Unsupervised deformable image registration in a landmark scarcity scenario: choroid OCTA *

Emilio López-Varela[1,2][0000−0002−0764−1164], Jorge Novo[1,2][0000−0002−0125−3064], José Ignacio Fernández-Vigo[3,4][0000−0001−8745−3464], Francisco Javier Moreno-Morillo[3][0000−0001−7748−0506], and Marcos Ortega[1,2][0000−0002−2798−0788]

[1] VARPA Group, Biomedical Research Institute of A Coruña (INIBIC), University of A Coruña, A Coruña (Spain)
[2] CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña (Spain)
[3] Department of Ophthalmology, Hospital Clínico San Carlos. Instituto de Investigación Sanitaria (IdISSC). Madrid, Spain
[4] Centro Internacional de Oftalmología Avanzada, Madrid, Spain.
e.lopezv@udc.es, jnovo@udc.es, jfvigo@hotmail.com, javimorenomorillo@gmail.com, mortega@udc.es

Abstract. Recent advances in OCTA allow the imaging of blood flow deeper than the retinal layers at the level of the choriocapillaris (CC), where a pattern of small dark areas represents the absence of flow, called flow voids. The distribution of flow voids can be used as a biomarker to diagnose and monitor the progression of relevant pathologies or the efficacy of applied treatments. A pixel-to-pixel comparison can help to carry out this monitoring effectively, although in order to carry out this comparison, the used images must be perfectly aligned. CC images are characterized by their granularity, presenting numerous and complex local deformations, so a deformable registration is necessary to carry out a reliable comparison. However, CC OCTA images also present a characteristic absence of visually significant anatomical structures. This landmark scarcity hardens drastically the identification of points of interest to achieve an accurate registration. Based on this context, we designed a methodology to accurately perform this deformable registration in this

* This research was funded by Instituto de Salud Carlos III, Government of Spain, DTS18/00136 research project; Ministerio de Ciencia e Innovación y Universidades, Government of Spain, RTI2018-095894-B-I00 research project; Ministerio de Ciencia e Innovación, Government of Spain through the research project with reference PID2019-108435RB-I00; Consellería de Cultura, Educación e Universidade, Xunta de Galicia, Grupos de Referencia Competitiva, grant ref. ED431C 2020/24; Axencia Galega de Innovación (GAIN), Xunta de Galicia, grant ref. IN845D 2020/38; CITIC, Centro de Investigación de Galicia ref. 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%). Emilio López Varela acknowledges its support under FPI Grant Program through PID2019-108435RB-I00 project.

challenging scenario. Hence, we propose a convolutional neural network model trained by unsupervised learning to register images in a real clinical scenario, being obtained at different time instants from patients with central serous chorioretinopathy (CSC) treated with photodynamic therapy. Our methodology produces superior alignment to those achieved with other proven methods, helping to improve the monitoring of the efficacy of photodynamic therapy applied to patients with CSC. Our robust and adaptable methodology can also be exploited in other similar scenarios of complex registrations with anatomical landmark scarcity.

Keywords: Ophthalmology · OCTA imaging · choriocapillaris · Deformable image registration · Flow voids · Convolutional Neural Networks

1 Introduction

Optical coherence tomographic angiography (OCTA) is a noninvasive imaging modality characterized by its capability to show a detailed visualization of the retinal vascularity. OCTA uses the variation in signal intensity of the OCT image over time as a contrast mechanism to obtain images [10, 15]. Recent advances in OCTA allow imaging of blood flow deeper than the retinal layers, at the level of the choriocapillaris (CC), providing new dynamic information about choroidal physiology. OCTA is currently the only noninvasive modality available for imaging the CC in a clinical setting. Also, CC flow is not adequately visualized in traditional angiography so OCTA is the only usable alternative. When OCTA is used to image the choroidal innermost thickness, a granular image is obtained [24] showing a pattern of bright areas, representing flow, and a pattern of small dark regions showing areas of no flow called flow voids. Several examples of this type of image can be seen in Figure 1. Several studies [8, 19, 6] reported a close correlation between abnormal flow void distribution and multiple retinal and choroidal diseases such as age-related macular degeneration, diabetic retinopathy, glaucoma, etc. Therefore, this different distribution of flow voids can be used as a biomarker to diagnose and monitor the progression of certain pathologies. In addition, it can also be used to monitor the efficacy of an applied treatment, as is the case, for example, in the treatment of central serous chorioretinopathy (CSC) [12].

In order to carry out an accurate and objective monitoring over time of the evolution of a pathology or the effectiveness of a treatment, it is necessary to perform an adequate comparison of the different progressive obtained images. A pixel-to-pixel comparison can adequately reflect these changes in the flow voids distribution over time as long as the images are properly aligned. CC images are characterized by a very rough and grainy appearance, presenting numerous complex local deformations that make pixel-to-pixel comparisons unfeasible and inappropriate. In order to carry out this type of comparison, these local deformations must be corrected as much efficiently and accurately as possible without distorting the original image by means of a deformable registration.

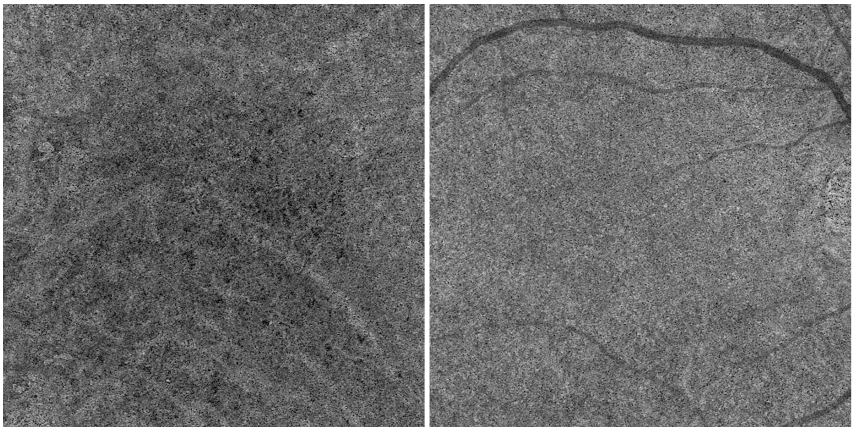

Fig. 1. Two examples of OCTA cc images.

However, there are several factors that hinder this registration process. Firstly, the roughness of the images discussed above, which gives rise to numerous complex deformations. Secondly, the characteristic absence of visually significant anatomical structures. This landmark scarcity hardens drastically the identification of points of interest to achieve an accurate registration. Lastly, the regular use of high-resolution imaging (1024×1024 pixels) in the clinical setting. The use of high-resolution images complicates the alignment process in terms of time and accuracy. To the best of our knowledge, there are no works that try to solve this deformable image registration problem in this novel image modality.

From the perspective of deformable registration in other medical imaging modalities, several traditional optimization methods such as B splines [22] or dense vector fields [25] deal with matching pairs of images. Unfortunately, these methods often require considerable time and computer resources to register a given pair of images. In the state of the art, several contributions have proposed the use of neural networks to perform medical image registration issues [17, 9]. There are some perspectives that use unsupervised training that does not rely on any ground truth directly, although it can be used in a complementary way [18, 4]. In general, these works propose the use of a convolutional neural network such as U-net [21] and a spatial transformation function [14] that warps images to one another. These architectures proved to be useful in solving registration tasks in modalities such as magnetic resonance imaging (MRI) where there are clearly defined anatomical structures that facilitate the registration process in a natural(due to the presence of characteristic points on the image) or guided way (as complementary ground truth). In contrast, CC OCTA images are characterized by their granularity, and unlike other imaging modalities as MRI, there do not present clear and defined anatomical structures, making it complicated to accurately align images. In addition, these architectures are designed to work with relatively small image resolutions (typically 256×256). The use of high-

resolution images poses several problems, the main one being the loss of model efficiency due to factors such as insufficient receptive field. To adequately perform a registration task, the receptive field of the convolutional kernel in the smallest layer of the network must be at least as large as the expected maximum displacement. An increase in image resolution causes the maximum expected displacement to be larger than the cases of smaller images, resulting in a loss of quality in the registration. To increase the receptive field of the network and solve this problem, several works in related tasks such as semantic segmentation have used techniques such as dilated convolutions [29], wider networks [28] or multiscale images [30].

In this work, a novel approach is proposed using a convolutional neural network trained by unsupervised learning to register CC OCTA images. The proposal aims at achieving an efficient registration by overcoming the limitations imposed by the characteristics of using rough images, without clearly defined structures and presenting high resolution. Our methodology is tested with cases of a real clinical study about monitoring the response to a treatment.

2 Materials and Methods

2.1 Dataset

The dataset consists of a total of 821 CC OCTA images (1024×1024 pixels) from 52 patients with chronic CSC obtained using the Zeiss Plex Elite capture device. All the images were acquired by two well-trained clinical experts, being preliminarily used in a clinical study to evaluate the efficacy of studying changes in CC and choroid (CH) flow signal voids as a biomarker for monitoring the response of photodynamic therapy applied to CSC. Within the dataset, there are two different types of images, CC slab and CH slab images. For each patient, images were obtained at different times during treatment. The time instants correspond to pre-treatment, 2 to 4 days after treatment, 1 month after treatment, 3 months after treatment and 6 months after treatment. An example of the first 4 time instants for the CC image modality and the CH modality is shown in Figure 2.

2.2 Network architecture

We created an architecture (PyTorch 1.6 [20]) composed of convolutional layers and a spatial transform. The input of our network consists of the reference image and the image to be registered. The network is divided into 3 branches that use the input images at different scales (low, medium and high resolution images). In each of these branches, convolutions are applied with kernels of size 3×3 and a stride of 2. Each convolution is followed by a batch normalization layer that helps the convergence of the model acting as a regulating factor [13] and a LeakyReLU layer. The low resolution branch allows to make convolutions in a fast and efficient way, being able to accumulate several consecutive convolutions

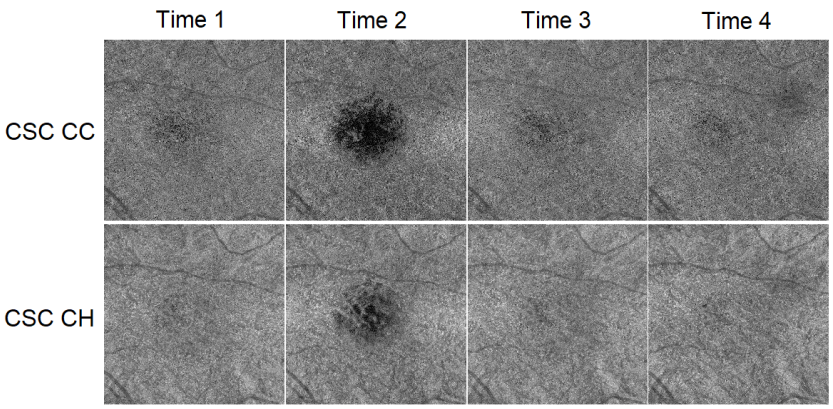

Fig. 2. Images for the first 4 temporal instants of a patient.

which increases the receptive field. In order to further increase the receptive field, the last convolution of this branch is applied using a dilation. This branch allows us to obtain most of the image features in a coarse but efficient way. This allows the independent medium and high resolution branches to concentrate specifically on fine details that refine the final result without having to provide a complete representation of all the features that make up the image so that far fewer convolutions are needed. This reduces the number of model parameters that must be optimized, which reduces the risk of overfitting and aids in the convergence [5]. The 3 branches are merged using transpose convolutions, also using skip connections. Several extra convolutions are used to refine the result at the final resolution. The output of this plus the image to be registered serve as input to the spatial transformer which produces the final registered image. Figure. 3 shows an overview of the neural architecture.

2.3 Training details

The dataset was randomly divided into training (533 images), validation (128) and test (160) subsets. All the sets were independent from each other and all the images of a patient belonged to a single set. Local normalized cross correlation [3], a popular metric that is robust to intensity variations, was used as loss function. Network parameters were initialized using the He et al. [11] method. A batch size of 5 was used as it offered the best results in previous tests. As optimizer, we used the stochastic gradient descent with an initial learning rate of 0.001. It used a dynamic learning rate that was reduced by a factor of 0.7 if the loss of validation did not fall after 40 epochs. Early stopping was performed based on the validation loss. To make the model more robust and avoid overfitting, an exhaustive data augmentation process has also been applied [7]. Some transformations were grouped so that only one from the group could be applied at a time. Table 1 shows all the transformations applied to the images.

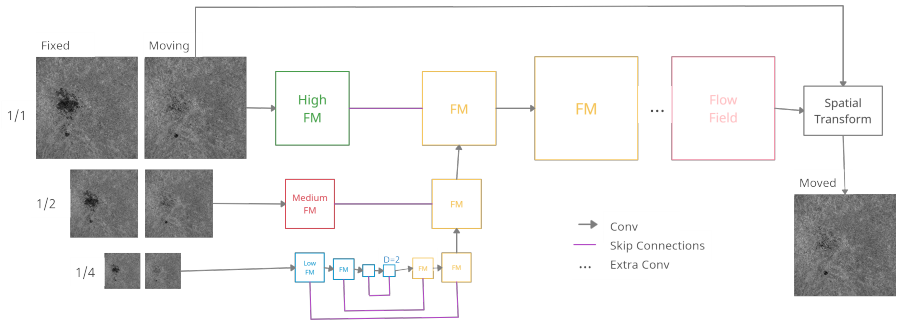

Fig. 3. Overview of the neural architecture. The network is composed of 3 branches (green, red and blue) that accept different sizes of inputs formed by convolutional layers, an expansive part that ends with some extra convolutional layers (yellow) and finally a spatial transform that outputs the registration image. FM=Feature maps.

Table 1. Transformations applied to fixed and moving images in each batch. From each group only one transformation can be applied at a time.

Groups	Transformations
Group 1	Coarse Dropout
Group 2	Elastic Transform, Piecewise Affine
Group 3	Shift Scale Rotate
Group 4	Horizontal Flip
Group 5	Vertical Flip
Group 6	Random Rotate 90^{o}
Group 7	Blur, Gaussian Blur, Motion Blur, Median Blur
Group 8	CLAHE, Brightness Contrast
Group 9	Gauss Noise, Image Compression, Multiplicative Noise

2.4 Baseline Methods

In this work, we selected representative baseline methods to compare and remark the suitable performance of the proposal. In particular, we use as first baseline Symmetric Normalization (Syn) [1], one of the best performing registration algorithms [16]. We use the version of this algorithm in the publicly available software package Advanced Normalization Tools (ANTs) [2], using mutual information (Syn) and cross-correlation (SynCC) as optimization metrics. We also tested other registration algorithms belonging to this package, such as the time-varying diffeomorphism using mean square metrics (TVMSQ) or some simple affine transformations.

2.5 Evaluation metric

As metrics to quantitatively evaluate the alignment of this proposal, we selected different complementary statistics that are frequently used in these issues. For example, MSE is a traditional and simple method for measuring point distances between two images. The two images are compared pixel by pixel and the average of the square of the difference between the error of the two images is calculated. Other types of metrics such as SSIM [26] are also commonly used to evaluate image quality, being a type of metric that correlates well with human visual perception, as it evaluates structural differences between images by comparing local statistics rather than measuring point distances. In particular, we considered: MSE, NRMSE, SSIM, MSSIM[27], and VIF[23].

3 Results and Discussion

3.1 Quantitative Evaluation

Table 2 shows the results obtained for the different similarity metrics by the registration algorithms that were tested. In the table header, the value before registration is shown below the metric name and the time is given in seconds. All the experiments were performed using an intel(R) Core(TM) i5-6300HQ processor and NVIDIA GeForce GTX 950M graphics. As can be seen, our proposal achieves the best values for all the tested similarity metrics, demonstrating the ability to achieve better image alignment than the other tested algorithms. Also, the registration efficiency of our method is superior to the rest of the tested algorithms involving lower execution times than the rest of the tested methods using a CPU or even better exploiting GPU capabilities.

3.2 Qualitative Evaluation

While the used metrics show that our proposal is efficient in achieving image alignment, we must check if our method is able to align the images obtained at different time instants in such a way that no strange deformations are produced, the noise produced by a bad alignment is mitigated whereas the truly significant

Table 2. Performance of different registration methods and the proposal in terms of time and image similarity. The header of each metric: the pre-registration value.

Methods	SSIM 0.0931	MSSIM 0.2103	VIF 0.0394	MSE 1482.1	NRMSE 0.3180	CPU seconds	GPU seconds
Affine	0.1564	0.3901	0.0527	1384.7	0.3068	361.2	
TVMSQ	0.1130	0.2230	0.0390	1437.3	0.3132	15787.6	
Syn	0.2920	0.5772	0.0828	1109.3	0.2744	1372.8	
SynCC	0.4452	0.6659	0.1180	965.3	0.2550	47772.6	
Our Model	**0.7228**	**0.7113**	**0.1694**	**473.0**	**0.1796**	**233.7**	**38.1**

differences between the images are maintained. We must also check if our method allows a pixel-to-pixel comparison and that this ultimately produces a benefit for the clinician. Figure. 4 shows illustrative results of applying the different registration algorithms to the same pair of images. Column one shows the fixed image, column two the registered image and column three the pixel by pixel differences between the fixed and registered images with a color code where the red represents a darker value in the fixed image, blue a darker value in the moving image and white the same value. Two values whose intensity value is less than 20 units apart are counted as equal as being not representative. As can be seen, our method substantially improves the visualization of pixel-to-pixel differences compared to the other tested methods. In addition, it manages to register the images without producing strange and undesired deformations.

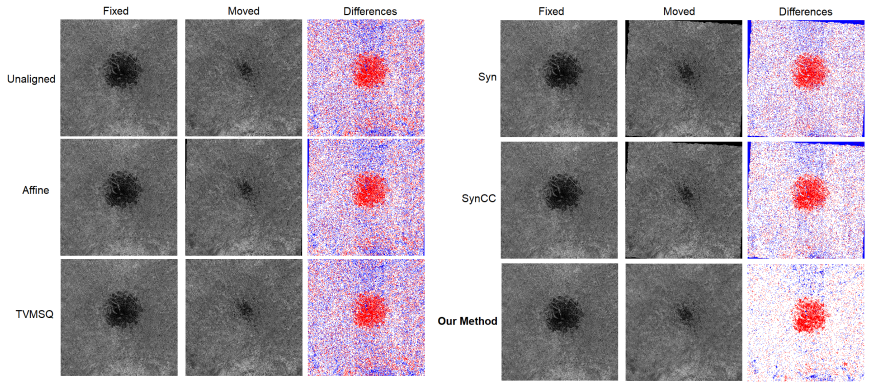

Fig. 4. Comparison of the registration of the different tested algorithms. Column 1, the fixed image; column 2, the image after the registration (except the first row, where the default unregistered images are shown); and column 3, pixel by pixel differences between the fixed and the moving image. The difference image shows in red the darkest areas in the first image, in blue the darkest areas in the second image and in white the areas that are considered equal in both images.

As can be seen in Figure. 5, our proposal is able to eliminate a large part of the noise resulting from local deformations without altering the truly significant changes in the image such as changes in the distribution of the flow voids. Regarding the clinical case that was used to test our methodology, this produces an improvement in the monitoring of the efficacy of photodynamic therapy treatment in patients with CSC, and can also be used as a complementary preliminary step to other computational and clinical analyses. Although our experiment was based on the monitoring of photodynamic therapy in patients with CSC, our methodology is expandable to other pathologies associated with this type of image. Therefore, our methodology has a direct clinical relevance. Also, our proposal presents the potential of being applied to other complex medical image modalities, specially to other depths of OCTA imaging, of a great complexity and interest.

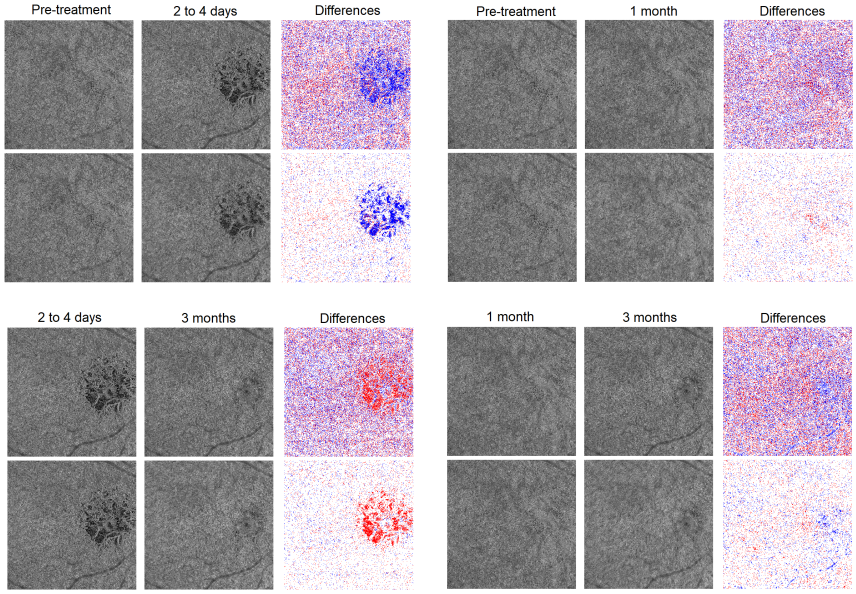

Fig. 5. Images and pixel-to-pixel differences obtained with our proposal for different time instants of the same patient. First row of each set, the fixed image and the unregistered moving image; second row, the fixed image and the registered image.

4 Conclusions

In this work, we have presented a robust registration methodology that employs a convolutional neural network trained by unsupervised learning to register CC OCTA images. The quantitative and qualitative obtained results have

demonstrated that our methodology is capable of achieving efficient and effective image registration in a current clinical problem such as CSC treatment monitoring, overcoming the limitations imposed by the characteristics of using high-resolution, rough images without clearly defined anatomical structures. Furthermore, our methodology can be applied to other pathologies associated with this type of imaging and being adapted to other medical imaging modalities of great complexity and lack of representative visual anatomical structures. As future work, a more detailed ablation study of the proposed architecture could be performed to see which of its components most affect the performance of the architecture.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)
2. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9252–9260 (2018)
4. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
5. Battiti, R.: Using mutual information for selecting features in supervised neural net learning. IEEE Transactions on neural networks **5**(4), 537–550 (1994)
6. Bhutto, I., Lutty, G.: Understanding age-related macular degeneration (amd): relationships between the photoreceptor/retinal pigment epithelium/bruch's membrane/choriocapillaris complex. Molecular aspects of medicine **33**(4), 295–317 (2012)
7. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information **11**(2) (2020). https://doi.org/10.3390/info11020125, https://www.mdpi.com/2078-2489/11/2/125
8. Cao, J., McLeod, D.S., Merges, C.A., Lutty, G.A.: Choriocapillaris degeneration and related pathologic changes in human diabetic eyes. Archives of Ophthalmology **116**(5), 589–597 (1998)
9. Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D.: Deformable image registration based on similarity-steered cnn regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 300–308. Springer (2017)
10. De Carlo, T.E., Romano, A., Waheed, N.K., Duker, J.S.: A review of optical coherence tomography angiography (octa). International journal of retina and vitreous **1**(1), 5 (2015)
11. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034 (2015)

12. Ho, M., Lai, F.H.P., Ng, D.S.C., Iu, L.P.L., Chen, L.J., Mak, A.C.Y., Yip, Y., Cheung, C., Young, A.L., Brelen, M.: Analysis of choriocapillaris perfusion and choroidal layer changes in patients with chronic central serous chorioretinopathy randomised to micropulse laser or photodynamic therapy. British Journal of Ophthalmology (2020)
13. Horwath, J.P., Zakharov, D.N., Megret, R., Stach, E.A.: Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. npj Computational Materials **6**(1), 1–9 (2020)
14. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial transformer networks. arXiv preprint arXiv:1506.02025 (2015)
15. Jia, Y., Tan, O., Tokayer, J., Potsaid, B., Wang, Y., Liu, J.J., Kraus, M.F., Subhash, H., Fujimoto, J.G., Hornegger, J., Huang, D.: Split-spectrum amplitude-decorrelation angiography with optical coherence tomography. Opt. Express **20**(4), 4710–4725 (Feb 2012). https://doi.org/10.1364/OE.20.004710, http://www.opticsexpress.org/abstract.cfm?URI=oe-20-4-4710
16. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. Neuroimage **46**(3), 786–802 (2009)
17. Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F.C., Miao, S., Maier, A.K., Ayache, N., Liao, R., Kamen, A.: Robust non-rigid registration through agent-based action learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 344–352. Springer (2017)
18. Li, H., Fan, Y.: Non-rigid image registration using fully convolutional networks with deep self-supervision. arXiv preprint arXiv:1709.00799 (2017)
19. Lutty, G., Grunwald, J., Majji, A.B., Uyama, M., Yoneya, S.: Changes in choriocapillaris and retinal pigment epithelium in age-related macular degeneration. Mol Vis **5**(35), 35 (1999)
20. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
22. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast mr images. IEEE transactions on medical imaging **18**(8), 712–721 (1999)
23. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Transactions on image processing **15**(2), 430–444 (2006)
24. Spaide, R.F., Fujimoto, J.G., Waheed, N.K.: Image artifacts in optical coherence angiography. Retina (Philadelphia, Pa.) **35**(11), 2163 (2015)
25. Thirion, J.P.: Image matching as a diffusion process: an analogy with maxwell's demons. Medical image analysis **2**(3), 243–260 (1998)
26. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)

27. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. vol. 2, pp. 1398–1402. Ieee (2003)
28. Wu, Z., Shen, C., Van Den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognition **90**, 119–133 (2019)
29. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
30. Zhao, H., Qi, X., Shen, X., Shi, J., Jia, J.: Icnet for real-time semantic segmentation on high-resolution images. In: Proceedings of the European conference on computer vision (ECCV). pp. 405–420 (2018)