Kalpa Publications in Computing
Volume 14, 2023, Pages 127–129
Proceedings of V XoveTIC Conference. XoveTIC 2022

Deep learning-based detection of domain-specific retinal landmarks for color fundus image registration*

David Rivas-Villar[1,2], Álvaro S. Hervella[1,2], José Rouco[1,2], and Jorge Novo[1,2]

[1] Centro de investigación CITIC, Universidade da Coruña, A Coruña 15071, Spain
[2] Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruna, A Coruña 15006, Spain
{david.rivas.villar, a.suarezh, jrouco, jnovo} @udc.es

Abstract

Retinal imaging is widely used to diagnose and monitor eye-specific pathologies as well as some systemic diseases. Registering retinal images is crucial to compare the relevant structures within the eye. In this work, we propose a deep learning-based method to register color fundus images using domain-specific landmarks. We employ a deep neural network to detect bifurcations and crossovers of the retinal arterio-venous vessel tree. Then, these keypoints are directly matched using RANSAC. Our method was tested using the public FIRE dataset obtaining a registration score of 0.657, which is comparable to the best state of the art methods although our proposal is comparably much faster. Furthermore, our method improves the results of the state of the art deep learning methods.

1 Introduction

Image registration consists in aligning a pair of images whose content is partly coincident but were captured under different conditions. In retinal imaging, registration allows to compare key structures and monitor disease progression. Moreover, it is a key component of computer-aided diagnosis systems, which cannot rely on the labor-intensive manual registration. Currently, the best automatic methods for retinal image registration are still classical approaches [1, 2]. Deep learning methods provide desirable advantages over classical pipelines like end-to-end training. However, they still cannot compete with classical methods in terms of results. In this work, we present a deep learning-based registration method [3]. It uses a neural network to locate domain-specific keypoints (blood vessel crossovers and bifurcations) which are then matched using RANSAC (Random Sample Consensus) to create a transformation that aligns the images.

*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 2021/147]; postdoctoral grant [ED481B-2022-025]; 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%).

2 Materials and Methods

The presented method is divided in two separate steps, keypoint detection and keypoint matching. The goal of the keypoint detection is to accurately locate the crossovers and bifurcations of the blood vessel tree. To do so we employ an U-Net. The marked class imbalance between the background and the keypoints makes unsuitable the direct prediction of the keypoint location from the binary ground truth. Instead, we follow the approach in [5], which trains the network employing heatmaps created from the binary ground truth. The heatmaps have peak values located in the same position as the binary keypoints, with said values decreasing progressively in the surrounding pixels. These heatmaps are created convolving the binary ground truth. For this convolution, we test two separate kernels, a Gaussian kernel and a Radial Hyperbolic Tangent kernel (Radial Tanh) [5]. The network is trained using the Mean Squared Error between the predicted and the ground truth heatmap as the loss function. In order to distinguish between the two types of keypoints, the network creates two separate heatmaps. However, using just two output channels heavily penalizes misidentifying a landmark in the wrong category. Therefore, we add a third channel, containing both types of keypoints, which encourages the detection of any landmark regardless of their type [5]. From these predicted heatmaps, the specific location of each keypoint can be recovered using a local maxima filter. Additionally, we use an intensity threshold to prevent background noise being mistakenly detected as keypoints. After the keypoint detection, the points are directly matched via RANSAC to infer a transformation capable of aligning the images. Due to the low number of highly-specific detected keypoints, we do not need to compute and match descriptors. In addition, our approach only matches each of the points with those of the same type, decreasing the execution time.

Currently, there is no retinal image dataset with both keypoints and registration ground truth. Therefore, we train and test our keypoint detector using the DRIVE dataset which has a crossovers and bifurcations ground truth. Additionally, we test the whole registration pipeline using the FIRE dataset which has registration ground truth. FIRE is itself divided in three categories, S with high overlapping, P with low overlapping and A, with high overlapping and pathology progression. Both datasets are very different in terms of resolutions. Therefore, we need to scale the keypoints from the DRIVE dataset resolution, in which they are detected, to the FIRE resolution to properly evaluate the registration methodology. We test two separate methods, the simpler approach consists on upscaling the point locations (point scaling). The second alternative consists in upscaling the predicted heatmap and then calculate the local maxima to find the keypoint locations in the FIRE resolution (heatmap scaling).

We evaluate our work using the registration score, a standard metric proposed in conjunction with the FIRE dataset. Moreover, to enable comparison with other works [4], we also evaluate our work using Root Mean Square Error (RMSE) between the registration control points of the FIRE dataset at a lower resolution (256×256 pixels).

3 Results and Conclusions

The proposed keypoint detector network obtained 81.22% of precision and 70.80% of recall in the test set of DRIVE using the Gaussian kernel. Similarly, the Tanh kernel obtained 79.63% of precision and 71.58% of recall in the same set. The results for the experiments regarding the registration performance of the scaling methods are shown in Table 1. All the approaches are accurate and produce very similar results to one another.

We compare our proposals with the best classical state of the art works in Table 1. Our results are comparable to those of the best methods in categories A and S while in category P

Name	Registration Score (AUC)				Execution Time
	S	P	A	FIRE	(in seconds)
VOTUS [2]	0.934	**0.672**	0.681	0.812	106*
REMPE [1]	**0.958**	0.542	0.660	0.773	198†
Point scaling Gauss (Ours)	0.900	0.298	0.662	0.655	0.65
Point scaling Tanh (Ours)	0.904	0.296	0.666	0.656	0.65
Heatmap scaling Gauss (Ours)	0.908	0.293	0.660	0.657	0.65
Heatmap scaling Tanh (Ours)	0.905	0.288	*0.658*	0.654	0.65

Table 1: Results for the best state of the art methods and our proposals. Best results highlighted in bold. * Indicates execution time extracted from [2] and † from [1].

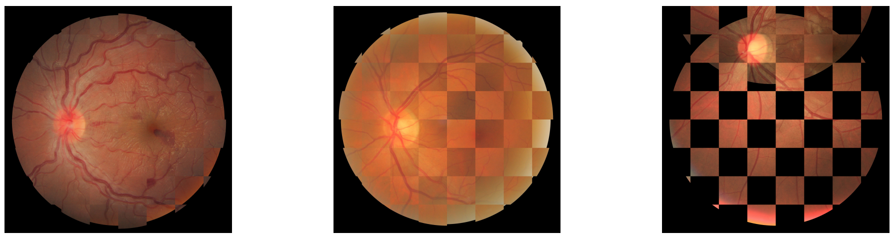

Figure 1: Registered images from FIRE dataset, from left to right: Category A, S and P [3]

the results are worse than the state of the art. Our method is the fastest by orders of magnitude, as evidenced by the execution times shown in Table 1. These times are only indicative, as they were not obtained with the same hardware, but they illustrate the clear difference in execution times. Finally, we also compared our approach to the only other deep learning method [4]. We improve its results as our method successfully registers images from the category P and improves the combined RMSE for categories A and S, obtaining 0.299 while [4] obtains 0.915.

Overall, we can conclude that our approach is satisfactory. It produces similar results to the top classical approaches in the state of the art, while being much faster. Furthermore, our proposal also improves the results of the other deep learning-based approach.

References

[1] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros. Rempe: Registration of retinal images through eye modelling and pose estimation. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3362–3373, 2020. https://doi.org/10.1109/JBHI.2020.2984483.

[2] D. Motta, W. Casaca, and A. Paiva. Vessel optimal transport for automated alignment of retinal fundus images. *IEEE Transactions on Image Processing*, 28(12):6154–6168, 2019. https://doi.org/10.1109/TIP.2019.2925287.

[3] David Rivas-Villar, Álvaro S. Hervella, José Rouco, and Jorge Novo. Color fundus image registration using a learning-based domain-specific landmark detection methodology. *Computers in Biology and Medicine*, 140:105101, 2022. https://doi.org/10.1016/j.compbiomed.2021.105101.

[4] Beiji Zou, Zhiyou He, Rongchang Zhao, Chengzhang Zhu, Wangmin Liao, and Shuo Li. Non-rigid retinal image registration using an unsupervised structure-driven regression network. *Neurocomputing*, 404:14–25, 2020. https://doi.org/10.1016/j.neucom.2020.04.122.

[5] Álvaro S. Hervella, José Rouco, Jorge Novo, Manuel G. Penedo, and Marcos Ortega. Deep multi-instance heatmap regression for the detection of retinal vessel crossings and bifurcations in eye fundus images. *Computer Methods and Programs in Biomedicine*, 186:105201, 2020. https://doi.org/10.1016/j.cmpb.2019.105201.