Enhancing Retinal Blood Vessel Segmentation through Self-Supervised Pre-Training †

José Morano [1,2,*], Álvaro S. Hervella [1,2], Noelia Barreira [1,2], Jorge Novo [1,2] and José Rouco [1,2]

1   CITIC-Research Center of Information and Communication Technologies, University of A Coruña, 15071 A Coruña, Spain; a.suarezh@udc.es (Á.S.H.); noelia.barreira@udc.es (N.B.); jnovo@udc.es (J.N.); jrouco@udc.es (J.R.)
2   Department of Computer Science, University of A Coruña, 15071 A Coruña, Spain
*   Correspondence: j.morano@udc.es
†   Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Abstract: The segmentation of the retinal vasculature is fundamental in the study of many diseases. However, its manual completion is problematic, which motivates the research on automatic methods. Nowadays, these methods usually employ Fully Convolutional Networks (FCNs), whose success is highly conditioned by the network architecture and the availability of many annotated data, something infrequent in medicine. In this work, we present a novel application of self-supervised multimodal pre-training to enhance the retinal vasculature segmentation. The experiments with diverse FCN architectures demonstrate that, independently of the architecture, this pre-training allows one to overcome annotated data scarcity and leads to significantly better results with less training on the target task.

Keywords: self-supervised learning; transfer learning; multimodal; retinal vasculature segmentation

1. Introduction

Retinal vasculature segmentation represents a key step in the analysis of multiple common diseases like glaucoma and diabetes. However, its manual completion is arduous and partly subjective, so automatic methods have emerged as an advantageous alternative. State-of-the-art vasculature segmentation is based on Fully Convolutional Networks (FCNs). Nonetheless, using FCNs requires addressing two major difficulties: (1) Determining the network architecture and (2) gathering a large amount of annotated training data. The first issue can be partly overcome by reviewing similar problems. Annotated data, however, are usually scarce in medical imaging, as they require experts to be involved in a tedious process. This motivates the proposal of self-supervised multimodal pre-training (SSMP) to learn the relevant patterns from unlabeled data and reduce the required amount of annotated data [1–3]. Specifically, the proposed SSMP consists of training an FCN to predict fluorescein angiographies (a grayscale modality that enhances the vasculature) from retinographies.

In this work, we present a novel application of SSMP to enhance vasculature segmentation in a transfer learning setting, performing a comparative analysis of several FCN architectures.

2. Methodology

The main objective of this work is the segmentation of the retinal vasculature using FCNs. To enhance the results, we propose a transfer learning setting that consists of using SSMP followed by a fine-tuning in the segmentation task [4]. To appraise our proposal, we evaluated the results of the same networks using the SSMP or training from scratch and with different training set sizes (1, 5, 10, and 15). In all of the cases, we used the following FCN architectures: U-Net [5], FC-DenseNet [6], and ENet [7,8].

In order to perform the SSMP, we aligned the 59 retinography–angiography pairs of the publicly available Isfahan MISP dataset [9] using the method proposed in [10]. Then, inspired by [1,2], we used SSIM function to compute the reconstruction loss between the network output and its ground truth.

To train the networks for the vasculature segmentation task, we employed the DRIVE dataset [11], which consists of 40 retinographies and their corresponding vasculature segmentation masks. As the loss, we used Binary Cross-Entropy. For testing, we included the 20 annotated images of the STARE dataset [12].

The networks were trained using the Adam optimization algorithm with learning rate decay and data augmentation through affine transformations and color and intensity variations.

3. Results and Conclusions

Table 1 shows the best AUC-ROC and AUC-PR values of the different networks trained from scratch (FS) and using SSMP for the STARE dataset. Moreover, in Figure 1 is depicted an example of the segmentation masks predicted by the U-Net trained with 15 images, with and without SSMP. As observed, the use of SSMP has significant benefits in both quantitative and qualitative terms; mainly due to the fact that the vessel continuity is better preserved and the pathological structures are better handled. This improvement, in addition, is achieved with less training in the target task. These results demonstrate that the use of SSMP emerges as a valuable option when annotated data in the target task are scarce.

Regarding the diverse FCN architectures, both qualitative and quantitative results (see Table 1) demonstrated that the U-Net provided the best performance.

Table 1. Best AUC-ROC and AUC-PR values of the different networks trained from scratch (FS) and using self-supervised multimodal pretraining (SSMP) for the STARE dataset.

U-Net				FC-DenseNet				ENet			
SSMP		FS		SSMP		FS		SSMP		FS	
ROC	PR	ROC	PR	ROC	PR	ROC	PR	ROC	PR	ROC	PR
0.9834	0.9051	0.9728	0.8590	0.9794	0.8924	0.9699	0.8468	0.9694	0.8434	0.8349	0.4472
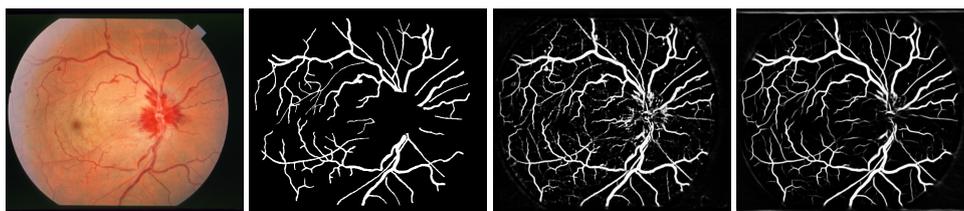
Figure 1. Example of the predicted vasculature mask. From left to right: Original STARE retinography, vasculature segmentation ground truth, vasculature segmentation mask predicted by U-Net trained with 15 images without SSMP, vasculature mask predicted by the same network with SSMP.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Hervella, Á.S.; Rouco, J.; Novo, J.; Ortega, M. Retinal Image Understanding Emerges from Self-Supervised Multimodal Reconstruction. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI, Granada, Spain, 16–20 September 2018; Volume 11070, pp. 321–328.

2. Álvaro S. Hervella.; Rouco, J.; Novo, J.; Ortega, M. Self-supervised multimodal reconstruction of retinal images over paired datasets. *Expert Syst. Appl.* **2020**, *161*, 113674.

3. Hervella, Á.S.; Rouco, J.; Novo, J.; Ortega, M. Learning the retinal anatomy from scarce annotated data using self-supervised multimodal reconstruction. *Appl. Soft Comput.* **2020**, *91*, 106210. doi:10.1016/j.asoc.2020.106210.

4. Morano, J.; Hervella, Á.S.; Barreira, N.; Novo, J.; Rouco, J. Multimodal Transfer Learning-based Approaches for Retinal Vascular Segmentation. In Proceedings of the European Conference on Artificial Intelligence (ECAI), Santiago de Compostela, Spain, 29 August–5 September 2020.

5. Ronneberger, O.; P.Fischer.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; Volume 9351, pp. 234–241.

6. Jégou, S.; Drozdzal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1175–1183. doi:10.1109/CVPRW.2017.156.

7. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. *CoRR* **2016**, *abs/1606.02147*.

8. Canziani, A.; Culurciello, E.; Paszke, A. Evaluation of neural network architectures for embedded systems. In Proceedings of the 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, USA, 28–31 May 2017; pp. 1–4. doi:10.1109/ISCAS.2017.8050276.

9. Kashefpur, M.; Kafieh, R.; Jorjandi, S.; Golmohammadi, H.; Khodabande, Z.; Abbasi, M.; Fakharzadeh, A.A.; Kashefpoor, M.; Rabbani, H. Isfahan MISP Dataset. *J. Med. Signals Sens.* **2016**, *7*, 43–48.

10. Hervella, Á.S.; Rouco, J.; Novo, J.; Ortega, M. Multimodal registration of retinal images using domain-specific landmarks and vessel enhancement. *Procedia Comput. Sci.* **2018**, *126*, 97 – 104.

11. Staal, J.; Abramoff, M.D.; Niemeijer, M.; Viergever, M.A.; van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. doi:10.1109/TMI.2004.825627.

12. Hoover, A.D.; Kouznetsova, V.; Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **2000**, *19*, 203–210. doi:10.1109/42.845178.