2019

# Hyperspectral Image Segmentation of Retinal Vasculature, Optic Disc and Macula

## Garifullin, A

IEEE

# Hyperspectral Image Segmentation of Retinal Vasculature, Optic Disc and Macula

Azat Garifullin*, Peeter Kööbi†§, Pasi Ylitepsa†, Kati Ådjers†, Markku Hauta-Kasari‡,
Hannu Uusitalo†§ and Lasse Lensu*

* Machine Vision and Pattern Recognition Laboratory Lappeenranta University of Technology, Finland
† SILK, Department of Ophthalmology, University of Tampere, Finland
‡ University of Eastern Finland, Finland
§ Tays Eye Center, Tampere University Hospital, Finland
Emails: {azat.garifullin, lasse.lensu}@lut.fi, {peeter.koobi, pasi.ylitepsa, hannu.uusitalo}@uta.fi,
kati.adjers@pshp.fi, markku.hauta-kasari@uef.fi

*Abstract*—The most common approach for retinal imaging is the eye fundus photography which usually results in RGB images. Recent studies show that the additional spectral information provides useful features for automatic retinal image analysis. The current work extends recent research on the joint segmentation of retinal vasculature, optic disc and macula which often appears in different retinal image analysis tasks. Fully convolutional neural networks are utilized to solve the segmentation problem. It is shown that the network architectures can be effectively modified for the spectral data and the utilization of spectral information provides moderate improvements in retinal image segmentation.

## I. Introduction

Retinal diseases like diabetic retinopathy, age-related macular degeneration and glaucoma are the leading causes of blindness worldwide [1]. A diagnostic process to recognize the signs of these diseases is traditionally based on retinal RGB images. Recent developments in machine vision technologies provide various methods for automatic RGB images analysis. In [2], it has been shown that additional spectral features introduced to machine learning methods may improve the performance of lesions classification and enable new ways to analyze the retinal tissue layers. Thus, spectral retinal imaging can be treated as a useful alternative to traditional color fundus imaging. This work studies the joint segmentation of retinal vasculature, optic disc and macula for hyperspectral retinal images.

Deep convolutional neural networks is a common trend in both retinal and hyperspectral image analysis. Deep architectures similar to U-Net are extensively used for the vasculature [3], optic disc and cup segmentation [4] purposes. Tan et al. [5] studied the segmentation of optic disc, fovea and retinal vasculature using a single model trained on the DRIVE dataset [6]. All the mentioned approaches have been tested on RGB images. To the best of the authors' knowledge, this work is the first work which studies spectral retinal image segmentation using deep fully-convolutional neural networks.

One way to build deep architectures for hyperspectral image (HSI) segmentation is to combine dimensionality reduction methods and convolutional neural networks. Jiao et al. [7] proposed to use the feature fusion from VGG16 encoder [8] and principal component analysis for HSI segmentation. Yu et al. [9] showed that dimensionality reduction blocks can be trained end-to-end altogether with the convolutional neural networks. Other approaches are based on 3D convolutional neural networks [10], [11] that can effectively extract both spatial and spectral features. However, it is more difficult to scale the 3D convolutions on high-resolution images. In this paper, we followed ideas similar to Yu et al. [9] and Jiao et al. [7], and adapted SegNet [12] and dense fully-convolutional neural networks (Dense-FCNs) [13] for the spectral retinal segmentation task. These architectures are trained and evaluated by using a spectral image dataset with manual ground truth for the vasculature, optic disc and macula.

## II. Spectral retinal image dataset

Several spectral fundus imaging setups with different optical principles have been proposed. A typical hyperspectral imaging setup is an adapted fundus camera with a light source with a broadband illumination, and a spectral device for selecting a spectral band. Fält et al. [14] modified a Canon CR5-45NM fundus camera to spectral fundus camera by replacing the standard light source with a fibre optic illuminator including a halogen lamp with the illumination spectrum from 380 to 780 nm and 30 interference filters with 10 nm interval were used for the wavelength selection. As the detector, a grayscale charge-coupled device (CCD) camera with the sensor array size of $2048 \times 2048$ pixels and $2 \times 2$ binning was used.

The resulted dataset is a set of $1024 \times 1024$ images with the 30 channels where each channel corresponds to the certain wavelength. For each image in the dataset, the field-of-view (FOV) mask is provided. The FOV masks are binary images where the white areas correspond to regions of the fundus of the eye. Manual segmentation masks for the vasculature, optic disc and macula are available (Fig. 2). The dataset consists of 55 spectral retinal images acquired from patients with diabetic retinopathy: randomly selected 25 images are in the training set, and the rest are used as the testing set. Thus, the amount of data in the dataset is comparable to the amount of labelled data in DRIVE [6] and STARE [15] datasets which are typically used for blood vessel segmentation algorithms benchmarking. Examples of the mean and variation of the spectra are presented in Fig. 1.
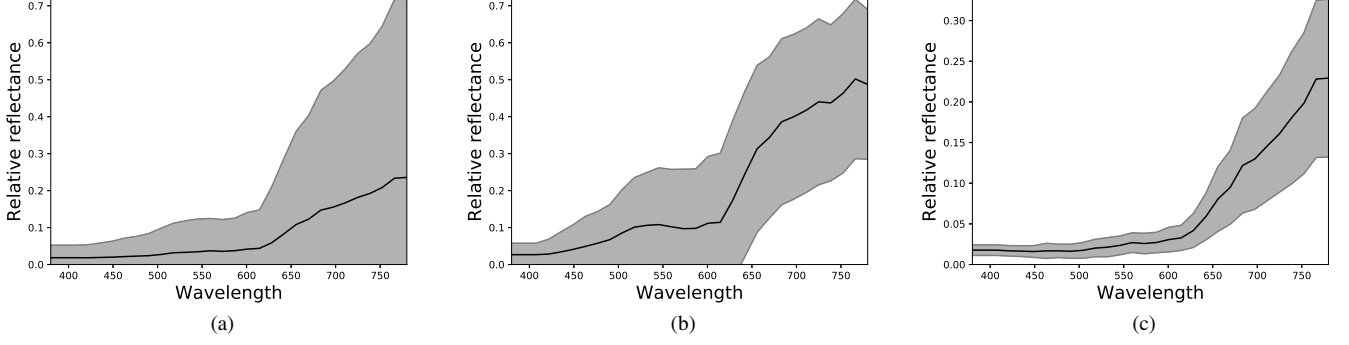
Fig. 1: Visualization of the mean spectrum and $3\sigma$ range of (a) blood vessels, (b) optic disc and (c) macula
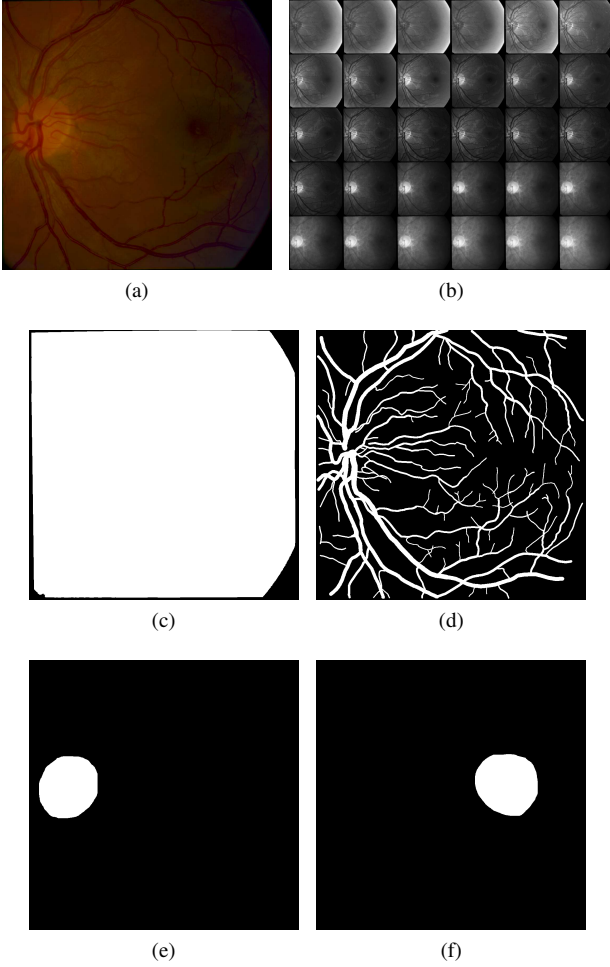


Fig. 2: (a) An RGB image from the spectral retinal image dataset. (b) Montage of the channel images composing an example hyperspectral retinal image. The image was normalized for the visualization purpose. (c) FOV mask and the corresponding segmentation masks for the (d) vessels, (e) optic disc and (f) macula.

## III. SEMANTIC SEGMENTATION ARCHITECTURES

### A. SegNet

SegNet is an encoder-decoder architecture for semantic segmentation. In this work, we used basic-SegNet with small modifications for the HSI segmentation. The modified architecture consists of a dimensionality reduction block, encoder and decoder. The dimensionality reduction block is a sequence of blocks consisting of $1 \times 1$ convolutional layer and rectified linear unit (ReLU) activation. As the encoder, VGG16 pretrained on ImageNet [16] was used. The decoder is a sequence of transposed convolutions and convolutional layers followed by batch normalization (BN), ReLU activation and dropout. The scheme of the architecture is shown in Fig. 3.
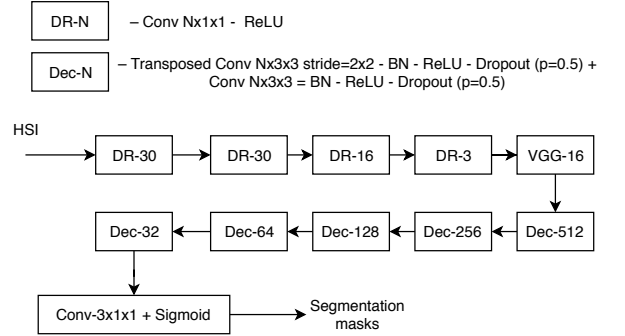


Fig. 3: DR-SegNet architecture.

### B. Dense-FCN

It has been shown that Dense-FCNs have less parameters and may outperform the SegNet architecture in a variety of different segmentation tasks [13]. Here we adapted the Dense-FCN architecture for the retinal HSI segmentation task.

The main building block of Dense-FCN is a dense convolutional block (DCB) where the input of each layer is a concatenation of the outputs of previous layers. The block consists of repeating BN, ReLU, convolution and dropout $p = 0.5$ layers resulting in $K$ feature maps (growth rate).

The main concept of Dense-FCN is similar to SegNet in a sense that the input is first compressed to a hidden

representation by the downsampling part, and then the segmentation masks are recovered by an upsampling part. The downsampling part consists of DCBs and downsampling transitions (DT) with skip connections to the upsampling part. The upsampling part consists of DCBs and upsampling transitions (UT). The scheme of the utilized architecture is given in Fig. 4.
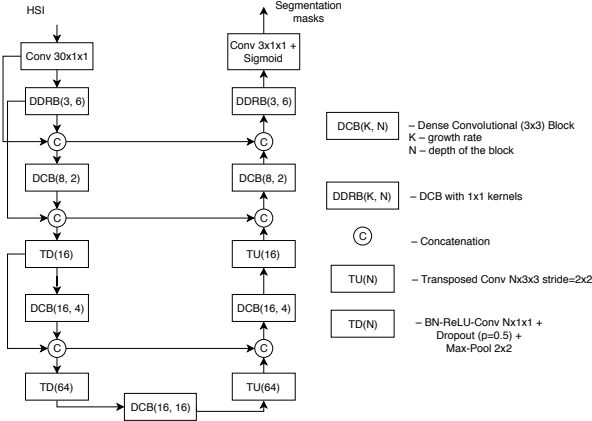


Fig. 4: DR-6-Dense-FCN architecture.

### C. Image preprocessing

First, each channel of a spectral image is normalized to values between 0 and 255. After the normalization step, contrast limited adaptive histogram equalization [17] with the clip limit of 2 and the grid size of $8 \times 8$ is applied to each channel of the input. The described preprocessing scheme was applied to both RGB and spectral images. The preprocessing scheme was used to reduce the effects of uneven illumination fields of the channel images and indirectly affect the inter-person variation in the limited dataset. The scheme was found to improve the convergence and performance of the trained models. Examples of the preprocessed RGB and spectral images are given in Fig. 5.
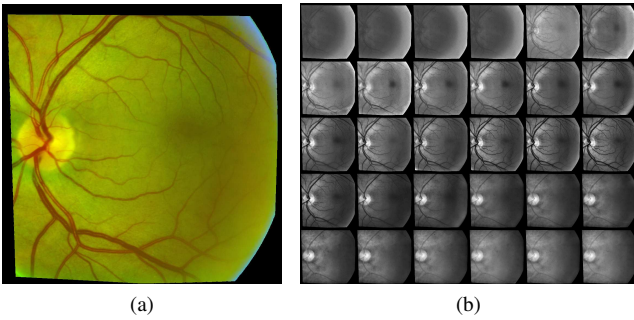


Fig. 5: Preprocessed (a) RGB and (b) spectral images.

### D. Training details

Dense-FCNs are pretrained for 50 epochs with 500 steps per epoch on random patches $512 \times 512$ with the batch size equal to 3. Each batch consists of examples for blood vessels,

optic disc, and macula. The vasculature examples are sampled uniformly, whereas the patch centers of the examples of the optic disc and macula are sampled from the normal distribution the parameters of which are estimated as the sample mean and covariance of true label coordinates.

The common training step for both architectures is fine-tuning on full size images. The number of epochs used was 50 for Dense-FCN and 100 for SegNet.

In both cases, the weights were initialized using HeNormal [18]. Binary cross-entropy was used as the loss function. In addition to dropout, $l2$ regularization with the weight decay factor $10^{-4}$ was used. As the optimizer, Adadelta [19] with the learning rate $l = 1$ and the decay rate $\rho = 0.95$ was used for the both pretraining and fine-tuning. The learning rate was dropped by a factor of 10 if the training loss was not decreased by 0.005 for 5 epochs. Data augmentation through flipping, reflecting and rescaling (with scale rates 0.8 and 1.2) was applied in both cases. The parameter values were determined empirically based on initial experiments.

## IV. EXPERIMENTS AND RESULTS

The trained networks were evaluated on the full size images from the testing set using Monte Carlo dropout [20] in the test phase with 100 forward passes. The standard F1 measure, intersection over union (IoU) for each class and mean IoU over the classes were used as the evaluation metrics and they are presented in Table I. The evaluation metrics were calculated only inside the FOV. In order to distinguish between the RGB and spectral architectures, the DR prefix was added for the architectures for spectral images and the RGB prefix for the architectures for RGB images. The Dense-FCN architecture was also tested for the hyperspectral images without the dimensionality reduction layers (HSI prefix) and with a different number of output channels of the DR layers (DR-6 means 6 output channels).

From the table, it is clear that the vessel segmentation is the easiest task for all the architectures, since F1 score is comparable in all the cases, whereas the macular region is the most difficult to segment. The latter can be explained by the fact that there are normally no clearly defined structural characteristic in the macula. Furthermore, there are numerous images where the macula is only partly visible, and these images happen to be present only in the test set, because of what all the considered models were unable to generalize to a partly visible macula. Another interesting fact is that the region may have different shapes in different spectral channels.

The VGG encoder pretrained on ImageNet allows to significantly improve the performance on the optic disc and macula segmentation tasks. In the case of Dense-FCNs, it is difficult to achieve satisfactory performance for both the optic disc and macula segmentation, and the performance depends on the number of output channels of the dimensionality reduction layers. The segmentation results for the image presented in Fig. 5 are presented in Fig. 6 – 8.

Comparing the segmentation results for hyperspectral and RGB images in Fig. 6 and Fig. 7 shows that in some cases the models trained on RGB fail to localize the macula properly, whereas the utilization of spectral information may help to

TABLE I: Evaluation results. The best F1 scores are in bold.

| Architecture | Vessels | | Optic Disc | | Macula | | Mean IoU | # Parameters |
|---|---|---|---|---|---|---|---|---|
| | F1 | IoU | F1 | IoU | F1 | IoU | | |
| DR-SegNet | 0.8091 | 0.6802 | **0.8947** | 0.8356 | **0.6566** | 0.5291 | 0.6816 | 21795786 |
| RGB-SegNet | 0.7925 | 0.6571 | 0.8802 | 0.8149 | 0.6033 | 0.4657 | 0.6458 | 21793379 |
| DR-1-Dense-FCN | 0.8125 | 0.6853 | 0.7128 | 0.6122 | 0.3962 | 0.3223 | 0.5399 | 730028 |
| DR-2-Dense-FCN | **0.8243** | 0.7019 | 0.7311 | 0.6257 | 0.3084 | 0.2414 | 0.5230 | 730383 |
| DR-3-Dense-FCN | 0.8069 | 0.6776 | 0.6986 | 0.5947 | 0.2822 | 0.2120 | 0.4948 | 730738 |
| DR-4-Dense-FCN | 0.8200 | 0.6958 | 0.6880 | 0.5867 | 0.3389 | 0.2665 | 0.5163 | 731093 |
| DR-5-Dense-FCN | 0.8006 | 0.6772 | 0.6977 | 0.5949 | 0.3843 | 0.2954 | 0.5225 | 731448 |
| DR-6-Dense-FCN | 0.8021 | 0.6714 | 0.7427 | 0.6394 | 0.4244 | 0.3402 | 0.5503 | 731803 |
| HSI-Dense-FCN | 0.7974 | 0.6647 | 0.7323 | 0.6202 | 0.2932 | 0.2282 | 0.5043 | 745837 |
| RGB-Dense-FCN | 0.8112 | 0.6840 | 0.7154 | 0.6153 | 0.1543 | 0.1070 | 0.4688 | 729043 |
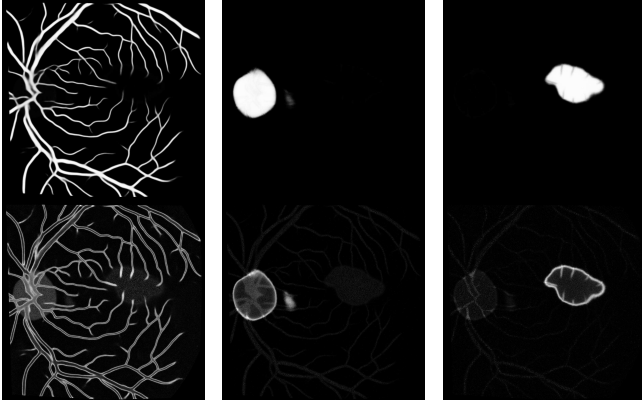


Fig. 6: Top row: example segmentation results with DR-SegNet. Bottom row: standard deviations of the activations.
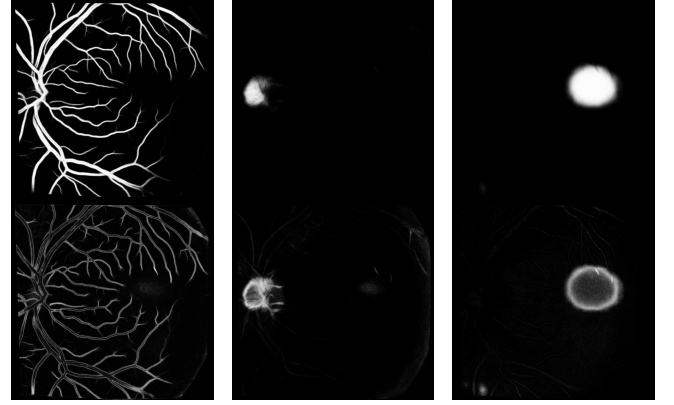


Fig. 8: Top row: example segmentation results with DR-6-Dense-FCN. Bottom row: standard deviations of the activations.
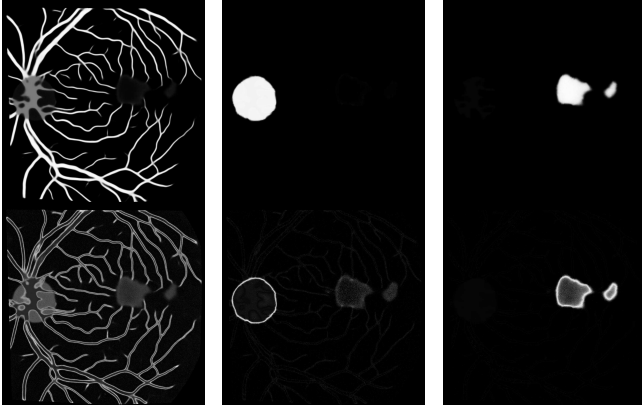


Fig. 7: Top row: example segmentation results with RGB-SegNet. Bottom row: standard deviations of the activations.

avoid such problems. It is worth to mention that in both cases the segmentation artefacts are present in areas where the labels overlap. This is mainly caused by the VGG encoder pretrained on ImageNet, since there were no signs of such artefacts when it was not used or used without pretraining. It is also clear that the results obtained with DR-SegNet show less artefacts compared to RGB-SegNet.

The dimensionality reduction layers allowed to adapt standard convolutional architecures for the hyperspectral image segmentation without significantly increasing the amount of parameters of the model. In addition, the utilization of the dimensionality reduction layers may slightly improve the performance of vessels and macula segmentation. In Fig. 9, the outputs of the dimensionality reduction layers for the both spectral architectures are shown.

In Fig. 9 one can see that the models try to emphasize areas where certain labels are most visible. For example, in the first image of the bottom row the macula is clearly seen, whereas in the second and third images the optic disc is visible more clear compared to the first image. Nevertheless, training the dimensionality reduction layers is a challenging task, and it is not always possible to train them to extract useful features. In the case of DR-SegNet, the pretrained VGG encoder makes the training easier. However, if the network is trained from the scratch, it is difficult to get results comparable to the results obtained with a pretrained model. We also tried to add skip connections to DR-SegNet in a manner similar to DR-Dense-FCNs, but in the case of DR-SegNet, it just confuses the model more and the performance decreases, whereas in the case of DR-Dense-FCNs, it boosts performance significantly.

Previously, Laaksonen [2] has shown that diabetic retinopathy lesions classification algorithms trained on spectral data
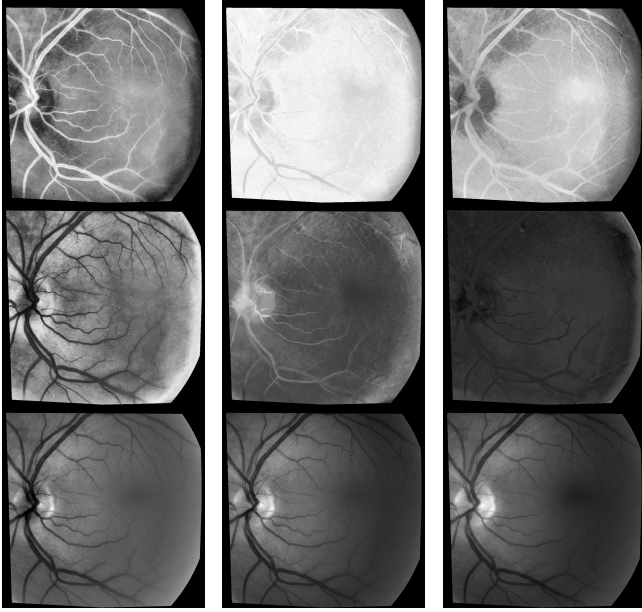
Fig. 9: Visualization of dimensionality reduction layers. The top two rows illustrate the results for DR-6-Dense-FCN and the bottom row for DR-SegNet.

outperform algorithms trained on RGB data. Furthermore, spectral information can also be utilized for the histological analysis of fundus images [2]. From the presented results, it also clear that the utilization of additional spectral information may also improve the segmentation results compared to RGB images.

## V. CONCLUSIONS

In this work, multilabel segmentation of retinal vasculature, optic disc and macula for spectral retinal images was studied. It was shown that spectral information may give additional advantages in optic disc and macula segmentation moderately improving the segmentation performance.

The results also show that it is necessary to study more the dimensionality reduction layers to find a way to train them effectively on small datasets. The future work will be concentrated on further improvements of training the Dense-FCN architecture in order to achieve comparable performance with the architectures pretrained on ImageNet. Another direction of the future work is the development of a gold standard based on label data from multiple experts.

## REFERENCES

[1] "World Health Organization: causes of blindness and visual impairment," http://www.who.int/blindness/causes/en/, accessed: 2018-06-09.

[2] L. Laaksonen, "Spectral retinal image processing and analysis for ophthalmology," Ph.D. dissertation, Lappeenranta University of Technology, 2016.

[3] L. Giancardo, K. Roberts, and Z. Zhao, "Representation learning for retinal vasculature embeddings," in *Fetal, Infant and Ophthalmic Medical Image Analysis*, M. J. Cardoso, T. Arbel, A. Melbourne, H. Bogunovic, P. Moeskops, X. Chen, E. Schwartz, M. Garvin, E. Robinson, E. Trucco, M. Ebner, Y. Xu, A. Makropoulos, A. Desjardin, and T. Vercauteren, Eds. Cham: Springer International Publishing, 2017, pp. 243–250.

[4] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, "Joint optic disc and cup segmentation using fully convolutional and adversarial networks," in *Fetal, Infant and Ophthalmic Medical Image Analysis*, M. J. Cardoso, T. Arbel, A. Melbourne, H. Bogunovic, P. Moeskops, X. Chen, E. Schwartz, M. Garvin, E. Robinson, E. Trucco, M. Ebner, Y. Xu, A. Makropoulos, A. Desjardin, and T. Vercauteren, Eds. Cham: Springer International Publishing, 2017, pp. 168–176.

[5] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, and S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network," *Journal of Computational Science*, vol. 20, pp. 70–79, 2017.

[6] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[7] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5585–5599, Oct 2017.

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

[9] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, 2017.

[10] M. He, B. Li, and H. Chen, "Multi-scale 3d deep convolutional neural network for hyperspectral image classification," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 3904–3908.

[11] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2018.

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, Dec 2017.

[13] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 1175–1183.

[14] P. Fält, J. Hiltunen, M. Hauta-Kasari, I. Sorri, V. Kalesnykiene, J. Pietilä, and H. Uusitalo, "Spectral Imaging of the Human Retina and Computationally Determined Optimal Illuminants for Diabetic Retinopathy Lesion Detection," *Journal of Imaging Science and Technology*, vol. 55, no. 3, pp. 253–263, 2011.

[15] M. Michael Goldbaum, "Structured analysis of the retina," http://www.cecas.clemson.edu/ ahoover/stare, 2003, online.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[17] K. Zuiderveld, "Graphics gems iv," P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press Professional, Inc., 1994, ch. Contrast Limited Adaptive Histogram Equalization, pp. 474–485. [Online]. Available: http://dl.acm.org/citation.cfm?id=180895.180940

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[19] M. D. Zeiler, "Adadelta: An adaptive learning rate method," *CoRR*, vol. abs/1212.5701, 2012.

[20] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016, pp. 1050–1059.