SIMULTANEOUS SEGMENTATION OF MULTIPLE STRUCTURES IN FUNDAL IMAGES USING MULTI-TASKING DEEP NEURAL NETWORKS

Name of author

Address - Line 1
Address - Line 2
Address - Line 3

ABSTRACT

Fundal imaging is the most commonly used non-invasive technique for early detection of many retinal diseases like diabetic retinopathy. An initial step in automatic processing of fundal images for detecting diseases is to identify the various landmark regions like optic disc, blood vessels and fovea. In addition to these, various abnormalities like exudates that help in pathological analysis are also visible in fundal images. In this work, we propose a multi-tasking deep learning architecture for segmenting optic disc, blood vessels, fovea and exudates simultaneously. Our experimental results on publicly available datasets show that simultaneous segmentation of all these structures results in significant improvement in the performance. For segmentation performance on blood vessels we got Dice score of 80.31%, 78.11% and % on the datasets DRIVE, HRF, and CHASE_DB respectively. On Exudates, we got a Dice score of 65% on IDRiD dataset when trained in combination with Optic Disc (OD) and Blood Vessels using multi-tasking loss function, whereas the Dice score when trained individually is 50%. To the best of our knowledge, we are the first one to evaluate the effectiveness of multi-task learning for segmenting multiple structures in fundal images. We obtained a state of the art Dice score of 79.76% for blood vessel segmentation on 20 DRIVE test images which is 3.76% higher than one of the recently reported studies.

Index Terms— Fundal Image Segmentation, Multi-task learning, Deep Learning

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, which is the most common cause for blindness, and many other cardiovascular diseases. Blood vessels, Optic Disc (OD) and Macula are the major structures visible in a fundal image. However, manual identification and demarcation of fine structures like blood vessels take a lot of time and effort. Hence automatic detection of major landmarks in fundal image has become an active research area.

Figure 1 shows a fundal image with various structures like blood vessels, optic disc and fovea marked. The optic disc 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 blood vessels enter the eye.Since there are no photo sensors (rods and cones) present in the optic disc, it corresponds to a blind spot in the retina.Fovea is a small region with a lot of cone cells packed together and hence this region is responsible for sharp vision. Blood vessels that carry blood to the eye are spread across the entire region of the retina and vary in thickness and density.

Figure 2 shows a fundal image with Exudates and corresponding ground truth. Exudates are the fluid (pus) leaking out of blood vessels. It is found mostly in Diabetic Retinopathy patients. Exudates appear as white patches in the fundal image.

Since the break-through success of deep learning in solving tasks in domains like computer vision for classification [1] [2] and segmentation [3], many deep learning architectures have been tried for segmenting important structures, such as optic disc and blood vessels, in fundal images [4] [5] [6] [7]. One of the challenges of using deep learning architecture for medical images is the lack of annotated training data. Many approaches, like taking multiple training patches from a single image [4] and transfer learning, where a model trained on a dataset such as the Imagenet [8] is fine tuned for the task at hand, are proposed and found to be successful.

In this work, we propose a multi-tasking deep learning architecture for simultaneous detection of blood vessels, optic disc, fovea and exudates. Our results show that a single network that predicts multiple structures performs much better compared to detecting each structure independently using different networks as the single network can make use of the correlation between the two tasks. This correlation is evident from Figure 1, where one can see that near and inside the optic disc blood vessels are thicker and denser. So the knowledge of the optic disc can help the prediction task for blood vessels and vice versa. We perform experiments and report results on

three popular datasets DRIVE [**?**], HRF [9] and IDRID [10].

The contributions of our work are:

1. Propose a multi-tasking model for simultaneous segmentation of blood vessels, optic discs, fovea and exudates.
2. Propose a method for data augmentation of fundal images which enables the training and prediction on whole images, rather than using image patches. Using the entire image provides more contextual information that can help the segmentation task.
3. We report a state of the art accuracy of 79.76% on DRIVE test images which is 3.76% higher than the results reported by Adapa et. al. [11].
4. We could improve the F1- score on DRIVE test dataset to 80.31% by optimizing the dimension of latent representation and number of channels in the bottleneck layer of U-Net architecture.

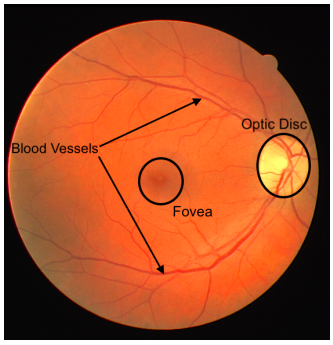


Fig. 1. Sample fundal image showing important structures Blood Vessels, Optic Disc and Macula

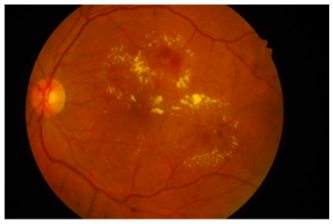


Fig. 2. Sample fundal image with Exudates

2. RELATED WORK

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 [12] [13] [14] and morphological operations based [15][16]. Image processing based techniques have the advantage that they don't need the ground truth images, but their performance is far behind the deep learning based approaches.Further, these algorithms are based on many tunable thresholds that vary from dataset to dataset and they work based on some assumptions like gaussian distribution.

Several deep learning architectures that were successful in segmentation tasks [3] in natural images were tried for segmenting blood vessels in retinal images and the results were significantly better than using conventional image processing techniques. In [4] a state-of-the-art segmentation model deeplab [3], which is pre-trained on natural images for semantic segmentation, was used to segment blood vessels at pixel level.Jiang et.al. proposes [6] a pre-trained fully convolutional network for segmenting blood vessels and report accuracy of cross-dataset test on four different datasets.In M-GAN [7], introduced by Kyeong et.al., a multi-kernel pooling block added between stacked convolutional layers supports scale-invariance which is a highly desirable feature for blood vessel 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 [17] model was introduced specifically for medical image segmentation which has multiple skip connections. In their recent work, Joshua et.al. [18] used a modified version of U-NET for segmenting blood vessels in retinal images and reported state of the art accuracies. In Laddernet, introduced by Zhuang et.al. in 2018 [5], is a sequence of multiple U-Nets cascaded together.

3. PROPOSED METHOD

3.1. CNN architecture

A modified version of the UNET architecture [17], shown in Fig 3, is used for segmenting various structures. The modification are:

1. The proposed network retains the original dimensions of the input image, whereas the original U-Net mentioned in [17] reduces the original input size from 572×572 to 388×388.
2. We used de-convolutional layers with stride for upsampling as opposed to upsampling layers used in original U-Net architecture. Our method is preferred as we learned the interpolation weights using a deconvolutional layer.
3. We added 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 blood vessel and optic disc features are mutually inclusive. These features share common connections in the fundal image, and hence, their simultaneous segmentation also yields the best results.

For experiments on individual prediction 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, an additional channel for predicting the optic disc in combination with another structure is added at the output layer. We chose to combine optic disc with other structures as optic disc annotation was available for most of the datasets.

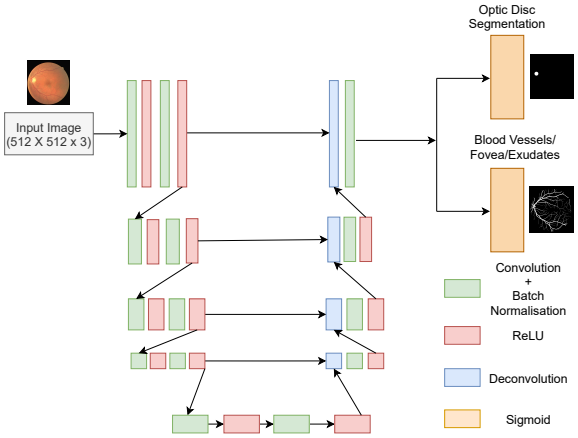


Fig. 3. Architecture of the proposed multi-tasking U-Net model

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

3.2. Dataset

We used the DRIVE [**?**], HRF [9], CHASE_DB[**?**] and IDRiD[**?**], 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 blood vessel 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 blood vessel segmentation.

IDRID [10] localization dataset, which contains 413 training images and 103 test images along with fovea ground truth, was used for fovea localization. For exudates segmentation, we used IDRID [10] segmentation dataset which contains a total of 81 images.

Data-augmentation - horizontal and vertical flipping, grid and elastic distortion provided by the library Albumentations [19], was used to increase the number of training samples by a factor of 4.

3.3. Experiments performed

We used 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 fovea. We performed multiple experiments, for individual and simultaneous prediction of blood vessels, optic disc, fovea and exudates.

Firstly, we performed blood vessel 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 utilized simultaneous segmentation of blood vessels and optic disc. Apart from these experiments, we also modified 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 blood vessel segmentation using training set containing images from HRF, DRIVE and CHASE datasets and we evaluated the performance of this trained model on 1) on a held out validation dataset which comprises images from HRF, DRIVE and CHASE 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 blood vessel results on the IDRiD dataset is further utilized to improve the optic disc, fovea 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 were trained independently one for each of the structures: blood vessel, optic disc, fovea and exudates.

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

1. Blood vessel and optic disc
2. Fovea and optic disc

3. Exudates and optic disc

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

4. RESULTS AND DISCUSSION

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 individual training. In all these cases, Optic Disc is added as an auxiliary task along with other main tasks like Blood Vessels, Macula and Exudates. The row for optic disc shows comparison of segmentation performance of optic disc when trained with optic disc alone Vs results of multi-tasking loss with blood vessels. 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 blood vessel and optic disc segmentation with an improved Dice score of 6% (for HRF dataset) and 6% (for IDRiD dataset) for Optic Disc and Blood Vessels respectively. For Macula, the segmentation results decreased by 2% upon addition of Optic Disc as an additional task during training. This happens because the fovea and optic disc are mutually exclusive as they appear at different locations in the image.

However, when trained with blood vessels, Macula and OD together we observed a 3% increase in the score for Macula. Similarly, when multi-tasking was done with three tasks (Optic Disc, Exudates and Blood Vessels segmentation) the score for Exudates segmentation also increased by 15% to a value of 65%.

Figure **??** shows sample results for blood vessel segmentation task for HRF, DRIVE and CHASE_DB when the model is trained onlly for that task and Figure **??** shows the results with multi-tasking.

The increase in Dice score is due to less number of false positives in the prediction(resulting in larger precision).

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 optic disc can easily be confused as exudates. In simultaneous segmentation, the network learns to discriminate optic disc and exudates which improves the segmentation results for both.

Figure 5 shows sample segmentation output for blood vessels, optic disc, fovea and exudates with multi-tasking loss function.

Figure **??** shows the plot of validation Dice score Vs latent representation dimension and Fig **??** shows Dice score as a function of number of channels in the bottleneck layer. As evident from the figure....

Figure shows the results of blood vessel segementtion on IDRiD dataset using the model trained using images from DRIVE, HRF and CHASE_DB datasets.

5. CONCLUSION

In this work, we illustrate the effectiveness of multi-tasking deep learning models for segmenting blood vessels, optic disc, fovea and exudates in fundal images. We build a multi-tasking model, and corresponding loss function, based on the famous U-Net architecture for simultaneous segmentation of multiple structures. Using the proposed approach we report a Dice score of 78.5%, 94%, 66% and 60% on segmentation of blood vessels, optic disc, fovea and exudates respectively. The proposed approach resulted in a peak increase of 10% in Dice score for exudates segmentation compared to the individual segmentation result with the same architecture. Using the proposed approach we were able to get a state of the art accuracy of 78% on DRIVE test images which is 2% greater than one of the recently reported results.

6. COMPLIANCE WITH ETHICAL STANDARDS

No ethical approval was required for this study as we used only publicly available datasets.

7. REFERENCES

[1] Alex Krizhevsky and Sutskever et al., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[2] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[3] Liang-Chieh Chen and Papandreou et al., "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[4] Sunil Kumar Vengalil, Neelam Sinha, Srinivas SS Kruthiventi, and R Venkatesh Babu, "Customizing cnns for blood vessel segmentation from fundus images," in *2016 International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2016, pp. 1–4.

[5] Juntang Zhuang, "Laddernet: Multi-path networks based on u-net for medical image segmentation," *arXiv preprint arXiv:1810.07810*, 2018.

[6] Zhexin Jiang, Hao Zhang, Yi Wang, and Seok-Bum Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *Computerized Medical Imaging and Graphics*, vol. 68, pp. 1–15, 2018.

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

		Individual		**Multi-tasking**	
	Dataset	***Dice(%)***	***IoU(%)***	***Dice(%)***	***IoU(%)***
Blood Vessel	DRIVE	77	62.78	80.31	67.35
	HRF	78.11	64.29		
	CHASE_DB				
Optic Disc	DRIVE	76.24	66.15	78.63	69.85
	IDRID	88(%)	81	94	90
Fovea	IDRID	70	61	66	57
Exudates	IDRID	50	34	60	45

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, Optic Dic and Blood Vessels.

Experiments Run	**Dice(%)**
Exudates Alone	50
Exudates and OD	53
Exudates and Blood Vessels	61
Exudates, OD and Blood Vessels	**65**

[7] Kyeong-Beom Park and Choi et al., "M-gan: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks," *IEEE Access*, vol. 8, pp. 146308–146322, 2020.

[8] Jia Deng and Dong et al., "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[9] Attila Budai, Rüdiger Bock, Andreas Maier, Joachim Hornegger, and Georg Michelson, "Robust vessel segmentation in fundus images," *International journal of biomedical imaging*, vol. 2013, 2013.

[10] Prasanna Porwal and Pachade et al., "Indian diabetic retinopathy image dataset (idrid)," 2018.

[11] Dharmateja Adapa, Alex Noel Joseph Raj, Sai Nikhil Alisetti, Zhemin Zhuang, and Ganesh Naik, "A supervised blood vessel segmentation technique for digital fundus images using zernike moment based features," *Plos one*, vol. 15, no. 3, pp. e0229831, 2020.

[12] Bob Zhang and Zhang et al., "Retinal vessel extraction by matched filter with first-order derivative of gaussian," *Computers in biology and medicine*, vol. 40, no. 4, pp. 438–445, 2010.

[13] Zafer Yavuz and Cemal Köse, "Retinal blood vessel segmentation using gabor filter and top-hat transform," in *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2011, pp. 546–549.

[14] Muhammet Fatih Aslan and Ceylan et al., "Segmentation of retinal blood vessel using gabor filter and extreme learning machines," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*. IEEE, 2018, pp. 1–5.

[15] Gehad Hassan and El-Bendary et al., "Retinal blood vessel segmentation approach based on mathematical morphology," *Procedia Computer Science*, vol. 65, pp. 612–622, 2015.

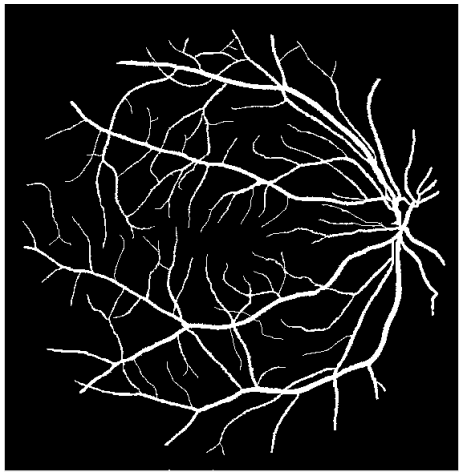
[16] Dalwinder Singh and Singh et al., "A new morphology based approach for blood vessel segmentation in retinal images," in *2014 annual IEEE India conference (INDICON)*. IEEE, 2014, pp. 1–6.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

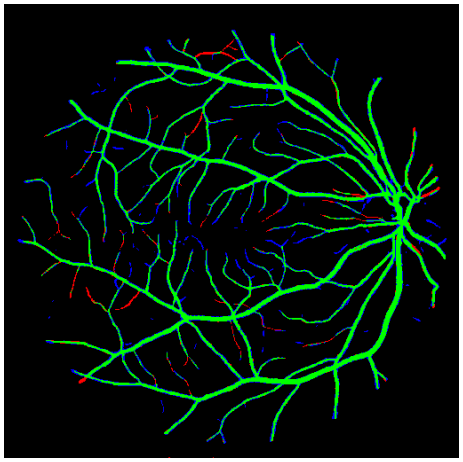
[18] Afolabi O Joshua and Nelwamondo et al., "Blood vessel segmentation from fundus images using modified u-net convolutional neural network," *Journal of Image and Graphics*, vol. 8, no. 1, pp. 21–25, 2020.

[19] Alexander Buslaev and Iglovikov et al., "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, pp. 125, 2020.

(a) Blood Vessels



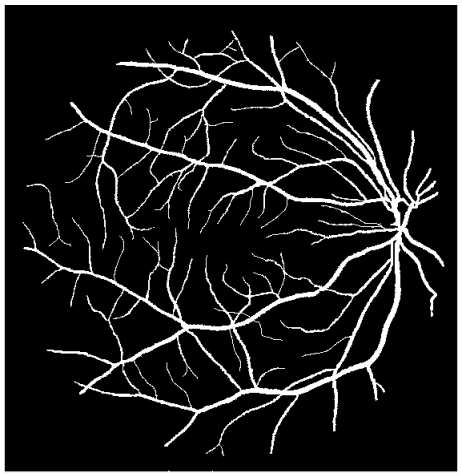
(b) Optic Disc



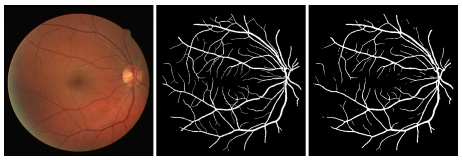
(c) Exudates

Fig. 4. Sample Blood Vessel segmentation results for DRIVE, HRF and CHASE_DB datasets with 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. F1-score is 86.01% with a precision of 84.63% and recall of 87.44%

(a) Blood Vessels



(b) Optic Disc



(c) Exudates

Fig. 5. Sample Blood Vessel segmentation results for DRIVE, HRF and CHASE_DB datasets 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. The F-score is 70.83% with a precision of 56.40% and recall of 91.12%