

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

2 ABSTRACT

3 Fundal imaging is the most commonly used non-invasive technique for early detection of many
4 retinal diseases like Diabetic Retinopathy(DR). An initial step in automatic processing of fundal
5 images for detecting diseases is to identify the normal landmarks: Optic Disc, Blood Vessels
6 and Macula. In addition to these, various abnormalities like exudates that help in pathological
7 analysis are also visible in fundal images. In this work, we propose a multi-tasking deep learning
8 architecture for segmenting Optic Disc, Blood Vessels, Macula and exudates simultaneously. The
9 objective is to improve the prediction results on Blood Vessels and exudates, which are relatively
10 more challenging, while utilizing segmentation of Optic Disc and Macula as auxiliary tasks. Our
11 experimental results on publicly available datasets show that simultaneous segmentation of all
12 these structures results in significant improvement in the performance. Among the segmented
13 structures maximal improvement of 15% is obtained on exudates when trained using multi-tasking
14 loss function. The proposed approach makes predictions for the whole image in a single forward
15 pass. We use modified U-Net architecture with only convolutional and de-convolutional layers
16 with comparatively less number of parameters leading to faster (by a factor of 12) prediction time,
17 compared to the current AUC-wise best performing exudate segmentation deep learning model,
18 ERUnet. The proposed approach has been evaluated on publicly available datasets DRIVE, HRF,
19 CHASE_DB and IDRiD datasets. Comparison of results with the state-of-the-art have also been
20 presented.

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

1 INTRODUCTION

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

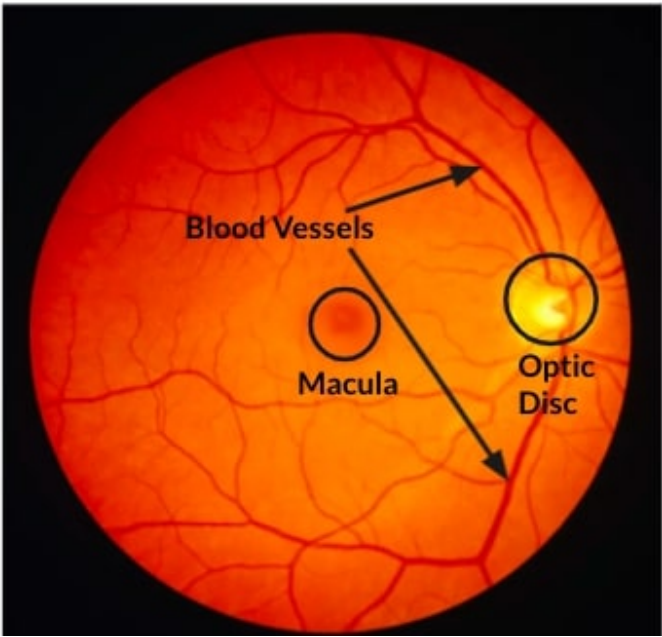


Figure 1. Sample fundal image showing normal landmarks BV, OD and Macula

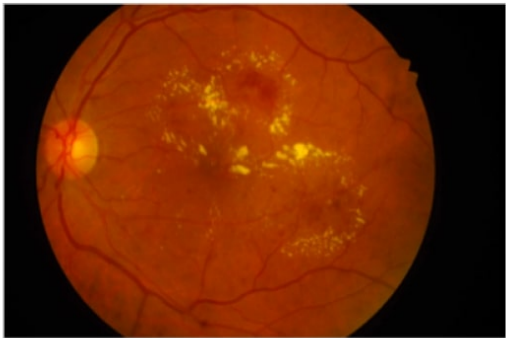

a) Original Image	b) Ground Truth
	

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

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. It is indicative of advanced stage of Diabetic Retinopathy. Exudates appear as unstructured, scattered, bright patches in the fundal image.

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) . 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. One of the issues in using deep learning architecture for medical images is limited availability of annotated training data. 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.

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. 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 was 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

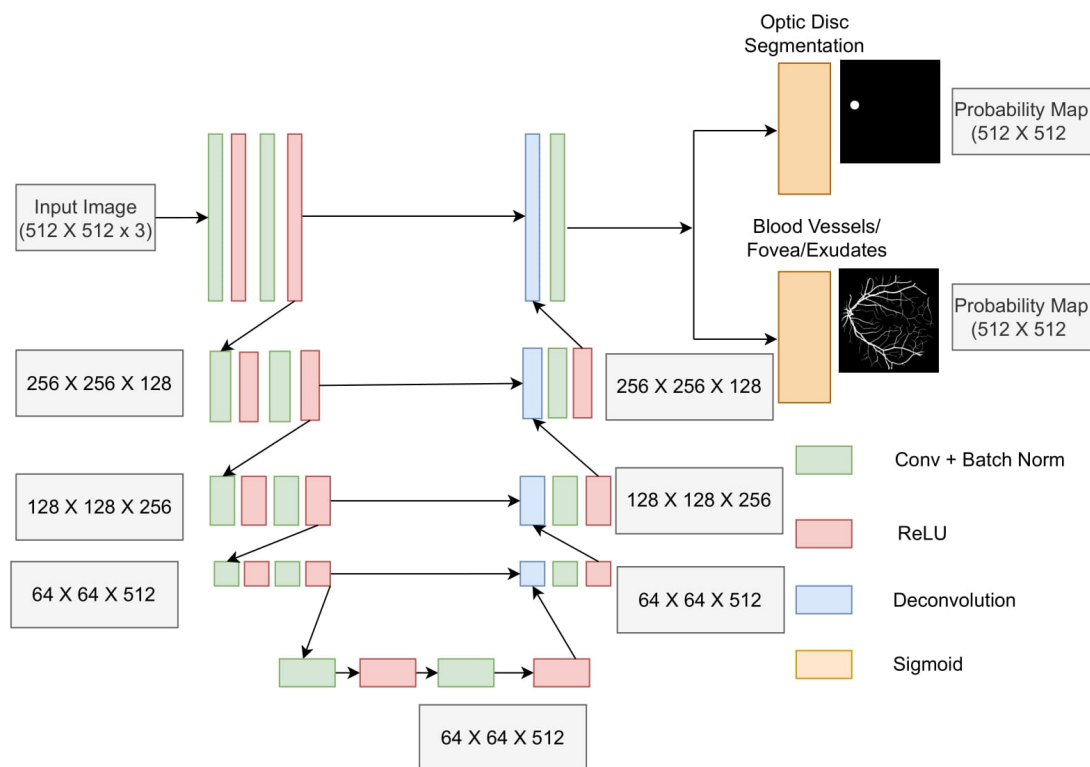


Figure 3. Architecture of the proposed multi-tasking U-Net model

123 share common connections in the fundal image, and hence, their simultaneous segmentation also yields the
124 best results.

125 For multi-tasking experiments, the auxiliary task chosen was Optic Disc segmentation. We also evaluate
126 models with different dimensions of latent representation and different number of channels in the bottleneck
127 layer and report the results for all combinations. For latent representation dimension, we experimented
128 with different values 16×16 , 32×32 , 64×64 and 128×128 . For the number of channels, experiments
129 are carried out with 384, 512, 768, 1024 and 1280.

130 3.2 Dataset

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

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

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.

	<i>Dataset</i>	Individual		Multi-tasking	
		<i>Dice(%)</i>	<i>IoU(%)</i>	<i>Dice(%)</i>	<i>IoU(%)</i>
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

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 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

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

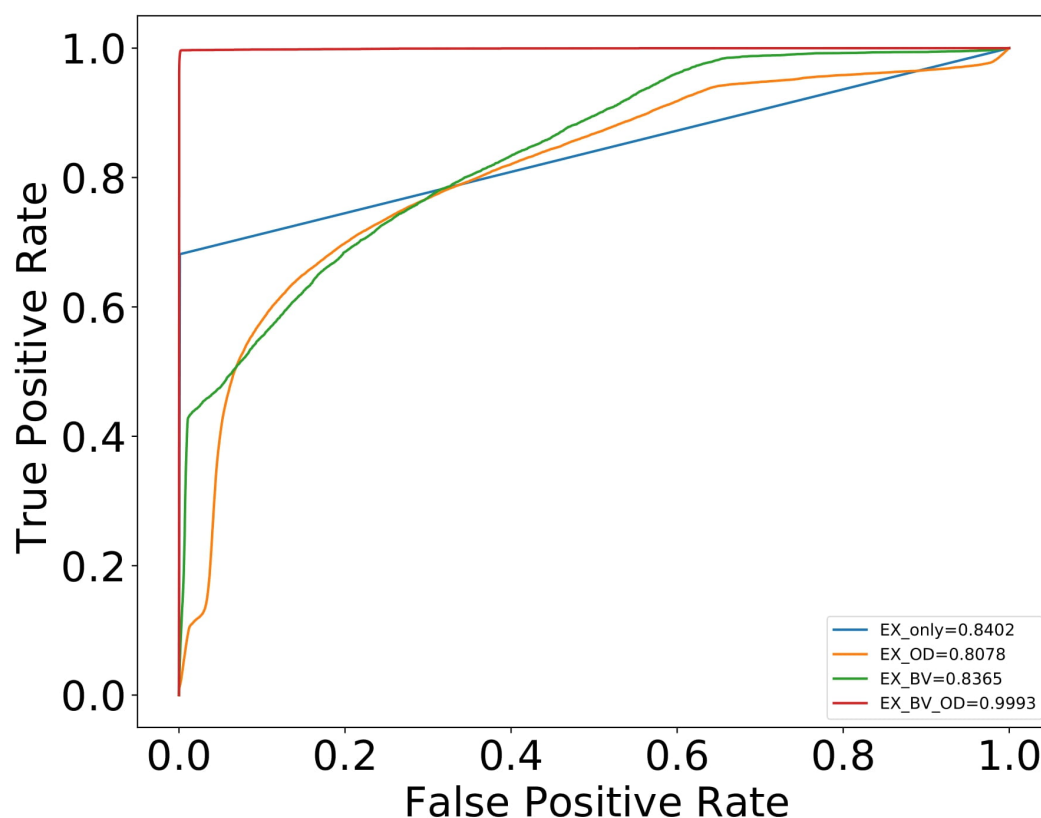


Figure 4. ROC-AUC curve for different experiments run on exudates segmentation.

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 (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.

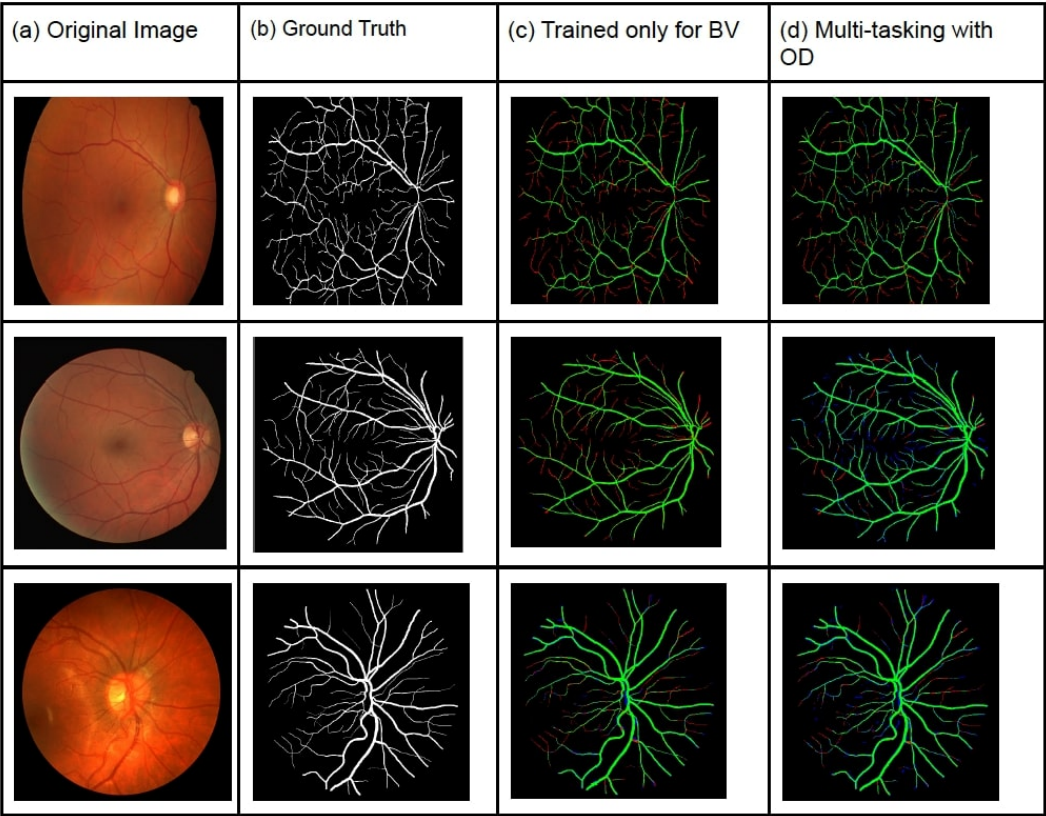


Figure 5. A comparison of Blood Vessels segmentation results for HRF, DRIVE and CHASE_DB (from top to bottom row) datasets with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false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

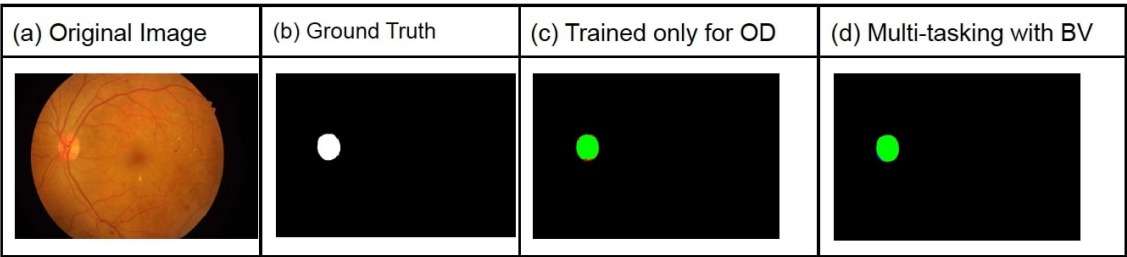


Figure 6. A comparison of OD segmentation results for the IDRiD dataset with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false negatives. An increase in F1-score score of 9% is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

221 Figure 12 shows the Receiver Operating Characteristics(ROC) curve of the best results we got for all
222 three structures BV, OD and Macula. Figure also shows the ROC curve for the pathological indicator
223 exudates. BV segmentation results are on DRIVE test images. OD, Macula and exudates results are on the
224 IDRiD dataset. As evident from the figure, the ROC curve for exudates, which is an indicator for early
225 detection of many retinal diseases like DR, is encouraging. It is worth noting that we got these improved
226 results as a consequence of multi-tasking with BV and OD.

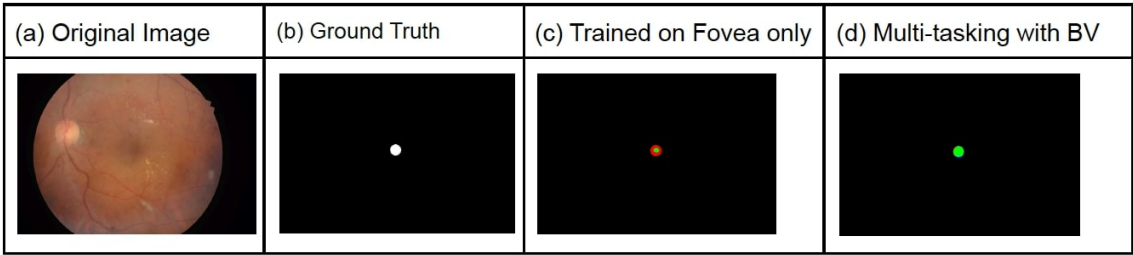


Figure 7. A comparison of Macula segmentation results for the IDRiD dataset with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false negatives. An increase in F1-score score of 58%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

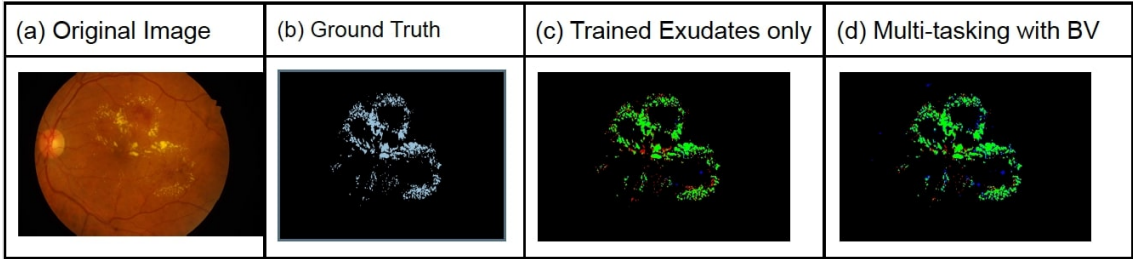


Figure 8. A comparison of exudates segmentation results for the IDRiD dataset with and without multi-tasking. The green pixels in the predicted image correspond to true positives, blue pixels correspond to false positives and red pixels correspond to false negatives. An increase in F1-score score of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

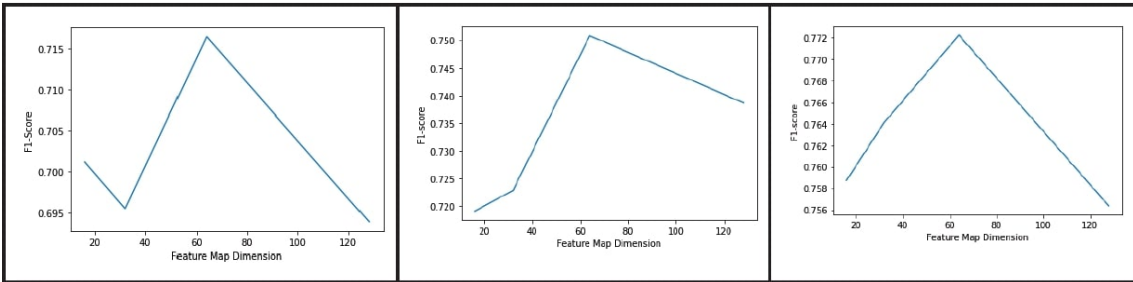


Figure 9. Plot of Dice score Vs Dimension of feature map in bottleneck layer for Blood vessel segmentation on DRIVE, CHASE_DB and HRF (from left to right) datasets.

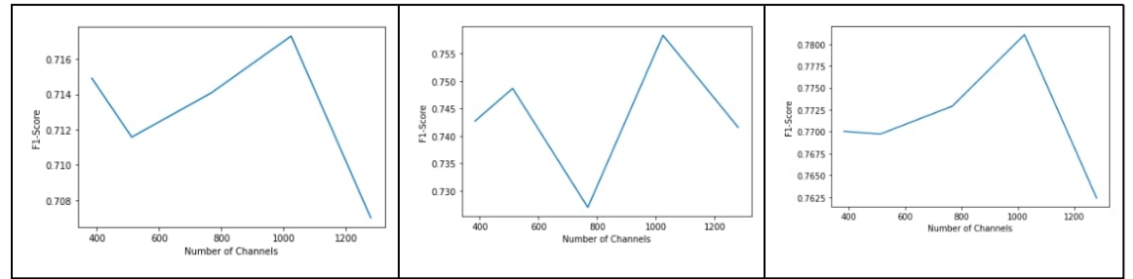


Figure 10. Plot of Dice score Vs number of channels in bottleneck layer for Blood vessel segmentation on DRIVE, CHASE_DB and HRF (from left to right) datasets.

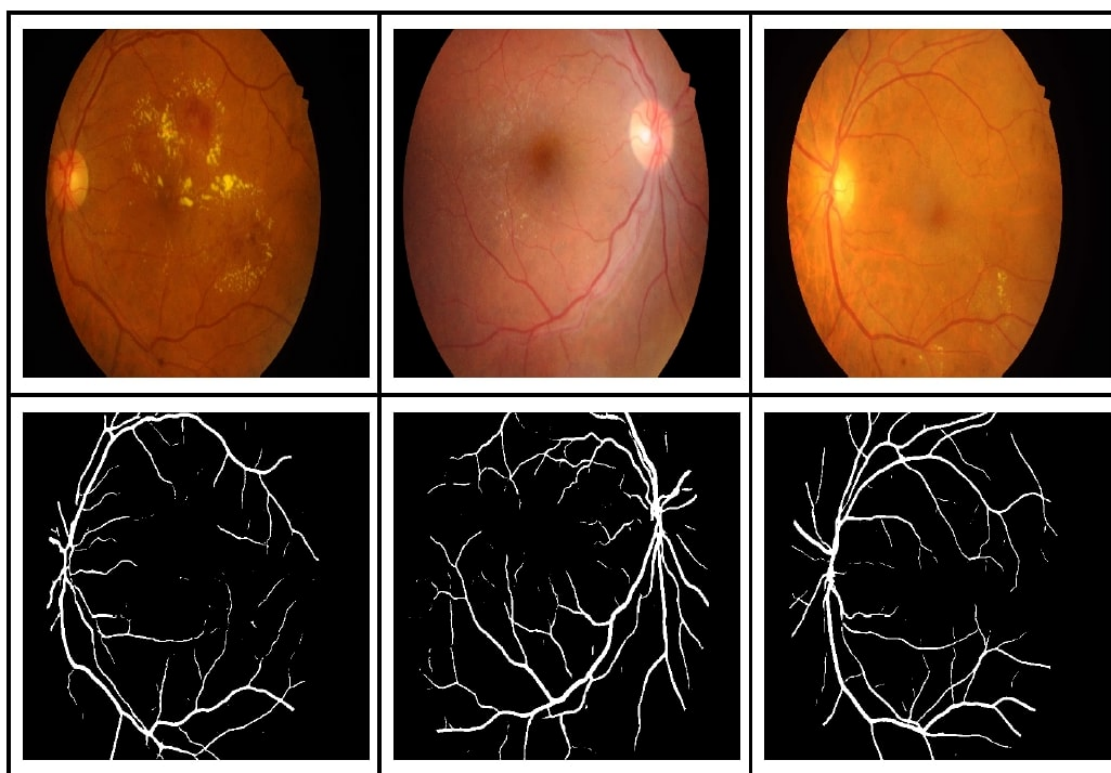


Figure 11. BV segmentation results on the IDRiD dataset using the model trained with images selected from DRIVE, HRF and CHASE.DB.

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 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

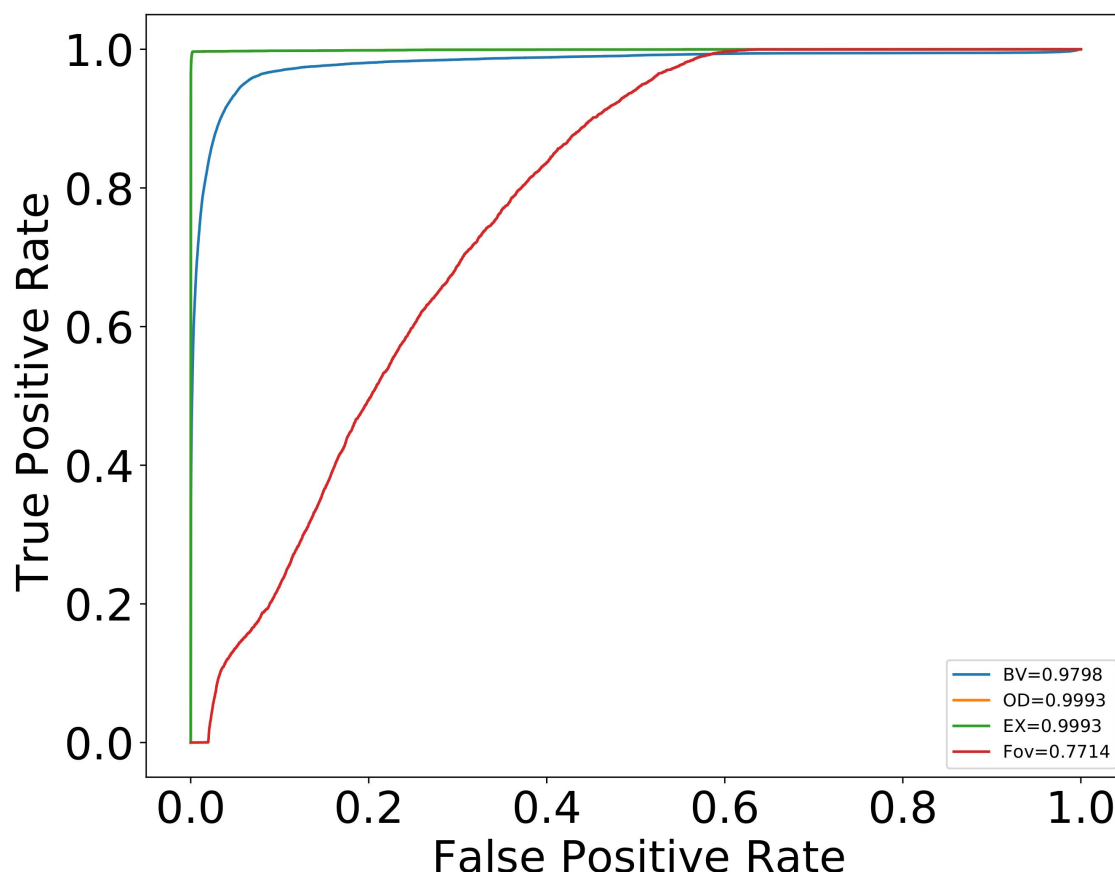


Figure 12. ROC curve for all three structures BV, OD and Macula and the pathological indicator exudates. BV results are on the DRIVE test dataset. OD, Macula (FOV) and exudates (EX) are on the IDRiD dataset.

Table 3. Comparison of exudates segmentation results of our approach with other state-of-the-art approaches on 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

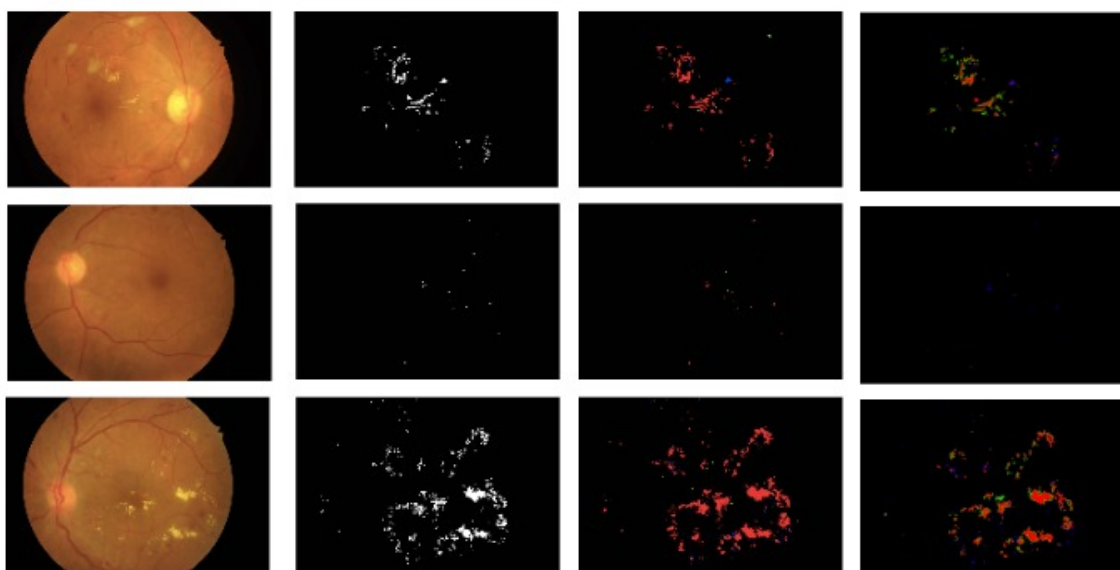


Figure 13. Comparison of exudates segmentation result of the proposed approach with Guo et al. (2020). Each column from left shows the original image, ground truth segmentation results of Guo et al. (2020) and segmentation result of proposed approach. Color code used is kept the same as in Guo et al.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

248 reported state-of-the-art and AUC is close to the state-of-the-art. Dice scores on all three dataset are lower
 249 than patch based state-of-the-art techniques. However, since we train on whole images, both training and
 250 prediction time is about 20 times faster than patch based deep learning approaches.

5 CONCLUSION

251 In this work, we illustrate the efficacy of modified multi-tasking U-Net for segmenting Blood Vessels,
 252 Optic Disc, Macula and exudates in fundal images. The proposed approach resulted in a peak increase of
 253 15% in Dice score for exudates segmentation compared to the individual segmentation result with the same
 254 architecture. Using the proposed approach, we are able to get a state-of-the-art accuracy of 95.89% on
 255 DRIVE test images which is 0.7% greater than one of the recently reported results. With our proposed
 256 method, the prediction times on the images are significantly lesser (12 times) compared to most other
 257 deep learning methods. In addition to this increased speed in prediction, we could improve the AUC and
 258 Accuracy for exudates from 0.9801 to 0.9993 and 99.33% to 99.42% on the IDRiD test dataset.

CONFLICT OF INTEREST STATEMENT

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

AUTHOR CONTRIBUTIONS

261 Sunil Kumar Vengalil and Bharath K contributed in conceptualization, investigation and to develop the
 262 methodology. Bharath K carried out the implementation and ran all the experiments mentioned in this work.
 263 Both Bharath K and Sunil Kumar Vengalil contributed equally in writing and formatting the manuscript.

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

264 Neelam Sinha supervised the work and was involved in writing-reviewing-editing. Neelam Sinha validated
265 the work and provided the resources.

FUNDING

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

DATA AVAILABILITY STATEMENT

269 1. Digital Retinal Images for Vessel Extraction Dataset (DRIVE) dataset is available at
270 <http://www.isi.uu.nl/Research/Databases/DRIVE/>

- 271 2. High-Resolution Fundus (HRF) Image Database
272 <https://www5.cs.fau.de/research/data/fundus-images/>
- 273 3. CHASE_DB1 Retinal Image Database [https://www.idiap.ch/software/bob/docs/](https://www.idiap.ch/software/bob/docs/bob/bob.db.chasedb1/master/index.html)
274 [bob/bob.db.chasedb1/master/index.html](https://www.idiap.ch/software/bob/docs/bob/bob.db.chasedb1/master/index.html)
- 275 4. Indian Diabetic Retinopathy Image Dataset (IDRiD) [https://idrid.grand-challenge.](https://idrid.grand-challenge.org/)
276 [org/](https://idrid.grand-challenge.org/)

REFERENCES

- 277 Adapa, D., Joseph Raj, A. N., Alisetti, S. N., Zhuang, Z., and Naik, G. (2020). A supervised blood vessel
278 segmentation technique for digital fundus images using zernike moment based features. *Plos one* 15,
279 e0229831
- 280 Aslan, M. F. and et al., C. (2018). Segmentation of retinal blood vessel using gabor filter and extreme
281 learning machines. In *2018 International Conference on Artificial Intelligence and Data Processing*
282 *(IDAP)* (IEEE), 1–5
- 283 Budai, A., Bock, R., Maier, A., Hornegger, J., and Michelson, G. (2013). Robust vessel segmentation in
284 fundus images. *International journal of biomedical imaging* 2013
- 285 Buslaev, A. and et al., I. (2020). Albumentations: fast and flexible image augmentations. *Information* 11,
286 125
- 287 Chen, L.-C. and et al., P. (2017). Deeplab: Semantic image segmentation with deep convolutional nets,
288 atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine*
289 *intelligence* 40, 834–848
- 290 Dash, J., Parida, P., and Bhoi, N. (2020). Retinal blood vessel extraction from fundus images using
291 enhancement filtering and clustering. *ELCVIA: electronic letters on computer vision and image analysis*
292 19, 0038–52
- 293 Feng, Z., Yang, J., Yao, L., Qiao, Y., Yu, Q., and Xu, X. (2017). Deep retinal image segmentation: a fcn-
294 based architecture with short and long skip connections for retinal image segmentation. In *International*
295 *conference on neural information processing* (Springer), 713–722
- 296 Guo, S., Wang, K., Kang, H., Liu, T., Gao, Y., and Li, T. (2020). Bin loss for hard exudates segmentation
297 in fundus images. *Neurocomputing* 392, 314–324
- 298 Hamad, H., Dwickat, T., Tegolo, D., and Valenti, C. (2020). Exudates as landmarks identified through fcm
299 clustering in retinal images. *Applied Sciences* 11, 142
- 300 Hassan, G. and et al., E.-B. (2015). Retinal blood vessel segmentation approach based on mathematical
301 morphology. *Procedia Computer Science* 65, 612–622
- 302 Hu, F. B., Satija, A., and Manson, J. E. (2015). Curbing the diabetes pandemic: the need for global policy
303 solutions. *Jama* 313, 2319–2320
- 304 Jiang, Z., Zhang, H., Wang, Y., and Ko, S.-B. (2018). Retinal blood vessel segmentation using fully
305 convolutional network with transfer learning. *Computerized Medical Imaging and Graphics* 68, 1–15
- 306 Joshua, A. O. and et al., N. (2020). Blood vessel segmentation from fundus images using modified u-net
307 convolutional neural network. *Journal of Image and Graphics* 8, 21–25
- 308 Kaur, J. and Kaur, P. (2022). Uniconv: An enhanced u-net based inceptionv3 convolutional model for
309 dr semantic segmentation in retinal fundus images. *Concurrency and Computation: Practice and*
310 *Experience* , e7138
- 311 Kou, C., Li, W., Yu, Z., and Yuan, L. (2020). An enhanced residual u-net for microaneurysms and exudates
312 segmentation in fundus images. *IEEE Access* 8, 185514–185525

- 313 Krestanova, A., Kubicek, J., and Penhaker, M. (2020). Recent techniques and trends for retinal blood
314 vessel extraction and tortuosity evaluation: A comprehensive review. *Ieee Access* 8, 197787–197816
- 315 Krizhevsky, A. and et al., S. (2012). Imagenet classification with deep convolutional neural networks.
316 *Advances in neural information processing systems* 25, 1097–1105
- 317 LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., et al. (1989). Handwritten digit
318 recognition with a back-propagation network. *Advances in neural information processing systems* 2
- 319 Liu, Z. (2021). Construction and verification of color fundus image retinal vessels segmentation algorithm
320 under bp neural network. *The Journal of Supercomputing* 77, 7171–7183
- 321 Nur, N. and Tjandrasa, H. (2018). Exudate segmentation in retinal images of diabetic retinopathy using
322 saliency method based on region. In *Journal of Physics: Conference Series* (IOP Publishing), vol. 1108,
323 012110
- 324 Park, K.-B. and et al., C. (2020). M-gan: Retinal blood vessel segmentation by balancing losses through
325 stacked deep fully convolutional networks. *IEEE Access* 8, 146308–146322
- 326 Perdomo, O., Arevalo, J., and González, F. A. (2017). Convolutional network to detect exudates in eye
327 fundus images of diabetic subjects. In *12th International Symposium on Medical Information Processing
328 and Analysis* (SPIE), vol. 10160, 235–240
- 329 [Dataset] Porwal, P. and et al., P. (2018). Indian diabetic retinopathy image dataset (idrid). doi:10.21227/
330 H25W98
- 331 Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical
332 image segmentation. In *International Conference on Medical image computing and computer-assisted
333 intervention* (Springer), 234–241
- 334 Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image
335 recognition. *arXiv preprint arXiv:1409.1556*
- 336 Singh, D. and et al., S. (2014). A new morphology based approach for blood vessel segmentation in retinal
337 images. In *2014 annual IEEE India conference (INDICON)* (IEEE), 1–6
- 338 Sun, X., Fang, H., Yang, Y., Zhu, D., Wang, L., Liu, J., et al. (2021). Robust retinal vessel segmentation
339 from a data augmentation perspective. In *International Workshop on Ophthalmic Medical Image Analysis*
340 (Springer), 189–198. Doi:https://doi.org/10.1007/978-3-030-87000-3_20
- 341 Tan, J. H., Fujita, H., Sivaprasad, S., Bhandary, S. V., Rao, A. K., Chua, K. C., et al. (2017). Automated
342 segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network.
343 *Information sciences* 420, 66–76
- 344 Vengalil, S. K., Sinha, N., Kruthiventi, S. S., and Babu, R. V. (2016). Customizing cnns for blood
345 vessel segmentation from fundus images. In *2016 International Conference on Signal Processing and
346 Communications (SPCOM)* (IEEE), 1–4
- 347 Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., et al. (2018). Understanding convolution for
348 semantic segmentation. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*.
349 1451–1460. doi:10.1109/WACV.2018.00163
- 350 [Dataset] Wikipedia (2022). https://en.wikipedia.org/wiki/Macula_of_retina
- 351 Xu, S., Chen, Z., Cao, W., Zhang, F., and Tao, B. (2021). Retinal vessel segmentation algorithm based on
352 residual convolution neural network. *Frontiers in Bioengineering and Biotechnology* 9
- 353 Yavuz, Z. and Köse, C. (2011). Retinal blood vessel segmentation using gabor filter and top-hat transform.
354 In *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU)* (IEEE),
355 546–549

- 356 Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., and Sang, N. (2018). Bisenet: Bilateral segmentation network
357 for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision*
358 (*ECCV*)
- 359 Zhang, B. and et al., Z. (2010). Retinal vessel extraction by matched filter with first-order derivative of
360 gaussian. *Computers in biology and medicine* 40, 438–445
- 361 Zheng, R., Liu, L., Zhang, S., Zheng, C., Bunyak, F., Xu, R., et al. (2018). Detection of exudates in fundus
362 photographs with imbalanced learning using conditional generative adversarial network. *Biomedical*
363 *optics express* 9, 4863–4878
- 364 Zhuang, J. (2018). Laddernet: Multi-path networks based on u-net for medical image segmentation. *arXiv*
365 *preprint arXiv:1810.07810*