Joint Optic Disc and Cup Segmentation Using Self-Supervised Multimodal Reconstruction Pre-Training †

Álvaro S. Hervella [1,2,*], Lucía Ramos [1,2], José Rouco [1,2], Jorge Novo [1,2] and Marcos Ortega [1,2]

[1]   Centro de Investigación CITIC, Universidade da Coruña, 15071 A Coruña, Spain; l.ramos@udc.es (L.R.);
      jrouco@udc.es (J.R); jnovo@udc.es (J.N.); mortega@udc.es (M.O.)
[2]   VARPA Research Group, Instituto de Investigación Biomédica de A Coruña (INIBIC),
      Universidade da Coruña, 15006 A Coruña, Spain
*   Correspondence: a.suarezh@udc.es; Tel.: +34-981-167-000
†   Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Abstract: The analysis of the optic disc and cup in retinal images is important for the early diagnosis of glaucoma. In order to improve the joint segmentation of these relevant retinal structures, we propose a novel approach applying the self-supervised multimodal reconstruction of retinal images as pre-training for deep neural networks. The proposed approach is evaluated on different public datasets. The obtained results indicate that the self-supervised multimodal reconstruction pre-training improves the performance of the segmentation. Thus, the proposed approach presents a great potential for also improving the interpretable diagnosis of glaucoma.

Keywords: deep learning; self-supervised learning; segmentation; eye fundus; glaucoma

1. Introduction

The detailed analysis of the optic disc and optic cup in retinal images is a key step for the diagnosis of glaucoma. In this regard, several biomarkers derived from the morphological analysis of these two structures have demonstrated to be useful for the diagnosis and screening of the disease. This has motivated the development of automated methods for the segmentation of optic disc and cup in retinography. Nowadays, the most successful segmentation approaches are those based on deep learning techniques. However, the training of deep neural networks requires large amounts of annotated data that can be difficult to obtain.

In this work, we present a novel approach for the joint segmentation of the optic disc and cup in retinography using deep learning [1]. Given the limited size of common datasets, we propose a novel self-supervised pre-training consisting in a multimodal reconstruction between complementary retinal image modalities. To validate this proposal, we perform experiments on different public datasets.

2. Methodology

The proposed pre-training consists of the self-supervised multimodal reconstruction of fluorescein angiography from retinography. This multimodal reconstruction leverages the availability of unlabeled multimodal image pairs for learning about the retinal anatomy [2]. In particular, the self-supervised multimodal reconstruction is trained as proposed in Reference [3] using a public dataset of retinography-angiography pairs. After the pre-training phase, the neural network is fine-tuned in the target task, that is, the joint segmentation of the optic disc and cup in retinography. This joint segmentation is approached as a pixel-wise multi-class classification where the neural network learns to predict the likelihood of the different classes [1].

3. Results and Conclusions

The proposed methodology is evaluated on two different public datasets, which present a representative variety of glaucomatous an healthy retinas. Additionally, we also perform a comparison against training the network from scratch in the segmentation task, which is the most common approach in the literature. The quantitative evaluation shows that the proposed methodology significantly improves the segmentation of the optic disc and optic cup. In particular, our proposal achieves a Jaccard index of 82.29 and 92.43 for the optic cup and disc, respectively, whereas the training from scratch achieves a Jaccard index of 75.36 and 88.19 for the optic cup and disc, respectively. Additionally, in comparison to other alternatives in the literature, our proposal achieves a competitive state-of-the-art performance while using less annotated data. Figure 1 depicts a representative example of predicted segmentations, where it is observed that the proposed approach produces more consistent segmentations than the alternative training from scratch. Conclusively, the self-supervised multimodal reconstruction pre-training demonstrates to be useful for improving the joint segmentation of the optic disc and cup in retinography.
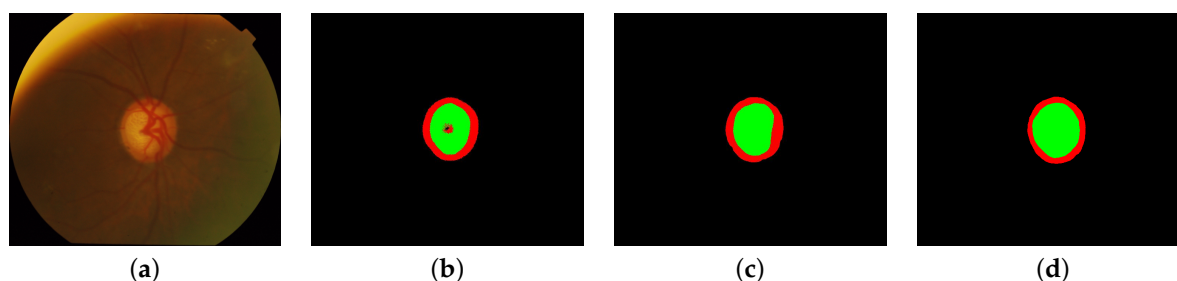

(**a**)	(**b**)	(**c**)	(**d**)

Figure 1. Representative example of predicted segmentations for a glaucomatous retina. (**a**) Retinography, (**b**,**c**) predicted segmentations, and (**d**) ground truth segmentation. The predicted segmentations correspond to (b) a neural network trained from scratch and (c) the proposed approach.

References

1. Hervella, Á.S.; Ramos, L.; Rouco, J.; Novo, J.; Ortega, M. Multi-Modal Self-Supervised Pre-Training for Joint Optic Disc and Cup Segmentation in Eye Fundus Images. In Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 961–965, doi:10.1109/ICASSP40776.2020.9053551.
2. 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.
3. Hervella, Á.S.; Rouco, J.; Novo, J.; Ortega, M. Self-supervised multimodal reconstruction of retinal images over paired datasets. *Expert Syst. Appl.* **2020**, *161*, 113674, doi:10.1016/j.eswa.2020.113674.