Segmentation of Pigment Signs in Fundus Images for Retinitis Pigmentosa Analysis by Using Deep Learning

Nadia Brancati[1], Maria Frucci[1], Daniel Riccio[1,2(✉)], Luigi Di Perna[3], and Francesca Simonelli[3]

[1] Institute for High Performance Computing and Networking National Research Council of Italy (ICAR-CNR), Naples, Italy
{nadia.brancati,maria.frucci,daniel.riccio}@cnr.it
[2] University di Napoli Federico II, Naples, Italy
daniel.riccio@unina.it
[3] Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, University of Campania "Luigi Vanvitelli", Naples, Italy
luigidiperna85@gmail.com,
francesca.simonelli@unicampania.it

Abstract. The adoption of Deep Learning (DL) algorithms into the practice of ophthalmology could play an important role in screening and diagnosis of eye diseases in the coming years. In particular, DL tools interpreting ocular data derived from low-cost devices, as a fundus camera, could support massive screening also in resource limited countries. This paper explores a fully automatic method supporting the diagnosis of the Retinitis Pigmentosa by means of the segmentation of pigment signs in retinal fundus images. The proposed approach relies on an U-Net based deep convolutional network. At the present, this is the first approach for pigment signs segmentation in retinal fundus images that is not dependent on hand-crafted features, but automatically learns a hierarchy of increasingly complex features directly from data.

We assess the performance by training the model on the public dataset RIPS and comparisons with the state of the art have been considered in accordance with approaches working on the same dataset. The experimental results show an improvement of 15% in F-measure score.

Keywords: Retinitis Pigmentosa · Pigment signs · Medical image analysis · Segmentation · Deep Learning

1 Introduction

Recent studies have shown that DL models are able to detect and diagnosing various retinal diseases interpreting ocular data derived from different diagnostic modalities including digital photographs, optical coherence tomography (OCT), and visual fields [1]. DL systems can already be applied to teleophthalmology programs to identify abnormal retinal images reducing the clinic workload for disease screening. Furthermore, DL tools could enable ophthalmic self-monitoring by patients via smartphone

retinal photography. Most of the initial studies have centered around automatic detection of diabetic retinopathy, age-related macular degeneration, and glaucoma [2–4, 14], while only a few methods have been developed for the automatic diagnosis of genetically heterogeneous retinal disorders. Many of these genetic eye disorders lead to blindness and an early diagnosis, even by means of a simple ophthalmoscopy, can reduce preventable vision loss. Automatic diagnostic systems are able to analyze ocular data and could be also used by non-ophthalmologists to screen patients who do not yet show signs of weakness in visual acuity. The aim of this work is to investigate a DL system for segmenting pigment signs (PSs) that represent a symptom of the Retinitis Pigmentosa (RP). The RP is one of the most common disease caused by genetic eye disorders and leads to night blindness and a progressive constriction of the visual field from the periphery to the center. Progression leads to central acuity loss and legal blindness in most patients. At present, no cure exists to stop the progression of the disease but, if an early diagnosis of RP is available, the progressive degeneration of RP can be delayed through the intake of vitamin A and other nutritional interventions [5]. Clinical diagnosis is possible through fundus examination revealing the presence of PSs, arteriolar attenuation, and pallor of optic disc. In Fig. 1, a healthy and a severely degenerated retina are shown.
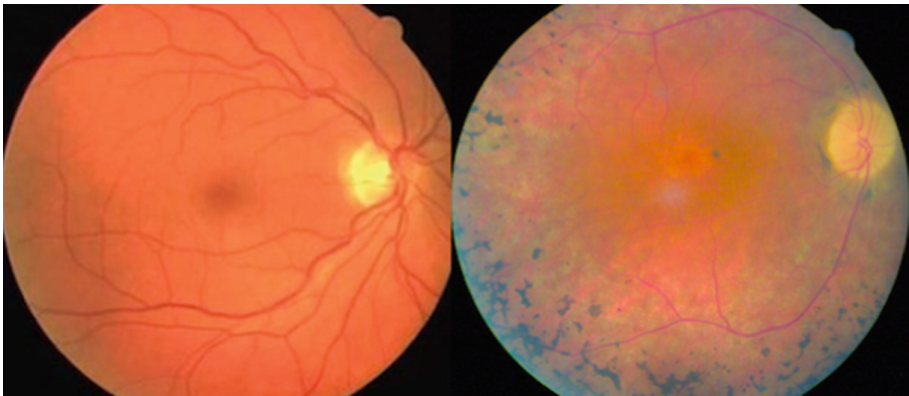

Fig. 1. Fundus images of a healthy retina (left) and a retina with Retinitis Pigmentosa (right). In the image of the diseased eye, peripheral pigment signs, attenuated retinal arterioles and optic-disc pallor are evident.

PSs are a consequence of a degeneration of the photoreceptors and accumulate over years, so they could not be present in younger individuals. However, PSs are the most easily identifiable signs on a retinal fundus image by a non-ophthalmologist and should be prompt referred. Further detailed ophthalmic examinations (visual test, OCT, electroretinography and fundus autofluorescence) are adopted to determine the severity of the disease and to monitor the disease progression [6]. Many automatic methods to quantify RP and to track its progression are based on the analysis of OCT [3, 7, 8]. The diagnosis by fundus camera represents the best solution for RP screening by

performing retinal image acquisition in resource-limited setting, since fundus images can be acquired with inexpensive devices.

Though many approaches to automatically analyze the retinal vessel structure and the pallor of the optic disk were developed [9–13], the literature about the automatic detection of PSs in fundus images is extremely limited [14–16]. In our previous work [16], we have proposed a supervised method to segment PSs in fundus images, which extract pixel-wise/region-wise hand-crafted features that are fed to machine learning techniques (i.e. Random Forests and AdaBoost.M1) to discriminate between PSs and normal fundus. Furthermore, we made publicly available a dataset of Retinal Images for Pigment Signs (RIPS) [17] for the evaluation of the performance of PSs segmentation algorithms.

DL based segmentation is a hot topic and has gained increasing attention, as deep neural networks learn a hierarchy of feature maps directly from data without requiring any hand-crafted features. Most of the early DL approaches for segmentation translate the segmentation task into a pixel-wise classification problem. However, in order to solve the image classification problems, DL models require a large number of images to be trained. Moreover, the classification of all the pixels in a test image is carried out by sliding a window on the image and classifying the current central pixel, that entails a slow prediction time. Other DL architectures specifically devoted to segmentation are based on an encoder-decoder scheme that learns to decode low-resolution images on pixel-wise predictions. In this work, the adopted DL model to segment PSs is a U-Net based convolutional neural network, that is an encoder-decoder network for pixel-wise prediction [18].

2   The Proposed Model

The proposed deep model is based on U-Net, which has been successfully used for segmenting medical images in several contests [9, 14]. This model is an encoder-decoder network implementing a contracting/expanding path consisting of convolutional, downsampling and upsampling layers.

In this work, the network has been modified with respect to its original architecture. Indeed, two of the five blocks have been dropped out (i.e. convolutions, pooling/upsampling) and the number of filters was halved. The architecture of the network is shown in Fig. 2.

In the encoding part of the network each feature map is downsampled by applying a pooling operation in order to spatially reduce the input as well as the number of parameters to be learned in the following layer. In our case, max-pooling has been adopted for all downsampling layers. Upsampling layers increase the dimension of feature maps by learning to deconvolve them. The decoder feature maps and the corresponding encoder feature maps are concatenated to produce the output. In order to stabilize the learning process and to reduce the number of training epochs, we also introduce batch normalization, while dropout (0.2) is introduced to prevent over-fitting.

In more details, both the encoder and decoder include five convolutional layers, whose filters have size 3×3, a stride of 1 and adopt the rectification non linearity (ReLU) and Batch Normalization. Moreover, dropout of 0.2 is applied alternately to

odd convolutional layers. In the contracting path, the second convolution of the first two blocks feeds a max-pooling layer that is computed on a window of size 2×2 with a stride equal to 2. In the expanding path, the first convolution of the last two blocks is preceded by an upsampling layer, which doubles the size of the feature map and concatenates it with the corresponding feature map coming from the contracting path.

At the end, a soft-max classifier computes the probability of each pixels of being a PS (foreground) or background.

Given the nature of PSs, choosing the right metric to be optimized represents a crucial aspect. Indeed, PSs represent a small percentage of the pixels of the image, that translates in very few positive pixels and a high number of true negatives in the segmented image. The most of works in literature consider the accuracy as the metric to be optimized during training. However, accuracy can be heavily contributed by a large number of true negatives, thus F1-score might be a better measure to use if one need to seek a balance between precision and recall and there is an uneven class distribution. For this reason, the F1-measure has been considered in our model. Furthermore, to increase the robustness of the training process, the Adadelta [19] optimizer has been selected, as it does not require manual tuning of the learning rate and has been shown to be robust to noisy gradient information and different model architectures.
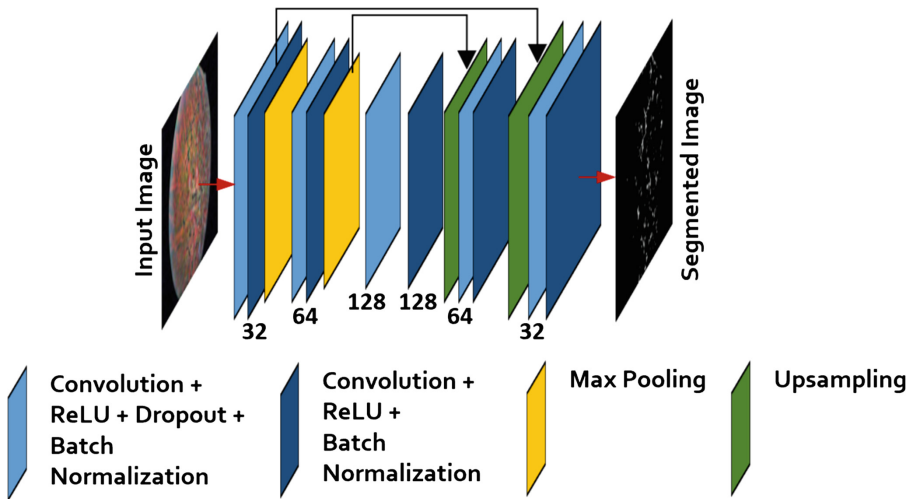

Fig. 2. Architecture of the proposed U-Net based network.

3   Experiments

3.1   Materials and Methods

The experiments have been performed on the Retinal Images for Pigment Signs dataset, namely RIPS. This dataset consists of 120 retinal fundus images with a resolution of 1440×2160 pixels captured from four patients, who underwent three different

acquisition sessions. During each session, five images per eye were acquired covering different regions of the fundus. The time lapse between two consecutive sessions is at least six months, while time interval between the first and last session always exceeds one year. Images were acquired using the digital retinal camera Canon CR4-45NM (Canon UK, Reigate, UK) and show a high variability in terms of color balancing, contrast and sharpness/focus, also for the same patient. Two binary masks are associated to each image, where the foreground representing PSs has been marked by two experts in the field of ophthalmology. Moreover, for each image, a mask image is provided to delineate the FOV.

3.2   Training Strategy and Image Prediction

The resolution of retinal images makes it unfeasible to train the existing DL architectures on the whole image. Most commonly used approaches to cope with this problem are either to reduce the image resolution or to partition the image in patches. The main drawback of a severe image downsampling is that small PSs could disappear. On the other hand, image partitioning produces a high number of patches with large size when working on high resolution images. For this reason, we adopted a compromise by reducing the image size of a factor equal to 0.5 that allows to extract patches with small size, but still representative enough. Only a subset of patches randomly extracted from each image is included in the training set. Indeed, PSs appears only in some regions of the image and could have a very small size, so considering only patches including at least one pixel marked as PS in the corresponding mask yield a training set sufficiently representative also for the background.

In the testing process, the input image is downsampled of a factor 0.5. The image prediction is performed by extracting patches according to a window sliding with a stride $s > 0$. Patches are fed to the network and pixels are assigned a probability of being PS. For values of s smaller than the window size, patch overlapping occurs, so that each pixel receives multiple predictions as belonging to several patches. The global score of a pixel is computed by summing up all predictions. Scores are normalized in the range [0, 1] and the foreground image is obtained by applying a threshold of 0.5.

3.3   Experimental Setup and Results

In the experiments, a per patient cross-validation protocol was applied considering samples from three out of the four patients for the training and the data of the fourth patient for the validation. The number of training epochs has been set to 30, while the batch size is equal to 32. To train our network, we used a NVIDIA GeForce GTX 1050 with 4 Gb of RAM.

The first experiment aims to verify that F-measure provides better performance than accuracy when used as loss function to train the network. In this experiment, the patch size is fixed to 48 × 48 pixels and the stride of the sliding window is set to 6 in the prediction process. Results are reported in Table 1.

Table 1. Performance measures of the proposed U-Net based network when F-measure or accuracy is used as loss function.

Loss function	Sensitivity		Specificity		Precision		Accuracy		F-measure	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
Accuracy	0.568	0.077	0.994	0.005	0.450	0.084	0.990	0.006	0.496	0.048
F-measure	0.557	0.035	0.994	0.005	0.480	0.086	0.990	0.006	0.506	0.035

The values in Table 1 show that the accuracy is very high in both cases, while the precision increases considerably when the F-measure is used as a loss function. This is because accuracy is heavily influenced by the number of true negatives, so it approaches to 1.0 even when the precision is very low. The experimental results confirm that the F-measure outperforms the accuracy for the PSs segmentation task.

In the second experiment we have analyzed the improvements obtained in terms of f-measure when the patch size increases. The proposed U-Net based architecture has been tested for three different patch sizes, namely 48×48, 72×72, and 96×96 and numerical results are reported in Table 2.

Table 2. Performance measures of the proposed model for different patch sizes.

Patch size	Sensitivity		Specificity		Precision		Accuracy		F-measure	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
48×48	0.557	0.035	0.994	0.005	0.480	0.086	0.990	0.006	0.506	0.035
72×72	0.626	0.067	0.993	0.005	0.465	0.081	0.990	0.006	0.528	0.049
96×96	0.552	0.070	0.996	0.003	0.561	0.076	0.992	0.005	0.551	0.033

Results in Table 2 mainly highlight that F-measure proportionally increases with respect to the patch size. In particular, we have observed that the larger the size of the patch is, the better the model performs in discriminating PSs from blood vessels.

Figure 3 shows the segmented images produced by the proposed model when it is trained with patches of size 48×48, 72×72, and 96×96 pixels, respectively. In Fig. 4, one image for each of the four patient is shown together with the corresponding ground truth and the segmented image produced by our model when patches of 96×96 pixels are considered.

The performance of the proposed U-Net based model has been compared with state of the art approaches. In particular, the machine learning based approach proposed in [16] was considered. Numerical results are reported in Table 3.

Table 3. Performance measures of different methods.

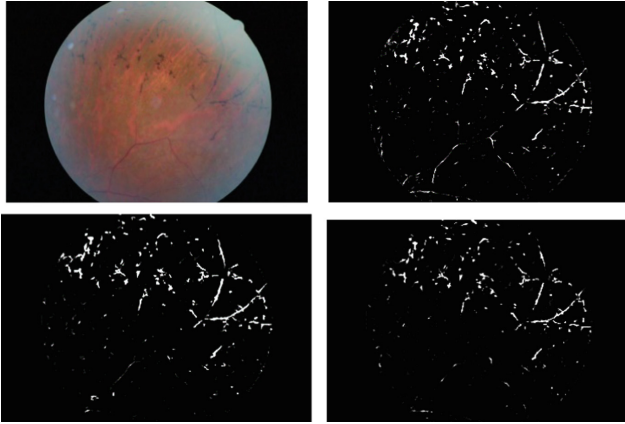| Method	Sensitivity	Specificity	Precision	Accuracy	F-measure
RF-PB [16]	0.582	0.994	0.461	0.991	0.479
AB-PB [16]	0.642	0.993	0.424	0.990	0.467
Proposed model	0.552	0.996	0.561	0.992	0.551

Fig. 3. Segmented images obtained for different patch sizes: input image (top-left), 48 × 48 pixels (top-right), 72 × 72 pixels (bottom-left), and 96 × 96 pixels (bottom-right).
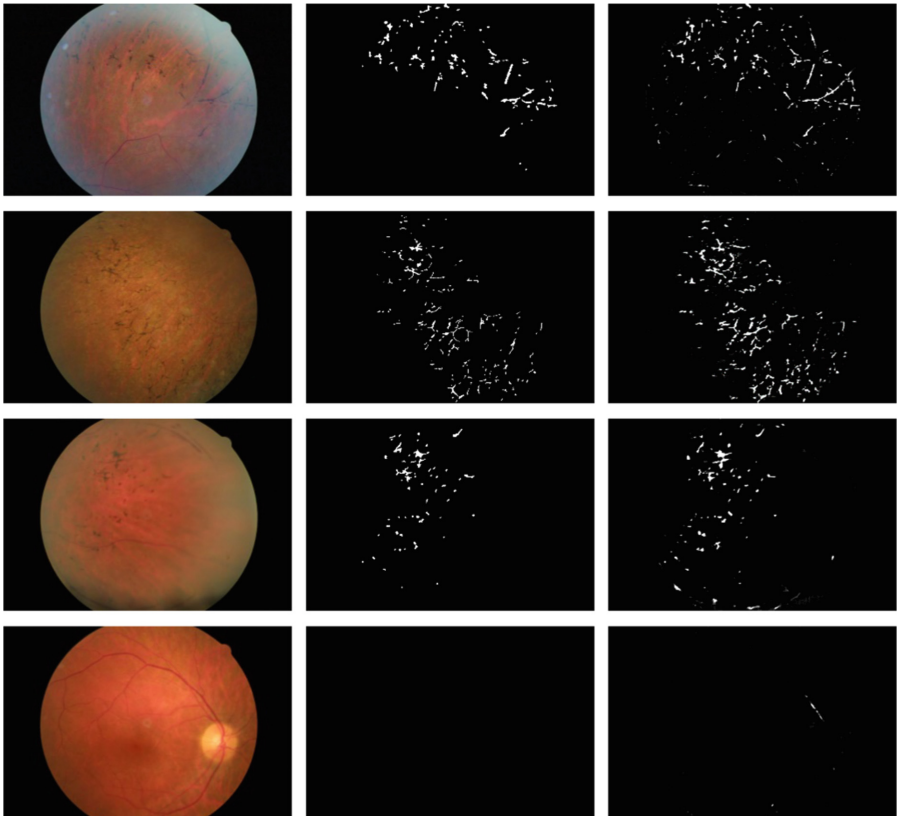

Fig. 4. Results from the RIPS dataset. Top to bottom: patients from 1 to 4. Left to right: the original image, the ground truth, the result of the proposed model.

4   Conclusions

In this study, a deep-learning based approach for segmenting PSs in retinal fundus images is presented. The segmentation has been performed in an end-to-end way by using a DL model. We have proposed a U-Net based model, since U-Net has been largely used for segmenting medical images in several contests. In particular, it has been successfully used to segment structures in retinal fundus images. We have modified the original architecture of U-Net to reduce the number of parameters, and consequently the computation time and memory requirements. The number of blocks has been reduced from five to three and the number of filters per block has been halved. The model implements a patch based strategy both for training and testing. The performance of the proposed model has been assessed on the publicly available RIPS dataset. Several experiments have been performed varying the size of the extracted patches and using different loss functions for the training phase. Experimental results show that using the F-measure in place of the accuracy improves the quality of segmentation. Moreover, the quality of the segmentation increases proportionally with the patch size. The proposed model also outperforms a pixel based machine learning method proposed in literature, as it produces an increment of 15% in terms of F-measure.

References

1. Grewal, P.S., Oloumi, F., Rubin, U., Tennant, M.T.: Deep learning in ophthalmology: a review. Can. J. Ophthalmol. **53**(4), 309–313 (2018)
2. Gargeya, R., Leng, T.: Automated identification of diabetic retinopathy using deep learning. Ophthalmology **124**(7), 962–969 (2017)
3. Meriaudeau, F.: Machine learning and deep learning approaches for retinal disease diagnosis. Procedia Comput. Sci. **135**, 2 (2018)
4. Tan, J.H., et al.: Age-related macular degeneration detection using deep convolutional neural network. Future Gener. Comput. Syst. **87**, 127–135 (2018)
5. Dias, M.F., et al.: Molecular genetics and emerging therapies for retinitis pigmentosa: basic research and clinical perspectives. Progress Retinal Eye Res. **63**, 107–131 (2018)
6. Fahim, A.: Retinitis pigmentosa recent advances and future directions in diagnosis and management. Curr. Opin. Pediatr. **30**(6), 725–733 (2018)
7. Cunefare, D., et al.: Deep learning based detection of cone photoreceptors with multimodal adaptive optics scanning light ophthalmoscope images of achromatopsia. Biomed. Opt. Express **9**(8), 3740–3756 (2018)
8. Ramachandran, R., Zhou, L., Locke, K.G., Birch, D.G., Hood, D.C.: A comparison of methods for tracking progression in x-linked retinitis pigmentosa using frequency domain oct. Translational Vis. Sci. Technol. **2**(7), 5 (2013)
9. Xiancheng, W., et al.: Retina blood vessel segmentation using a U-net based Convolutional neural network. In: Procedia Computer Science: International Conference on Data Science (ICDS 2018), Beijing, China, 8–9 June (2018)
10. Oliveira, A., Pereira, S., Silva, C.A.: Retinal vessel segmentation based on fully convolutional neural networks. Expert Syst. Appl. **112**, 229–242 (2018)

11. Brancati, N., Frucci, M., Gragnaniello, D., Riccio, D.: Retinal vessels segmentation based on a convolutional neural network. In: Mendoza, M., Velastín, S. (eds.) CIARP 2017. LNCS, vol. 10657, pp. 119–126. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75193-1_15

12. Yang, H.K., Oh, J.E., Han, S.B., Kim, K.G., Hwang, J.M.: Automatic computer-aided analysis of optic disc pallor in fundus photographs. Acta Ophthalmologica **97**, e519–e525 (2018)

13. Das, H., Saha, A., Deb, S.: An expert system to distinguish a defective eye from a normal eye. In: 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 155–158. IEEE, Ghaziabad (2014)

14. Sevastopolsky, A.: Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. Pattern Recogn. Image Anal. **27**(3), 618–624 (2017)

15. Brancati, N., Frucci, M., Gragnaniello, D., Riccio, D., Di Iorio, V., Di Perna, L.: Automatic segmentation of pigment deposits in retinal fundus images of Retinitis Pigmentosa disease. Comput. Med. Imaging Graph. **66**, 73–81 (2018)

16. Brancati, N., et al.: Learning-based approach to segment pigment signs in fundus images for Retinitis Pigmentosa analysis. Neurocomputing **308**, 159–171 (2018)

17. RIPS (2018). https://www.icar.cnr.it/sites-rips-datasetrips/

18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

19. Zeiler, M.D.: ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701 (2012)