Unsupervised Learning Model for Registration of Multi-phase Ultra-Widefield Fluorescein Angiography

Gyoeng Min Lee[1], Kwang Deok Seo[1], Hye Ju Song[1], Dong Geun Park[2], Ga Hyung Ryu[2], Min Sagong[2], and Sang Hyun Park[1(✉)]

[1] Department of Robotics Engineering, DGIST, Daegu, South Korea
{rud557,shpark13135}@dgist.ac.kr
[2] Department of Ophthalmology, Yeungnam University College of Medicine, Daegu, South Korea

Abstract. Registration methods based on unsupervised deep learning have achieved good performances, but are often ineffective on the registration of inhomogeneous images containing large displacements. In this paper, we propose an unsupervised learning-based registration method that effectively aligns multi-phase Ultra-Widefield (UWF) fluorescein angiography (FA) retinal images acquired over the time after a contrast agent is applied to the eye. The proposed method consists of an encoder-decoder style network for predicting displacements and spatial transformers to create moved images using the predicted displacements. Unlike existing methods, we transform the moving image as well as its vesselness map through the spatial transformers, and then compute the loss by comparing them with the target image and the corresponding maps. To effectively predict large displacements, displacement maps are estimated at multiple levels of a decoder and the losses computed from the maps are used in optimization. For evaluation, experiments were performed on 64 pairs of early- and late-phase UWF retinal images. Experimental results show that the proposed method outperforms the existing methods.

Keywords: Registration · Unsupervised learning · Deep learning · Vesselness map

1 Introduction

Ultra-Widefield (UWF) retinal imaging plays an essential role in the diagnosis and treatment of eye diseases with the peripheral retinal changes. This imaging device has multimodal capabilities including fundus photographs, fluorescein angiography (FA), indocyanine green angiography, and autofluorescence images.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59716-0_20) contains supplementary material, which is available to authorized users.

© Springer Nature Switzerland AG 2020
A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12263, pp. 201–210, 2020.
https://doi.org/10.1007/978-3-030-59716-0_20

Among them, the FA shows pathological changes in blood vessels such as leaks, new blood vessels, and ischemia, which can be distinguished through phase-by-phase comparison. Thus, for accurate identification and quantitative evaluation of these lesions, the registration of multi-phase images is inevitable. However, the registration is non-trivial since each image has peripheral distortion in the process of projecting a 3D fundus as a 2D image and intensity distributions of the early- and late-phase images are very different and the displacements with respect to eyeball movements are often large as shown in Fig. 1.

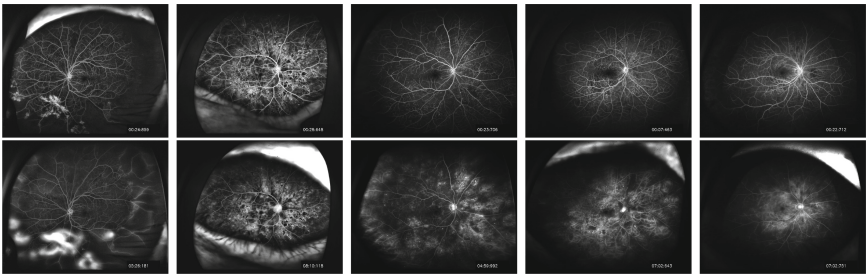


Fig. 1. Early-phase UWF FA images (Top) and late-phase UWF FA images (Bottom). It shows that the intensity distribution is different and the position of the vessel is changed by the movement between phases.

To address this problem, we propose an unsupervised learning model for aligning the multi-phase UWF FA retinal images. Recently, deep learning-based registration methods [1,15,18,21,22,27,28] that do not require ground truth displacements have been proposed, but they mostly performed intensity-based non-rigid registration to align relatively local displacements between moving and fixed images. However, these methods often fail to perform the registration of images with different intensity distribution and large displacements. We introduce a novel method that consists of an encoder-decoder style network for estimating displacements and spatial transformers to reconstruct the moved image as well as a vesselness map generated from the moving image. The moved image and vesselness map are estimated from multiple levels of a decoder and all losses from the multiple estimations are considered to address large displacements.

The main contributions of this work are as follows: (1) The proposed method provides an effective way to align inhomogeneous images, while taking advantages of the proposed unsupervised learning model by adding the loss computed with the vesselness feature map. (2) The proposed method can integrate any suitable features according to image characteristics. Though we use vesselness maps as features in this work, any advanced feature extractors that enable one to extract consistent features from inhomogeneous images can be used. (3) Large displacements are precisely estimated without pre-processing (*e.g.*, affine registration) or post-processing steps. We empirically show the benefit of using multi-level decoder predictions with improvements in performance. (4) Lastly,

in the UWF FA with multiple phases, changes between images can be identified and progress can be evaluated over time. To the best of our knowledge, this is the first work to address the registration of UWF FA images.

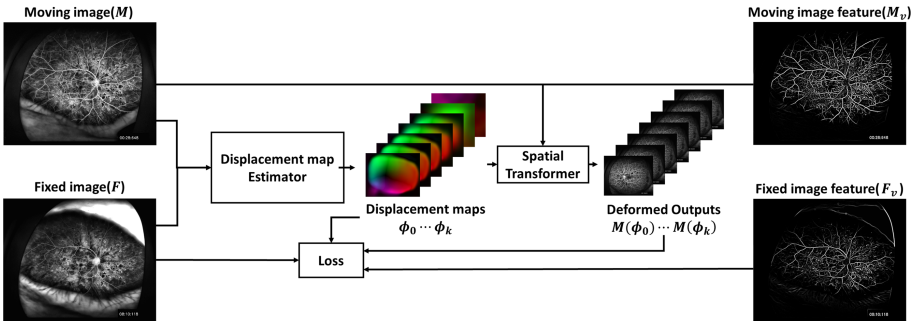


Fig. 2. Overview of our proposed method.

1.1 Related Works

Retinal Image Registration: Most of retinal image registration methods were feature-based approaches [4,7,13,20,26]. For example, vessels and bifurcation point detectors [20], SIFT [23], or edge-based Harris corner detector [6] were used to find correspondences, and then the registration was performed using iterative closest point or spline-based methods. However, these methods often fail to find robust correspondences. Recently, a deep learning-based method [14] has been proposed to find correspondences in inhomogeneous images and improve the registration performance. However, this method could not predict a dense displacement map in a single end-to-end deep learning framework.

Registration via Unsupervised Deep Learning Model: Several deep learning frameworks have been proposed for registration of medical images. Among them, supervised learning-based methods [5,12,19,24] often achieved limited results since it was difficult to produce the ground truth of displacements. Recently, registration methods [15,18,21,22] that do not require training data have been proposed. For example, Vos et al. [22] predicted a sparse displacement grid and then performed interpolation using a third-order B-spline kernel. Balakrishnan et al. first proposed a CNN network [1] that predicts dense displacement map end-to-end using spatial transformers [10]. Zhao et al. [28] proposed a cascaded framework [27] to address large displacements by gradually registering the images. However, these methods often did not work properly on the registration of images with different intensity distributions. The proposed method mainly addresses this limitation by minimizing a feature-based loss.

2 Method

Our proposed network consisting of a displacement map estimator and spatial transformers is shown in Fig. 2. In the displacement map estimator, a U-net style network with K-level encoders and decoders predicts displacement maps $\phi_0, \phi_1, ..., \phi_K$ to match the coarse to fine displacements between a moving image M and a fixed image F. The spatial transformers generate the moved images $M(\phi_0), M(\phi_1), ..., M(\phi_K)$ using the displacement maps. With the moved images, we also extract a vesselness feature maps M_v from M using Frangi filtering [8] and then generate the moved vesselness maps $M_v(\phi_0), M_v(\phi_1), ..., M_v(\phi_K)$. The network is learned to minimize the differences between $M(\phi_0), M(\phi_1), ..., M(\phi_K)$ and F as well as $M_v(\phi_0), M_v(\phi_1), ..., M_v(\phi_K)$ and the vesselness feature map F_v of F. After the model is sufficiently trained, $M(\phi_0)$ is considered as the final registration result.

2.1 Unsupervised Learning Model for Registration

Inspired by VoxelMorph [1], we follow the overall structure of that model using the spatial transformer. However, since the U-Net structure [17] is not designed to adjust large displacements, we predict the displacement maps at each level of decoder as shown in Fig. 3. M and F pass through 6 encoders and decoders (i.e., $K = 6$) with [16, 32, 32, 32, 64, 64] channels for encoding block and its reversed order for decoding block. The feature maps of each level of the decoder pass through a 3×3 convolution layer with LeakyReLU activation function to estimate the displacements with different resolutions. A convolution layer with 4×4 kernel size and 2×2 stride is used without pooling in the conv block of encoder, while a $\times 2$ upsampling is used in the conv block of decoder. The displacement maps in low resolution, i.e., ϕ_K or ϕ_{K-1}, are upsampled to the size of F by the linear interpolation.

In optimization using gradient-descent, the parameters should be differentiable with respect to the loss function. Thus, the spatial transformer proposed in [10] is used for this purpose. The spatial transformer uses bilinear interpolation to calculate values between pixels in an image. Then, the pixel values of $M(\phi)$ are sampled using ϕ. By using linear interpolation, we can transform images with discontinuities into a continuous grid. This makes it possible to perform differentiable transformations.

2.2 Loss Function

Spatial transformers allow the use of any deformation field in any of the derivative fields. Thus, we define the loss function using the feature maps obtained from the input image along with a similarity loss. Thus, similarity loss (L_{sim}) between $M(\phi)$ and F is defined as the negative cross-correlation(CC) of local regions of $M(\phi)$ and F. In particular, let $\hat{F}(p)$ and $\hat{M}(\phi(p))$ denote the mean

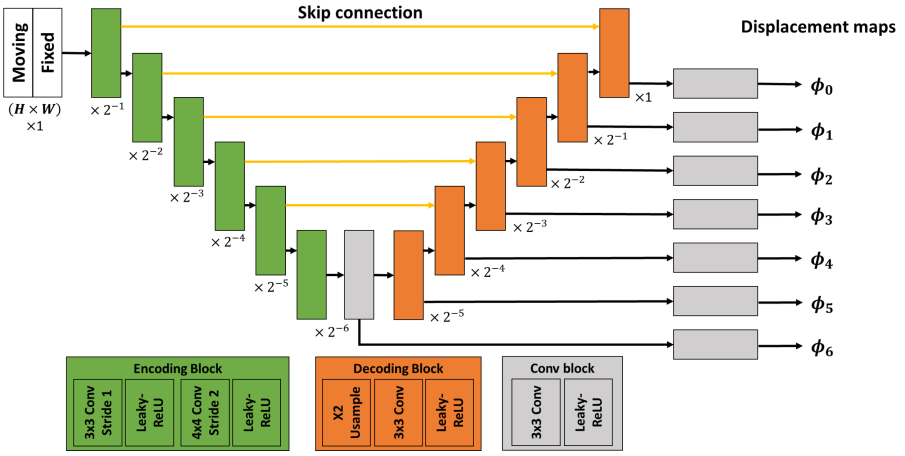


Fig. 3. Architecture of the proposed deformation field estimator network.

intensity in a $n \times n$ patch, the local cross-correlation of F and $M(\phi)$ is defined as:

$$CC(F, M(\phi)) = \sum_{p \in \Omega} \frac{(\sum_{p_i}(F(p_i) - \hat{F}(p))(M(\phi(p_i)) - \hat{M}(\phi(p))))^2}{(\sum_{p_i}(F(p_i) - \hat{F}(p)))(\sum_{p_i}(M(\phi(p_i)) - \hat{M}(\phi(p))))}, \quad (1)$$

where p_i iterates over a $n \times n$ patch around p. We set $n = 9$ in our experiments. The higher the CC value, the better the alignment. The loss function is used in the form of $L_{sim}(F, M(\phi)) = -CC(F, M(\phi))$. Furthermore, to match $M(\phi)$ with F with different characteristics, CC between $M_v(\phi)$ and F_v is also added to the loss function. For the feature maps, Frangi filter is applied to M and F and each vesselness map is extracted and scaled to [0, 1] range. Displacement maps obtained by network are applied to ϕ to generate $M_v(\phi)$. CC between $M_v(\phi)$ and F_v was calculated as: $L_{vessel}(F_v, M_v(\phi)) = -CC(F_v, M_v(\phi))$.

With L_{sim} and L_{vessel}, we also use the smoothness loss L_{smooth} to regularize unnatural displacements. It is based on the gradient of the deformation field as:

$$L_{smooth}(\phi) = \sum_{p \in \Omega} \| \bigtriangledown \phi(p) \|^2, \quad (2)$$

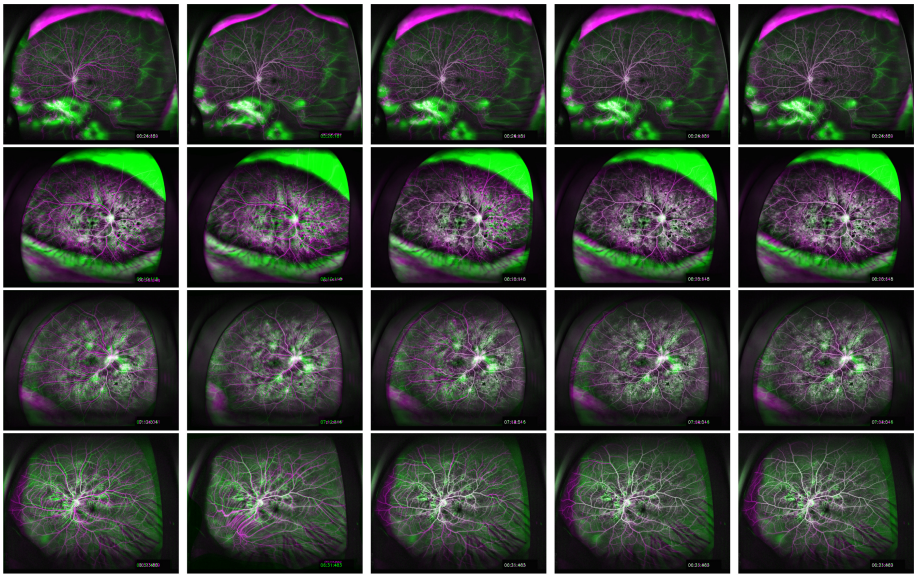
Finally, the final loss is defined as:

$$L_{total} = \sum_{k=0}^{K} (L_{sim}(F, M(\phi_k)) + L_{vessel}(F_v, M_v(\phi_k)) + \lambda_1 L_{smooth}(\phi_k)), \quad (3)$$

where λ is the regularization parameter.

2.3 Implementation Details

We used PyTorch [16] to implement the proposed method and experimented on Intel i7-8700K CPU, NVIDIA Geforce 1080Ti GPU with 64 GB RAM. The ADAM [11] optimizer was used for optimization and the learning rate was set to $1e^{-4}$. In each step, 16 mini-batches were randomly chosen from all 64 samples. The gradient regularization parameter λ_1 of the deformation field was 1 for VoxelMorph, and 0.001 for the proposed method. All models were trained until the losses converged.



(a) (b) (c) (d) (e)

Fig. 4. Qualitative results. Each image shows overlap of a deformed early-phase (purple) and a late-phase image (green). Well overlapped blood vessels appear white. (a) Affine, (b) Affine + B-Spline, (c) VoxelMorph, (d) Ours, (e) Ours+vesselness map (Color figure online)

3 Experimental Results

Dataset. For evaluation, we used 1) 30 pairs of early- and late-phase UWF FA images with size of 3900×3076 from 30 patients with diabetic retinopathy and 2) 5∼10 multi-phase UWF FA images with the size of 4000×4000 from 34 patients with other retinal vascular diseases. All these images were acquired from a university hospital. We resized these images to a same size and performed histogram equalization to adjust the intensity range in [0, 1]. We extracted vessels and bifurcation points from the pair of images to measure the registration accuracy. For the vessel extraction, we made a binary segmentation by thresholding

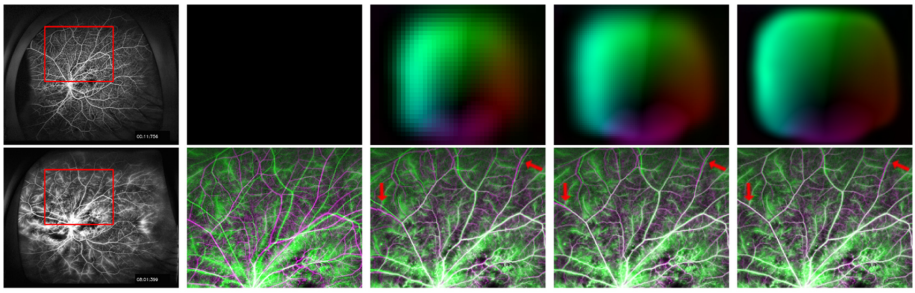
Table 1. Mean recall, precision, DSC, distance scores of the proposed method and the related methods. Moving denotes the accuracy scores between moving images and fixed images.

Method	Precision	Recall	DSC	Distance
Moving	0.11	0.19	0.12 ± 0.07	14.6 ± 19.19
Affine	0.21	0.23	0.21 ± 0.16	10.65 ± 15.27
Affine + B-Spline	0.18	0.15	0.16 ± 0.07	23.72 ± 28.55
SIFT-RANSAC	0.23	0.27	0.24 ± 0.16	50.60 ± 104.7
VoxelMorph	0.40	0.44	0.41 ± 0.18	10.13 ± 20.08
+Vesselness	0.41	0.45	0.42 ± 0.18	10.25 ± 20.32
Ours	0.43	0.46	0.43 ± 0.20	8.99 ± 20.46
+Vesselness	0.45	0.48	0.45 ± 0.19	8.25 ± 20.59

of vesselness scores obtained by Frangi filtering [8] and then manually corrected errors. The bifurcation points were also annotated manually.

Experimental Settings. To confirm the superiority of the proposed method, we compared our method with affine registration, B-spline, SIFT with RANSAC and VoxelMorph. The imregister function, a Matlab built-in function, was used for the affine registration with the mutual information similarity measurement. The transformation matrix was obtained by evolutionary optimization method. After the affine transformation, a B-spline non-rigid registration model implemented in SimpleITK [25] was used and optimized with a gradient based L-BFGS-B [3] optimization algorithm which minimizes the mutual information between a moved and fixed image. OpenCV [2] was used for implementing SIFT descriptor and RANSAC [9] with the perspective transform matrix. For VoxelMorph, we used the code provided by the authors. We predict the results after the model is learned until the losses converged to a certain level around 35,000 epochs. To verify the effect of each element of the proposed method, we also generated the results using the proposed method without feature loss, i.e., in other words, VoxelMorph with multi-level displacement estimation. For evaluation, we compared the average distance between corresponding bifurcation points and the precision, recall, and DSC scores between the vascular masks in 64 pairs of early- and late-phase.

Quantitative Results. Table 1 shows the average precision, recall, DSC between masks and the average pixel distances between correspondences of the proposed method and the related methods. The distances between correspondences of affine transform were not significantly different with the distances before transformation. In the case of affine+B-spline, there were many cases where matching failed due to a large intensity distribution difference. VoxelMorph achieved better performance both in distance and DSC scores, but the



(a) (b) (c) (d) (e)

Fig. 5. Example of the results of the multi-level displacement map of the proposed method. (a) shows early-phase image (top) and late-phase image (bottom) (b) is not a deformed image and (c)–(e) are images in which the deformed image and the late image are overlapped on the red rectangular region shown in (a) with different resolutions of displacement maps (ϕ_4, ϕ_3, and ϕ_0). (Color figure online)

improvement on the distances between correspondences was relatively low. On the other hand, our proposed network achieved the best performances for most cases. Note that the reduced distance between the branching points indicates that the result of our method obtained natural displacement maps. When displacement estimation was performed at multi-levels of the decoder, performance improved by 2% for DSC and 1 for distance compared to VoxelMorph. Furthermore, the performance improved by 3% for DSC and 2 for distance when the loss computed by vesselness was used.

Qualitative Results. Figure 4 shows images of early-phase (moving) and late-phase (fixed), and their qualitative results. The result of affine registration was limited since robust features could not be often extracted and it was difficult to consider non-rigid changes. VoxelMorph aligned most of the vessels well in the pupil area by performing non-rigid registration, but some errors occurred near the periphery with large displacements. On the other hand, the proposed method achieved robust results in most areas. Figure 5 shows the example of multi-level displacement map. In addition, it was shown that elaborate matching was performed as the resolution was changed from low to high resolution.

4 Conclusions

We have proposed a novel registration method for aligning early- and late-phase UWF FA retinal images with different characteristics. The proposed method does not only take advantage of unsupervised learning-based registration methods, but also works effectively for the registration of inhomogeneous images with the addition of a feature loss. Furthermore, we introduce a way to effectively estimate large displacements by predicting the displacement map at each level of

the decoder. The performance improvement of proposed method was significant on the evaluation of 64 pairs of UWF FA retinal images. We believe that the proposed technique will contribute the diagnosis and quantification of eye diseases since it is easy to find leaks, new blood vessels, and ischemia if multi-phase UWF FA images are well aligned.

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2019R1C1C1008727), and the grant of the medical device technology development program funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea)(20006006).

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)
2. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Inc. (2008)
3. Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. **16**(5), 1190–1208 (1995)
4. Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Trans. Pattern Anal. Mach. Intell. **24**(3), 347–364 (2002)
5. Cao, X., et al.: Deformable image registration based on similarity-steered CNN regression. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 300–308. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_35
6. Chen, J., Tian, J., Lee, N., Zheng, J., Smith, R.T., Laine, A.F.: A partial intensity invariant feature descriptor for multimodal retinal image registration. IEEE Transact. Biomed. Eng. **57**(7), 1707–1718 (2010)
7. Choe, T.E., Cohen, I.: Registration of multimodal fluorescein images sequence of the retina. In: Tenth IEEE International Conference on Computer Vision (ICCV 2005), vol. 1, pp. 106–113. IEEE (2005)
8. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0056195
9. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2003)
10. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems, pp. 2017–2025 (2015)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Krebs, J., et al.: Robust non-rigid registration through agent-based action learning. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 344–352. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_40

13. Laliberté, F., Gagnon, L., Sheng, Y.: Registration and fusion of retinal images-an evaluation study. IEEE Trans. Med. Imaging **22**(5), 661–673 (2003)
14. Lee, J.A., Liu, P., Cheng, J., Fu, H.: A deep step pattern representation for multimodal retinal image registration. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5077–5086 (2019)
15. Li, H., Fan, Y.: Non-rigid image registration using fully convolutional networks with deep self-supervision. arXiv preprint arXiv:1709.00799 (2017)
16. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. In: 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
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Sentker, T., Madesta, F., Werner, R.: GDL-FIRE^{4D}: deep learning-based fast 4D CT image registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 765–773. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_86
19. Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B.P.F., Išgum, I., Staring, M.: Nonrigid image registration using multi-scale 3D convolutional neural networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 232–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_27
20. Stewart, C.V., Tsai, C.L., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Trans. Med. Imaging **22**(11), 1379–1394 (2003)
21. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Med. Image Anal. **52**, 128–143 (2019)
22. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 204–212. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_24
23. Yang, G., Stewart, C.V., Sofka, M., Tsai, C.L.: Alignment of challenging image pairs: refinement and region growing starting from a single keypoint correspondence. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1973–1989 (2007)
24. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: fast predictive image registration-a deep learning approach. NeuroImage **158**, 378–396 (2017)
25. Yaniv, Z., Lowekamp, B.C., Johnson, H.J., Beare, R.: Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research. J. Digit. Imaging **31**(3), 290–303 (2018)
26. Zana, F., Klein, J.C.: A registration algorithm of eye fundus images using a Bayesian Hough transform. In: 7th International Conference on Image Processing and its Applications (1999)
27. Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al.: Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10600–10610 (2019)
28. Zhao, S., Lau, T., Luo, J., Eric, I., Chang, C., Xu, Y.: Unsupervised 3D end-to-end medical image registration with volume tweening network. IEEE J. Biomed. Health Inform. (2019)