Kalpa Publications in Computing

Volume XXX, 2022, Pages 51–53

Proceedings of V XoveTIC Conference. XoveTIC 2022

Multi-task Convolutional Neural Networks for the End-to-end Simultaneous Segmentation and Screening of the Epiretinal Membrane in OCT Images*

Mateo Gende[1,2], Joaquim de Moura[1,2], Jorge Novo[1,2], Manuel G. Penedo[1,2], and Marcos Ortega[1,2]

[1] Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC) Universidade da Coruña, 15006 A Coruña, Spain

[2] Centro de Investigación CITIC Universidade da Coruña, 15071 A Coruña, Spain

Abstract

The Epiretinal Membrane (ERM) is an ocular pathology that causes visual distortion. In order to detect and treat the ERM, ophthalmologists visually inspect Optical Coherence Tomography (OCT) images.This is a costly and subjective process. In this work, we present three different fully automatic, end-to-end approaches that make use of multi-task learning to simultaneously screen for and segment ERM symptoms in OCT images. These approaches were implemented into three architectures that capitalise on the way the models share a single architecture for the two complementary tasks.

1 Introduction

The Epiretinal Membrane (ERM) is an ocular condition characterised by the apparition of a layer of scar tissue over the retina. This layer can develop in the region known as the Inner Limiting Membrane (ILM), exerting a traction over the photosensitive tissue. If not detected and treated early, this traction may cause permanent deformations [1]. Ophthalmologists typically diagnose the ERM by visually inspecting Optical Coherence Tomography (OCT) images, a process that can prove to be costly and tiresome, leading to subjectivity in the diagnosis. This motivates the need for Computer-aided Diagnosis (CAD) systems that can aid clinicians in the detection and treatment of this relevant pathology. Initially, CAD methods for the ERM detection were semi-automatic, with more recent studies introducing automatised methods that

*This research was funded by Instituto de Salud Carlos III, Government of Spain, [DTS18/00136]; Ministerio de Ciencia e Innovación y Universidades, Government of Spain, [RTI2018-095894-B-I00]; Ministerio de Ciencia e Innovación, Government of Spain through the research project [PID2019-108435RB-I00]; Consellería de Cultura, Educación e Universidade, Xunta de Galicia, Grupos de Referencia Competitiva, [ED431C 2020/24], predoctoral grant [ED481A 2021/161]; Axencia Galega de Innovación (GAIN), Xunta de Galicia, [IN845D 2020/38]; CITIC, Centro de Investigación de Galicia [ED431G 2019/01], receives financial support from Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, through the ERDF (80%) and Secretaría Xeral de Universidades (20%).

make use of Convolutional Neural Networks (CNNs) to classify between healthy and pathological OCT images [2]. Other works have focused on characterising the appearance of the ERM in these images. [3, 4] This led to the development of methodologies for the ERM segmentation using classical machine learning techniques [5, 6] and CNNs [7, 8], with the latter outperforming the classical methods. However, these approaches work by employing a series of bespoke, purpose-specific steps in order to convert the segmentation problem into a classification of image fragments extracted by a sliding window. This reliance on a series of steps limits the reliability of systems that employ these methodologies, since an error at any of the stages can severely impact the performance of the system, or even cause it to fail. In this work, we present three approaches for the end-to-end simultaneous segmentation and screening of the ERM in OCT images [9]. These were implemented into three modular CNN architectures that leverage this multi-task context in different ways, by exploiting inter-task complementarity in order to guide the model training. Furthermore, since these models work in an end-to-end manner, they can contribute to a more robust and reliable system overall.

2 Materials and Methods

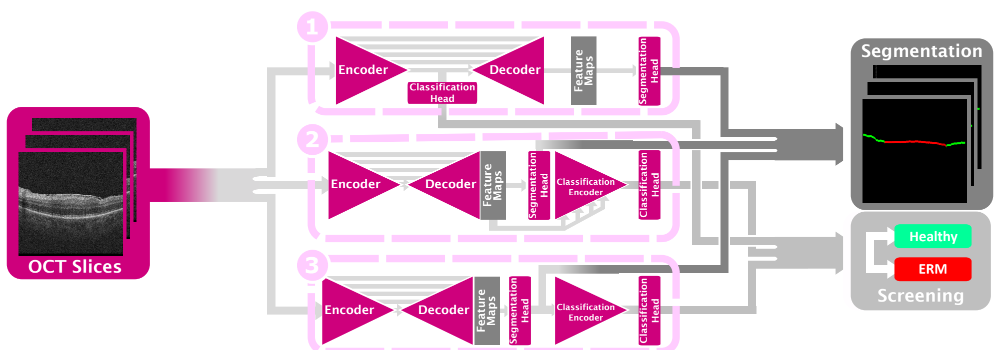

Figure 1: Summary of the three proposed architectures.

The three architectures herein presented take three different approaches to multi-task learning. These architectures share a basic structure based on the Feature Pyramid Network model [10] combined with a classification head in order to provide both outputs simultaneously. The first of these approaches takes advantage of the innermost, most-refined visual features extracted by the segmentation encoder to obtain the screening output (Fig. 1.1). The second approach makes all the visual feature maps employed in the segmentation process available to the classification head. To facilitate the processing of these feature maps, an additional classification-only encoder is added before the fully-connected layers. This way, the shared architecture takes on an extensive approach to multi-task learning (Fig. 1.2). The third and final approach takes a different focus by restricting the shared features to that which is common to both tasks. Here, the output segmentation maps are fed to the classification encoder, instead of the complete set of visual feature maps, helping both tasks to guide each other more closely (Fig. 1.3). Each of these approaches was tested by using three state-of-the-art encoder models commonly employed in similar domains in the current literature: Densely Connected Convolutional Networks, Residual Neural Networks and Google Inception Networks.

3 Results and Conclusions

Table 1: Results achieved by the three proposed approaches with each of the three architectures.

		First approach: inner features			Second approach: complete feature set			Third approach: segmentation maps		
		DenseNet-121	Resnet-18	Inception-v4	DenseNet-121	Resnet-18	Inception-v4	DenseNet-121	Resnet-18	Inception-v4
Segmentation	Sensitivity	0.659	0.768	0.703	0.643	0.674	0.725	0.704	0.673	0.454
	Specificity	0.965	0.945	0.910	0.943	0.960	0.920	0.937	0.965	0.933
Classification	Sensitivity	0.835	0.816	0.950	0.852	0.838	0.911	0.888	0.855	0.832
	Specificity	0.929	0.983	0.561	0.965	0.795	0.787	0.934	0.917	0.930

The proposed approaches achieved satisfactory results (Table 1). For the segmentation task, the first approach combined with a Resnet-18 attained the best balance between sensitivity and specificity, while the second and third approaches combined with a DenseNet-121 encoder produced the overall best results for the classification task. Summarising, the results highlight the advantages that may be gained from multi-task learning in the diagnosis of ocular diseases.

References

[1] Min Hee Suh, Jong Mo Seo, Kyu Hyung Park, and Hyeong Gon Yu. Associations between macular findings by optical coherence tomography and visual outcomes after epiretinal membrane removal. *American Journal of Ophthalmology*, 147(3):473–480.e3, March 2009.

[2] Esther Parra-Mora, Alex Cazañas-Gordon, Rui Proença, and Luís A. da Silva Cruz. Epiretinal membrane detection in optical coherence tomography retinal images using deep learning. *IEEE Access*, 9:99201–99219, 2021.

[3] Sergio Baamonde, Joaquim de Moura, Jorge Novo, José Rouco, and Marcos Ortega. Feature definition and selection for epiretinal membrane characterization in optical coherence tomography images. In *Image Analysis and Processing - ICIAP 2017*, pages 456–466. Springer International Publishing, 2017.

[4] Sergio Baamonde, Joaquim de Moura, Jorge Novo, and Marcos Ortega. Automatic detection of epiretinal membrane in oct images by means of local luminosity patterns. In Ignacio Rojas, Gonzalo Joya, and Andreu Catala, editors, *Advances in Computational Intelligence*, pages 222–235, Cham, 2017. Springer International Publishing.

[5] Sergio Baamonde, Joaquim de Moura, Jorge Novo, Pablo Charlón, and Marcos Ortega. Automatic identification and characterization of the epiretinal membrane in OCT images. *Biomedical Optics Express*, 10(8):4018, July 2019.

[6] Sergio Baamonde, Joaquim de Moura, Jorge Novo, Pablo Charlón, and Marcos Ortega. Automatic identification and intuitive map representation of the epiretinal membrane presence in 3d OCT volumes. *Sensors*, 19(23):5269, November 2019.

[7] Mateo Gende, Joaquim De Moura, Jorge Novo, Pablo Charlón, and Marcos Ortega. Automatic segmentation and intuitive visualisation of the epiretinal membrane in 3d oct images using deep convolutional approaches. *IEEE Access*, 9:75993–76004, 2021.

[8] Mateo Gende, Joaquim de Moura, Jorge Novo, Pablo Charlón, and Marcos Ortega. Automatic segmentation and visualisation of the epirretinal membrane in OCT scans using densely connected convolutional networks. In *The 4th XoveTIC Conference*, Basel Switzerland, September 2021. MDPI.

[9] Mateo Gende, Joaquim de Moura, Jorge Novo, and Marcos Ortega. End-to-end multi-task learning approaches for the joint epiretinal membrane segmentation and screening in OCT images. *Comput. Med. Imaging Graph.*, 98(102068):102068, June 2022.

[10] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017.