for retinal blood vessel extraction and tortuosity evaluation: A comprehensive review. *Ieee Access* 8, 197787–197816

Krizhevsky, A. and et al., S. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25, 1097–1105

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., et al. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems* 2

Liu, Z. (2021). Construction and verification of color fundus image retinal vessels segmentation algorithm under bp neural network. *The Journal of Supercomputing* 77, 7171–7183

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440

Nur, N. and Tjandrasa, H. (2018). Exudate segmentation in retinal images of diabetic retinopathy using saliency method based on region. In *Journal of Physics: Conference Series* (IOP Publishing), vol. 1108, 012110

Park, K.-B. and et al., C. (2020). M-gan: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks. *IEEE Access* 8, 146308–146322

Perdomo, O., Arevalo, J., and González, F. A. (2017). Convolutional network to detect exudates in eye fundus images of diabetic subjects. In *12th International Symposium on Medical Information Processing and Analysis* (SPIE), vol. 10160, 235–240

[Dataset] Porwal, P. and et al., P. (2018). Indian diabetic retinopathy image dataset (idrid). doi:10.21227/H25W98

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

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*

Singh, D. and et al., S. (2014). A new morphology based approach for blood vessel segmentation in retinal images. In *2014 annual IEEE India conference (INDICON)* (IEEE), 1–6

Sun, X., Fang, H., Yang, Y., Zhu, D., Wang, L., Liu, J., et al. (2021). Robust retinal vessel segmentation from a data augmentation perspective. In *International Workshop on Ophthalmic Medical Image Analysis* (Springer), 189–198. Doi:https://doi.org/10.1007/978-3-030-87000-3_20

Tan, J. H., Fujita, H., Sivaprasad, S., Bhandary, S. V., Rao, A. K., Chua, K. C., et al. (2017). Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Information sciences* 420, 66–76

Vengalil, S. K., Sinha, N., Kruthiventi, S. S., and Babu, R. V. (2016). Customizing cnns for blood vessel segmentation from fundus images. In *2016 International Conference on Signal Processing and Communications (SPCOM)* (IEEE), 1–4

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., et al. (2018). Understanding convolution for semantic segmentation. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1451–1460. doi:10.1109/WACV.2018.00163

[Dataset] Wikipedia (2022). https://en.wikipedia.org/wiki/Macula_of_retina

Xu, S., Chen, Z., Cao, W., Zhang, F., and Tao, B. (2021). Retinal vessel segmentation algorithm based on residual convolution neural network. *Frontiers in Bioengineering and Biotechnology* 9

Yavuz, Z. and Köse, C. (2011). Retinal blood vessel segmentation using gabor filter and top-hat transform. In *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU)* (IEEE), 546–549

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., and Sang, N. (2018). Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*

Zhang, B. and et al., Z. (2010). Retinal vessel extraction by matched filter with first-order derivative of gaussian. *Computers in biology and medicine* 40, 438–445

Zheng, R., Liu, L., Zhang, S., Zheng, C., Bunyak, F., Xu, R., et al. (2018). Detection of exudates in fundus photographs with imbalanced learning using conditional generative adversarial network. *Biomedical optics express* 9, 4863–4878

Zhuang, J. (2018). Laddernet: Multi-path networks based on u-net for medical image segmentation. *arXiv preprint arXiv:1810.07810*