



Multimodal affine registration for ICGA and MCSL fundus images of high myopia

GAOHUI LUO,^{1,4}  XINJIAN CHEN,^{1,2,4} FEI SHI,¹ YUNZHEN PENG,¹ DEHUI XIANG,¹ QIUYING CHEN,³ XUN XU,³ WEIFANG ZHU,^{1,5}  AND YING FAN^{3,6}

¹*School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China*

²*State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China*

³*Shanghai General Hospital, Shanghai 200080, China*

⁴*contributed equally*

⁵*wfzhu@suda.edu.cn*

⁶*mdfanying@sjtu.edu.cn*

Abstract: The registration between indocyanine green angiography (ICGA) and multi-color scanning laser (MCSL) imaging fundus images is vital for the joint linear lesion segmentation in ICGA and MCSL and the evaluation whether MCSL can replace ICGA as a non-invasive diagnosis for linear lesion. To our best knowledge, there are no studies focusing on the image registration between these two modalities. In this paper, we propose a framework based on convolutional neural networks for the multimodal affine registration between ICGA and MCSL images, which contains two parts: coarse registration stage and fine registration stage. In the coarse registration stage, the optic disc is segmented and its centroid is used as a matching point to perform coarse registration. The fine registration stage regresses affine parameters directly using jointly supervised and weakly-supervised loss function. Experimental results show the effectiveness of the proposed method, which lays a sound foundation for further evaluation of non-invasive diagnosis of linear lesion based on MCSL.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

Introduction

Pathological myopia is a major cause of blindness in many developed countries [1,2]. T. Tokoro et al. proposed that high myopia accompanied by visual dysfunctions might be defined as pathologic myopia [3]. According to [4,5], linear lesion (as indicated by the yellow arrow in Fig. 1) is an important clinical sign for evaluating the development from high myopia to pathological myopia. At present, indocyanine green angiography (ICGA) (shown in Fig. 1(a), (b), (c) and (d)) is considered to be the “Ground Truth” for the diagnosis of linear lesions in ophthalmology clinic [6,7], but it requires the injection of contrast agent indocyanine green (ICG), which may cause adverse reactions such as allergy, dizziness, and even shock [8]. So it is an urgent need to find a non-invasive imaging modality that can replace ICGA for the diagnosis of linear lesions. Multi-color scanning laser (MCSL) imaging is a non-invasive imaging technology, in which three lasers with different wavelengths (488nm, 515nm and 820nm) are used to scan the fundus simultaneously. MCSL image fused by several fundus images (shown in Fig. 1(e), (f), (g) and (h)) can reveal linear lesions more richly than other non-invasive modality such as color fundus imaging and red-free fundus imaging and some other invasive one such as fundus fluorescein angiography (FFA). Therefore, we try to investigate whether MCSL could replace ICGA as a non-invasive imaging for linear lesion diagnosis. At the beginning of this study, the ICGA and MCSL images need to be registered. As can be seen from Fig. 1, the multimodal registration between ICGA and MCSL fundus images is a big challenge due to two aspects: (1) the appearance differences between ICGA and MCSL are large. (2) The retinal vessels are so fuzzy in the late phase ICGA images that it can't be utilized as structure feature during the registration.

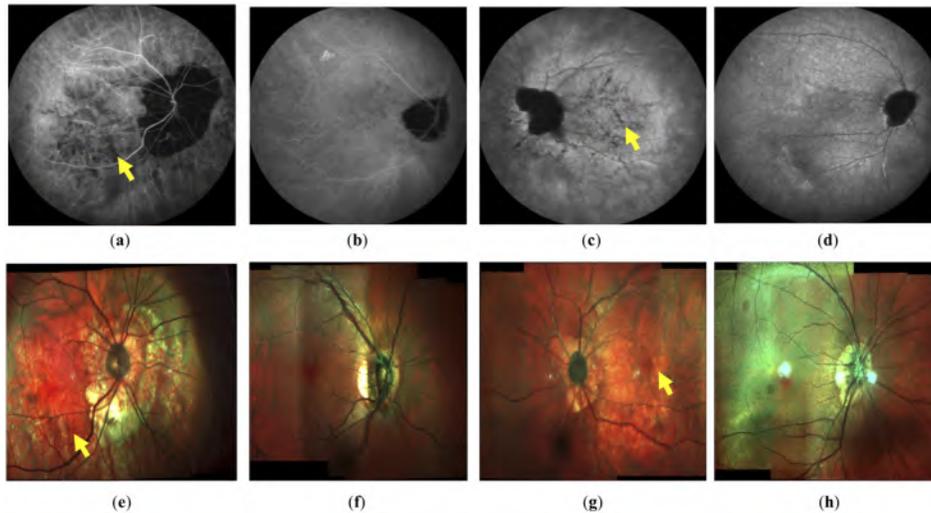


Fig. 1. ICGA images and MCSL images. (a) ICGA image with linear lesions (about 20 minutes after ICG injection). (b) ICGA image of normal fundus (about 20 minutes after ICG injection). (c) ICGA image with linear lesions (about 30 minutes after ICG injection). (d) ICGA image of normal fundus (about 30 minutes after ICG injection). (e), (f), (g) and (h) are the corresponding MCSL images for (a), (b), (c) and (d) respectively. The yellow arrow indicates linear lesions.

Traditional image registration methods can be roughly classified into two categories: intensity-based methods and feature-based methods [9,10]. The main idea of the intensity-based methods [11,12] are to search the geometric transformation parameters iteratively. The geometric transformation parameters are optimized by maximizing or minimizing a similarity metric between the transformed moving image and the corresponding fixed image. Common similarity metrics include cross-correlation (CC) [13], mutual information (MI) [14], normalized mutual information (NMI) [15], and sum of squared difference (SSD) [16], etc. Because of the iterative operations, the time and computational cost of registration is high. Feature-based methods [17,18] usually use features such as point, edge, line, contours surfaces and area, to establish the correspondence between fixed image and moving image. Algorithms such as SIFT [19], SUNSAN [20] and SURF [21] are often used in feature extraction. Generally, the retinal fundus image registration adopts the feature-based methods. Existing feature-based fundus image registration [22–25] mainly used vessel features for registration of color fundus images, FFA image and ICG image. However, it is not suitable to use retinal vessel features for late stage ICGA and MCSL image registration. As shown in Fig. 1(a-d), the retinal vessels in ICGA images are revealed as hyper-fluorescent with noisy background (Fig. 1(a)) or low contrast (Fig. 1(b)) at about 20 minutes and blurry hypo-fluorescent at the very late stage (about 30+ minutes, Fig. 1(c) and (d)).

In order to overcome the shortcomings of the traditional methods, deep learning based methods, which have achieved great success in image classification [26], segmentation [27] and object detection [28], are introduced in image registration and achieves better performances. In particular, the registration time is greatly reduced, making intraoperative real-time registration possible. The deep learning based methods can be classified as supervised learning, weakly supervised learning and unsupervised learning. The supervised methods [29,30] need the corresponding ground truth deformations obtained by traditional methods or manual registration. Reference [31] proposed a dual-supervised deep learning strategy that involves using both supervised and unsupervised loss

functions. There are also some works [32,33] using weakly supervised learning to merge the tissue segmentation label into the registration network and guide the network to learn registration parameters, which do not require registration ground truth. In recent years, to avoid the difficulties of registration ground truth obtainment and exhausting manual segmentation, the unsupervised learning method [34–36] has been developed. In this kind of methods, similarity metric is used as cost function to be maximized or minimized during network training. However, [37] pointed out that manually crafted similarity metrics have made very little success in multimodal registration. Recently, generation adversarial network (GAN) has been applied in registration [38–42], in which the discriminator is equivalent to the similarity metric.

Several studies [29], [43–45] have pointed out that a single convolutional neural network can not work well in registration task with large deformation. To solve this problem, [29] proposed CNN regressors to predict the affine parameters in a hierarchical manner. Reference [43] adopted the multi-stage strategy by stacking multiple stages of ConvNets, in which each stage has its own registration task. Work [44] proposed a multi-step affine framework which contains only one neural network. As in a recurrent network, the parameters of the affine network are shared by all steps. At each step, the network outputs the parameters to refine the previously predicted affine transformation. However, the best number of iterations needs a lot of experiments to be sure and recurrent training of a single neural network will increase the training time. Paper [45] proposed an unsupervised end-to-end learning with progressive alignment through deep recursive cascaded neural network. In this work, they trained a multi-network to transform images with large deformation gradually. Furthermore, [46] proposed a two-step registration method with traditional registration methods.

To the best of our knowledge, multimodal registration of ICGA and MCSL has not been reported yet. Since the high myopia cases to be analyzed are in the stationary phase, the key fundus features such as optic disc, retinal vessel and linear lesion do not change significantly in the MCSL and ICGA image pair, although there may be a certain time interval in the acquisition of image pair (usually no more than 1-2 weeks which is the common appointment interval of ICGA check in clinic). That is to say, the differences between these two modalities are mainly caused by the capture angle and visual field, which can be aligned through affine transformation such as scaling, rotation, etc. However, the registration task still faces great challenges because both the inter-subject differences of ICGA/MCSL and intra-subject differences between ICGA and MCSL are great, as shown in Fig. 1. This will make it difficult to optimize the required affine transformations at once, including scaling, rotation, translation and shearing transformations, especially large translations and scaling.

Therefore, in this paper, we propose a two-stage affine registration framework to achieve the registration of ICGA and MCSL images from coarse to fine. The affine registration parameters are regressed through the two-stage registration networks. In the first stage, the feature information of the image such as optic disc is used for coarse registration, which can greatly decrease the initial registration error. In the second stage, in order to avoid the over-fitting caused by the registration ground truth based supervised training and the under-fitting caused by the few prior structural feature (only optic disc) based weakly-supervised training, we adopts a dual-supervised training strategy by combining the supervised and weakly-supervised loss functions to achieve the fine registration of image pair. The main contributions of this paper are as follows:

- The framework of two-stage registration from coarse to fine is proposed.
- The supervised and weakly-supervised loss functions are jointly applied in this work.
- Image prior information such as optic disc is efficiently used in both stages of the registration.
- Multimodal registration for ICGA and MCSL images is explored for the first time.

The remainder of the paper is organized as follows. In Section 2, the proposed method is described in detail. In Section 3, experimental results are shown and analyzed, and followed by the conclusions and discussions in Section 4.

1. Methods

1.1. Overview

The overview of the proposed method is shown in Fig. 2, including coarse registration and fine registration in both training and test phases.

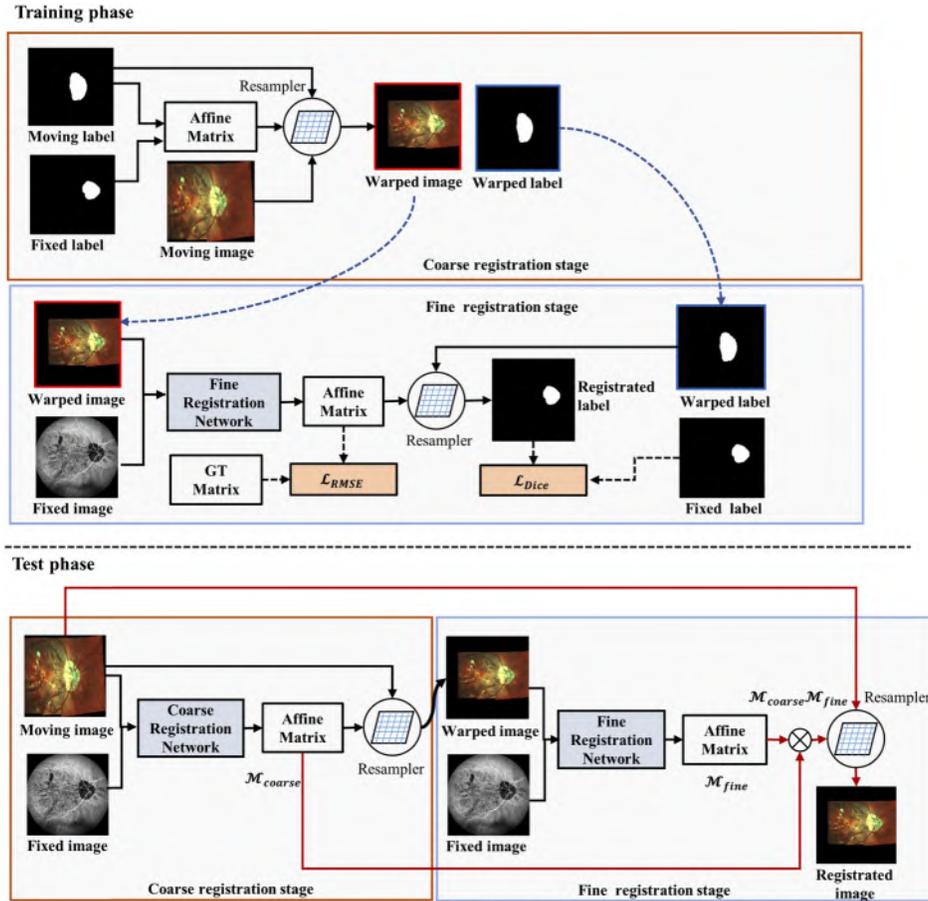


Fig. 2. The framework of proposed method. In the training phase, black dash lines represent the data flow of dual supervision. In the test phase, the red lines represent the data flow of final registration.

Due to the large scale and space differences between the original ICGA and MCSL image pair, a coarse registration consisting of translation and scaling can greatly reduce the initial registration error and reduce the difficulty of the subsequent fine registration. The translation parameters are calculated from the centroid coordinates of the paired optic disc labels. The scaling parameter is an empirical value which is statistically analyzed and estimated according to our experimental dataset. The fine registration network is improved from a Resnet18 model [47]. During the training of the fine registration network, the predicted affine parameters and the registered moving label are used to calculate the supervised loss function L_{rmse} (Eq. (10)) and the weakly-supervised loss function L_{dice} (Eq. (11)) respectively, which realize the dual-supervision loss function L (Eq. (9)).

The networks are trained to obtain the optimal network parameters $\hat{\xi}$ by minimizing the dual-supervision loss function L and predict the optimal affine transformation parameters \hat{M} :

$$\hat{\xi} = \arg \min_{\xi} L(x_f, x_m, M_{gt}, M, l_f, l_m; \xi) \quad (1)$$

$$\hat{M} = \Gamma(x_f, x_m, l_f, l_m; \xi) \quad (2)$$

Where Γ represents the proposed networks. ξ is the parameters of the neural network. x_f and x_m are the original fixed image and moving image respectively. l_f and l_m present the corresponding fixed label and moving label respectively. M_{gt} and M are the ground truth transformation parameters and predicted transformation parameters respectively.

In the test phase, the original image pair (x_f, x_m) and the corresponding label pair (l_f, l_m) are coarsely registered first, and then the coarsely registered image pair and the corresponding label pair are sent to the trained fine registration network to predict the parameters. In the coarse registration Γ_{coarse} stage, two trained U-Net networks [27] are used to automatically segment the optic disc label (l_f, l_m) . The translation parameters obtained from the label centroid coordinates alignment, combining with the scaling parameters, are adopted to obtain the coarse registration affine matrix M_{coarse} :

$$\Gamma_{coarse}(x_f, x_m, l_f, l_m) = M_{coarse} \quad (3)$$

And then the coarsely registered moving image x_m^w and the corresponding label l_m^w are obtained by Eq. (4):

$$\text{Re sampler}(x_m, l_m; M_{coarse}) = x_m^w, l_m^w \quad (4)$$

where the $\text{Re sampler}(\cdot)$ is the operation of affine transformation.

In the fine registration Γ_{fine} stage, the trained fine registration network is used to predict the affine transformation matrix M_{fine} :

$$\Gamma_{fine}(x_f, x_m^w, l_f, l_m^w) = M_{fine} \quad (5)$$

Finally, the original moving image x_m is interpolated only once to get the final registered image x_m^{reg} :

$$\text{Re sampler}(x_m; M_{coarse}M_{fine}) = x_m^{reg} \quad (6)$$

It can avoid information loss caused by multiple interpolations. No manual registration parameters or manually labeled optic disc labels are required during the test phase.

We summarize some advantages of the proposed framework as shown in Fig. 2. First, we make full use of optic disc label in both stages of registration, which solves the problem that retinal vessels are very insignificant in the late phase of ICGA imaging (shown as Fig. 1(a)-(d)). Second, the supervised and weakly-supervised loss function are effectively combined in the proposed network. Such dual supervision mechanism is the trend of deep learning based registration [48].

1.2. Affine registration

Affine registration is a linear and global transformation, in which the transformation parameters for each pixel are the same. That is to say the coordinates of the moving image can be mapped to the fixed image through a set of parameters. Equation (7) represents the mathematical expression of the affine transformation (rotate around image origin).

$$\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \lambda_x & 0 \\ \lambda_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} \quad (7)$$

where (X, Y) and (X', Y') represent the coordinates of the moving image and the fixed image respectively. a_{11} , a_{12} , a_{21} and a_{22} are the deformation parameters; b_1 and b_2 are translation

parameters along axis x , y respectively. Parameters a_{11} , a_{12} , a_{21} , a_{22} , b_1 and b_2 constitute the coordinate transformation matrix. Equation (8) also gives the detailed composition of the affine transformation matrix, where θ is the rotation angle, k_x , k_y are the shear coefficients in x and y direction, and λ_x , λ_y are the scaling factors in x and y direction, respectively.

1.3. Coarse registration

To make the deep learning network suitable for multimodal image registration with large deformation, it is necessary to reduce the initial deformation. In this paper, the prior information of the image is applied to coarsely register the image pairs. In ICGA and MCSL images, the most remarkable features are the retinal vessels and optic disc. But the retinal vessels are very faint in the late-stage ICGA images (as shown in Fig. 1(c) and (d)) or severely disturbed by the choroidal vessels. Moreover, the noise is serious in the middle-stage ICGA images (as shown in Fig. 1(a) and (b)). Therefore, the retinal vessels are not suitable to be used as prior information. On the contrary, optic disc with near-circular structure is robust in both modalities, so it can be used as the prior information in the coarse registration.

In the training phase, the manual-labeled optic disc label is used to calculate the affine matrix for the coarse registration of the image pair and the corresponding optic disc pair. In the test phase, as shown in Fig. 3, two individually trained U-Net networks are used to automatically segment the optic disc labels in ICGA and MCSL images respectively for coarse registration.

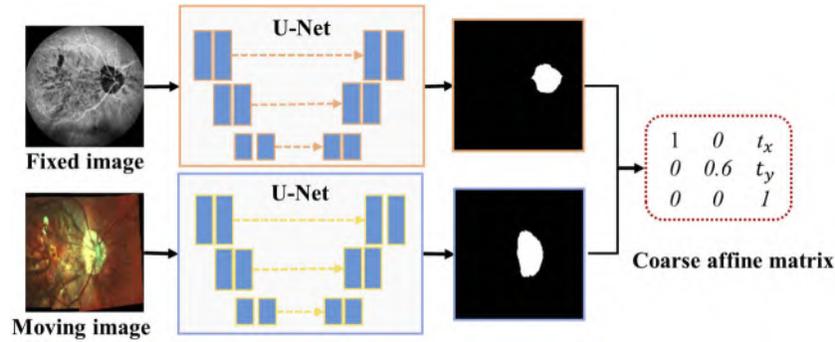


Fig. 3. Coarse registration stage.

The centroid coordinates of the fixed optic disc label (X' , Y') and the moving one (X_2 , Y_2) are used to calculate the coarse registration matrix M_{coarse} . The coarse registration transformation is applied to the moving image and its optic disc label. Equation (8) describes the process of coarse registration, where $t_x = X' - X$ and $t_y = Y' - 0.6Y$ are translation parameters in x and y directions and the 0.6 is the empirical value of the scaling parameter in the height direction.

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 0.6 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (8)$$

1.4. Fine registration

Figure 4 presents the structure of the fine registration network, which consists of two parts: feature extraction layers and regression layers. The feature extraction layers are implemented based on Resnet18 without classification layer. The regression layers contain four cascaded fully connected layers after the last global average pooling layer of Resnet18, with 512, 256, 64

and 6 output neurons, respectively. Leaky-Relu activation function is used after the first three fully connected layers to increase the network's non-linearity. The input of the network is the concatenation of the image pair on the channel, and the output is the regression of six parameters of the affine matrix denoted by the last fully connected layer.

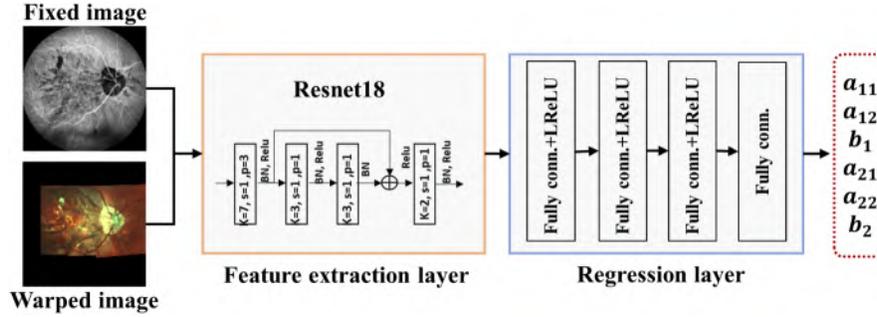


Fig. 4. The network of fine registration stage.

In order to reduce the network's dependence on the ground truth parameters and further improve the registration performance, the optic disc label information is used as an auxiliary supervision and the dual-supervised loss functions are adopted to optimize the fine registration network. Equation (9) shows the loss function of the network, which composes of root mean square error (RMSE) loss and Dice loss.

$$L_{dual} = \lambda_1 L_{rmse} + \lambda_2 L_{dice} \quad (9)$$

Where $\lambda_1 = \lambda_2 = 1$.

The network uses the RMSE loss function L_{rmse} shown in Eq. (10) to evaluate the difference between the ground truth affine parameters and the predicted affine parameters in the form of mini-batch data.

$$L_{rmse}(M_{gt}, \hat{M}) = \sqrt{\frac{1}{b} \sum_{i=1}^b \|v(M_{gt}^i) - v(\hat{M}^i)\|_2^2} \quad (10)$$

where b represents batch size, $v(\cdot)$ is the vector form of matrix, and M_{gt}^i and \hat{M}^i represent i^{th} ground truth affine parameter matrix and the corresponding predicted affine parameter matrix in a batch, respectively.

The predicted affine registration parameters are applied to the coarsely warped optic disc label l_m^w to obtain the registered label l_m^{reg} by spatial resampling. The Dice loss for l_m^{reg} and the fixed label l_f is shown in Eq. (11):

$$L_{dice}(l_f, l_m^{reg}) = 1 - 2 \frac{|l_f \cap l_m^{reg}|}{|l_f| + |l_m^{reg}|} \quad (11)$$

1.5. Implementation details

All the experiments were implemented with PyTorch on a Linux server running Ubuntu 16.04, with Intel Core i7-8700 CPU and 8 GB RAM. The networks were trained on a single NVIDIA GeForce GTX1060 GPU with 3 GB RAM. The initial learning rate is set to 1e-3 in the coarse registration stage with the SGD optimizer and Poly strategy. The learning rate of the fine registration stage is 1e-3, and the optimizer is the Adam. Batch size b is set as 16.

2. Experiments and results

2.1. Dataset

The collection and analysis of image data were approved by the Institutional Review Board of Shanghai General Hospital and adhered to the tenets of the Declaration of Helsinki. An informed consent was obtained from each subject. The medical records, ICGA (Heidelberg Retina Angiography 2, Heidelberg Engineering, Heidelberg, Germany, 596 X 496 pixels) and MCSL (SPECTRALIS, Heidelberg Engineering, Heidelberg, Germany, from 596 X 496 pixels to 960 X 496 pixels) database of Shanghai General Hospital from July 2018 to June 2019 were searched and reviewed. Totally 117 pairs of ICGA and MCSL images from 112 eyes (85 patients) were included, including 102 with linear lesions (such as Fig. 1(a) and (c)) and 15 without linear lesions (such as Fig. 1(b) and (d)). The heights of the original MCSL images are fixed at 496, while the widths vary from 596 to 960 because of different view angles. Previous study [49] reported that linear lesions are hypofluorescent in the late ICGA phase, which is 15 minutes after ICG dye injection. In the late phase, blood vessels have different morphologies and appear as bright white (such as Fig. 1(a) and (b)) or dark gray (such as Fig. 1(c) and (d)). In order to reduce the computational cost, all images are resampled to (256, 256), grayed and normalized to [0, 1]. Online data augmentation was applied during training, including rotation $[-5^\circ, 5^\circ]$, translation $[-6, 6]$ and scaling [0.9, 1.2]. Five-fold cross validation was adopted to evaluate the performance of the proposed method. All data were randomly split into five parts according to the subjects and initial registration errors, which contain 23, 23, 23, 24 and 24 image pairs.

Ground Truth. Under the supervision of the experienced ophthalmologists, three pairs of key points including the intersections and bifurcations of blood vessels are manually selected in the ICGA and the corresponding MCSL image. The ground truth of affine parameters are calculated by three pairs of key points. The ground truth of optic disc is manually labeled under the supervision of the experienced ophthalmologist.

2.2. Evaluation of optic disc segmentation

In the coarse registration stage, the optic disc is segmented based on the original U-Net both in ICGA and MCSL images. The indexes including IoU (Intersection over Union), Dice coefficient, sensitivity, specificity and accuracy of the optic disc segmentation are shown in Table 1, which indicate the good performance and ensure the feasibility of optic disc centroid calculation in the coarse registration stage and the prior image feature based weakly-supervision in the fine registration stage.

Table 1. The cross-validation performance of optic disc segmentation, measured with mean and standard deviation.

Dataset	IoU	Dice	Sensitivity	Specificity	Accuracy
ICGA	0.851 ± 0.083	0.914 ± 0.009	0.928 ± 0.063	0.997 ± 0.004	0.994 ± 0.004
MCSL	0.892 ± 0.104	0.894 ± 0.018	0.908 ± 0.102	0.994 ± 0.005	0.988 ± 0.010

2.3. Metrics

To evaluate the performance of the proposed method objectively, the RMSE of distance on five key points and Dice similarity coefficient (DCS) [50] and target registration error (TRE) [33] on the optic disc label pair are adopted in this paper. The DSC reflects the overlap degree of the optic disc label pair and the TRE reflects the center distance error of the optic disc label pair. Paired Wilcoxon signed-rank tests (significance level $\alpha_H = 0.05$) are applied to compare medians of the registration results between different methods.

2.4. Ablation experiments

In this section, the effect of two-stage network and the improvement of dual-supervised loss functions are investigated. Table 2 and Table 3 show the results of the ablation experiments, in which “*Baseline*” means single-stage registration (only fine registration network) and single supervision loss function with RMSE loss, “*TS*” means two-stage registration + single supervised loss function of RMSE loss, and “*TD*” means two-stage registration + dual supervised loss function of RMSE loss and Dice loss.

Table 2. The cross-validation registration results of ablation experiments, measured with percentiles [25th, 50th, and 75th].^a

Method	RMSE	DSC	TRE
Bef. reg.	-	-	-
Baseline	[6.09, 8.24, 11.90]	[0.630, 0.739, 0.842]	[4.52, 7.62, 10.98]
TS	[3.53, 4.94, 6.73]	[0.803, 0.868, 0.907]	[2.23, 3.81, 5.10]
TD	[3.13, 4.42, 5.91]	[0.847, 0.888, 0.919]	[1.58, 3.16, 4.30]

^a(Bef. reg. represents before registration; “*TS*” represents Two stage + Single supervised loss function; “*TD*” represents Two stage + Dual supervised loss function.)

Table 3. The cross-validation registration results of ablation experiments, measured with mean and standard deviation.^a

Method	RMSE	DSC	TRE
Bef. reg.	33.29 ± 3.27	0.270 ± 0.05	31.82 ± 3.21
Baseline	9.48 ± 5.34	0.704 ± 0.19	8.398 ± 5.6
TS	5.21 ± 2.32	0.855 ± 0.07	3.77 ± 1.92
TD	4.79 ± 2.18	0.874 ± 0.06	3.14 ± 1.69

^a(Bef. reg. represents before registration; “*TS*” represents Two stage + Single supervised loss function; “*TD*” represents Two stage + Dual supervised loss function.)

As can be seen from Table 2, the proposed “*TD*” framework achieved a median RMSE of 4.42 pixels on five key points with first and third quartiles being 3.13 and 5.91 pixels, a median DSC of 0.888 on optic disc label with first and third quartiles being 0.847 and 0.919, and a median TRE of 3.162 on label centroids with first and third quartiles being 1.58 and 4.30 pixels. More detailed results are summarized in Table 3 and shown in Fig. 5. Compared with original image pair, the “*Baseline*” method significantly decreased the RMSE of key points (p -value < 0.001) and the TRE of optic disc label centroids (p -value < 0.001), and significantly increased the DSC of the optic disc label pair (p -value < 0.001). The “*TS*” method significantly outperformed the “*Baseline*” method on all metrics of RMSE, DSC and TRE with p -values < 0.001. This result indicate that the optic disc label centroid alignment based coarse registration can reduce the initial error effectively and reduce the difficulty of fine registration. The proposed “*TD*” method significantly surpassed “*TS*” method in indexes of RMSE and TRE with p -values < 0.001 and in index of DSC with p -value = 0.0011, which indicates that the addition of auxiliary supervision of optic disc label can further refine the registration result. Specially, the improvement on RMSE index between “*TS*” and “*TD*” in Table 2 and Table 3 seems to be slight. We think the reason may be that the annotated key points are relatively sparse and fine registration (such as small-scaling, translation, and rotation) cannot be reflected well in the RMSE indicator. But as shown in Fig. 6(c) and (d), these fine tunings do improve the overall the registration result effectively.

Figure 6 shows an example of registration results with different methods (the corresponding original image pair are shown in Fig. 1(a) and (e)). It can be seen from Fig. 6 that the overlap

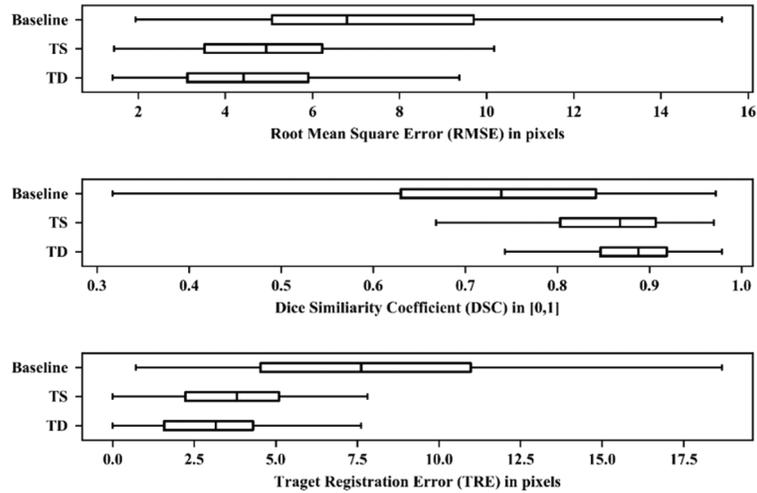


Fig. 5. Boxplot of the cross-validation results obtained from the networks described in section 2. The numerical results are also summarized in Table 2.

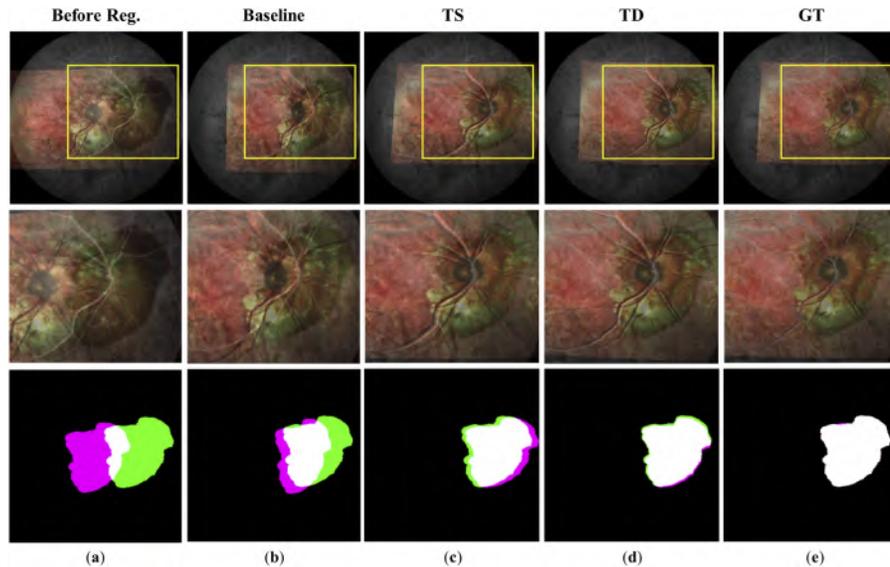


Fig. 6. The example of registration results with different methods. The corresponding original ICGA and MCSL image pair are shown in Fig. 1(a) and (e) respectively. The rows from up to down represent the overlap of the image pair after registration, the corresponding magnified sections in the yellow box and the overlap of the optic disc label pair (white region) after registration respectively. Each column represents the results of different methods: (a) Before registration. (b) Baseline. (c) Two stage + Single supervised loss function (TS). (d) Two stage + Dual supervised loss function (TD). (e) Ground Truth (GT).

degrees of both retinal vessels and optic discs gradually increase from left to right, which indicates that the proposed methods are gradually optimized through two-stage registration and dual-supervision strategies. By comparing with the overlap degree of the optic discs and the corresponding labels in Fig. 6(b) and Fig. 6(c), it can be seen that the coarse-fine registration strategy is effective. By comparing with overlap degree of the retinal vessels and the optic disc labels in Fig. 6(c) and Fig. 6(d), it can illustrate that combining the supervised and weakly-supervised loss function can improve the registration performance finely.

Figure 7 shows some registration results of two-modality image pair with linear lesions. It can be seen from Fig. 7 that the linear lesions in ICGA and MCSL images can be aligned well, which indicate the possibility of non-invasive diagnosis of linear lesion via MCSL imaging.

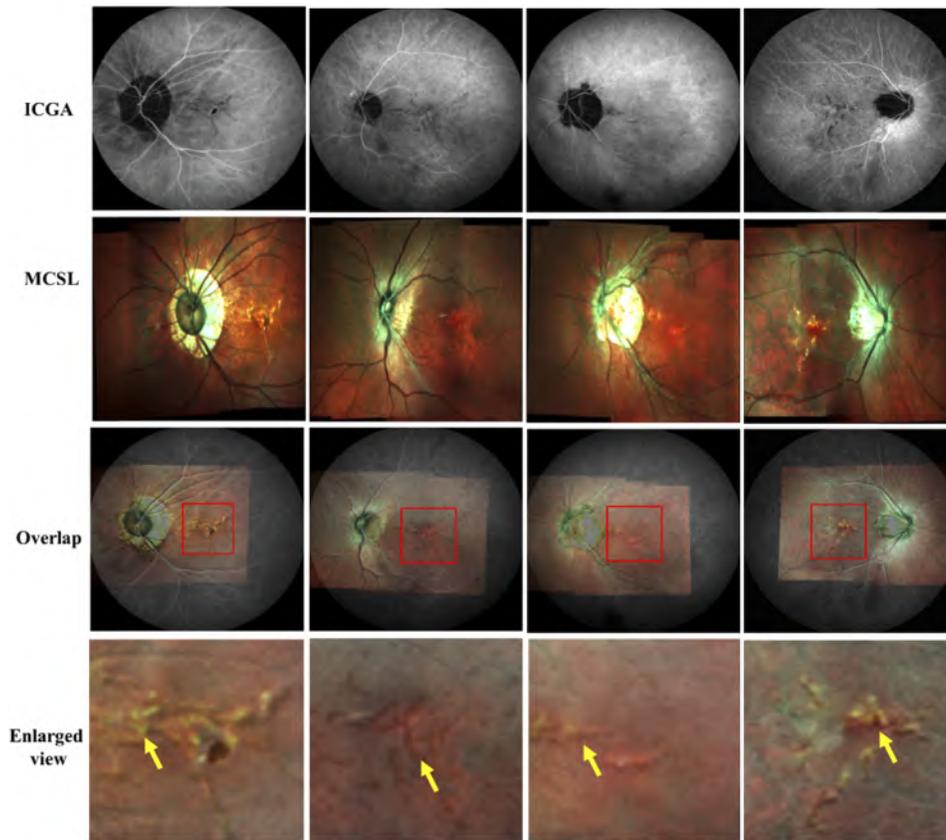


Fig. 7. The registration results of ICGA and MCSL images with linear lesions. The yellow arrows refer to linear lesions. The last row shows the enlarged view of the part of image in the red box.

The influences of model training based on data with or without lesions are also investigated in this section. Besides of the model which has been trained based on the mixed data (data with and without lesions, shown in Table 2 and Table 3 as “TD”), an additional model is trained only based on the abnormal data (102 pairs of images with lesions). Because there are only 15 pairs of normal data (data without lesions), we do not train the network only based on the normal data. Then we validate these two models with the mixed data, abnormal data and normal data, respectively. Table 4 shows the corresponding cross-validation registration results of the ablation experiments.

Table 4. The cross-validation registration results of ablation experiments, measured with percentiles [25th, 50th, and 75th] (the first 6 rows) and mean and standard deviation (the next 6 rows).^a

Model	Validation	RMSE	DSC	TRE
TD-M	Mixed data	[3.13, 4.42, 5.91]	[0.847, 0.888, 0.919]	[1.58, 3.16, 4.30]
	Abnormal data	[3.17, 4.59, 5.98]	[0.855, 0.892, 0.922]	[1.58, 2.96, 4.24]
	Normal data	[3.07, 3.87, 5.31]	[0.804, 0.874, 0.896]	[3.03, 3.60, 4.30]
TD-A	Mixed data	[3.59, 4.73, 6.47]	[0.804, 0.870, 0.910]	[2.24, 3.80, 5.01]
	Abnormal data	[3.49, 4.65, 6.17]	[0.814, 0.870, 0.910]	[2.24, 3.60, 4.94]
	Normal data	[4.47, 6.38, 7.34]	[0.701, 0.811, 0.878]	[2.55, 5.00, 6.28]
TD-M	Mixed data	4.79 ± 2.18	0.874 ± 0.06	3.14 ± 1.69
	Abnormal data	4.84 ± 2.21	0.879 ± 0.06	3.08 ± 1.70
	Normal data	4.54 ± 1.96	0.837 ± 0.08	3.58 ± 1.56
TD-A	Mixed data	5.14 ± 2.15	0.848 ± 0.08	3.84 ± 1.89
	Abnormal data	5.00 ± 2.09	0.857 ± 0.07	3.72 ± 1.78
	Normal data	6.10 ± 2.41	0.800 ± 0.12	4.78 ± 2.46

^a(“TD” represents Two stage + Dual supervised loss function. ‘TD-M’ and ‘TD-A’ represent the model trained with mixed data and the model trained with abnormal data, respectively.)

As can be seen from Table 4, the overall performances of the model trained with mixed data (TD-M) are generally better than that of the model trained with abnormal data (TD-A). We think the possible reason is that the relatively large quantity and good diversity of training samples (data with and without lesions) in model TD-M. In addition, on model TD-A, the validation results with normal data is worse than that with abnormal data. The reason may be that model TD-A enables the network to learn and take advantage of the feature of linear lesion, which can not be used in the normal data.

3. Conclusion and discussions

The pathological myopia developed from high myopia and its complications is one of the main causes of blindness worldwide. The timely detection and analysis of linear lesion is necessary and effective for the prevention, supervision and treatment of pathological myopia. In our previous related research [7], an improved cGAN based framework was proposed to segment linear lesions in ICGA images. As mentioned above, ICGA is invasive and a part of patients may suffer from allergic reactions. To solve this problem, our team focuses the study on the possibility evaluation for the replacement between non-invasive MCSL imaging and invasive ICGA imaging in linear lesion diagnosis and analysis. The evaluation conclusion will be drawn according to the results of linear lesion joint segmentation in MCSL and ICGA images, which is our ongoing and challenging research. The MCSL and ICGA registration research in this paper is the necessary premise of the linear lesion joint segmentation and the subsequent evaluation.

In this paper, we propose a deep learning based two-stage registration framework for the registration of ICGA and MCSL images, which contains the coarse registration stage and fine registration stage. The optic disc label information is fully used in both coarse registration and fine registration to increase the robustness and effectiveness of the network. We also combine supervised and weakly-supervised learning strategies to train the fine registration network, which are achieved through RMSE loss and Dice loss of optic disc label, respectively.

There are still some shortcomings in this paper: (1) The quantity of the experimental dataset is insufficient, which only includes 117 image pairs (102 with linear lesions and 15 without linear lesions). The generalization of the proposed registration network can be improved by increasing

the amount of dataset, especially the quantity of data without linear lesions. (2) To simplify the coarse registration stage, the original U-Net is used for the segmentation of the optic disc, whose segmentation error (shown in Table 1) may affect the registration performance gently. The optic disc segmentation accuracy should be further improved based on the improved U-Net or other advanced CNNs so that more information such as edge of optic disc can be fully used in the two-modality image registration. Introducing the adversarial training strategy and using discriminator as a similarity measurement function for multi-modality image registration are also a direction of our next work. (3) Although both ICGA and MCSL use confocal laser scanning imaging technology, the distortion of the retina's natural curvature cannot be unified because of the following two reasons: (a) The ICGA and MCSL images used in our experiments were acquired from two different devices. (b) The wavelength and number of lasers are different (ICGA: 795 nm, MCSL: 488 nm, 515 nm and 820 nm). The affine registration may not be sufficient to model the transformation between ICGA and MCSL. We will explore registration algorithm combining affine transformation and non-rigid transformation to achieve better registration performances in our near future work. On the foundation of further improvement of the registration accuracy, we will use the complementary characteristics of multi-modal information for the non-invasive detection and analysis of linear lesions in high myopia.

Funding

National Key R&D Program of China under Grant (2018YFA0701700); National Natural Science Foundation of China (61622114, 81401472); International Cooperation Project of Ministry of Science and Technology (2016YFE010770).

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. Y. Tano, "Lix Edward Jackson memorial lecture - Pathologic myopia: Where are we now?" *Am. J. Ophthalmol.* **134**(5), 645–660 (2002).
2. M. Ghafour, D. Allan, and W. S. Foulds, "Common causes of blindness and visual handicap in the west of Scotland," *Br. J. Ophthalmol.* **67**(4), 209–213 (1983).
3. T. Tokoro, "On the definition of pathologic myopia in group studies," *Acta Ophthalmol.* **66**(S185), 107–108 (1988).
4. K. Ohno-Matsui, T. Yoshida, S. Futagami, K. Yasuzumi, and M. Mochizuki, "Pachy atrophy and lacquer cracks predispose to the development of choroidal neovascularisation in pathological myopia," *Br. J. Ophthalmol.* **87**(5), 570–573 (2009).
5. X. Xu, Y. Fang, K. Uramoto, N. Nagaoka, K. Shinohara, T. Yokoi, N. Tanaka, and K. Ohno-Matsui, "Clinical features of lacquer cracks in eyes with pathologic myopia," *Retina* **39**(7), 1265–1277 (2019).
6. K. Ohno-Matsui, N. Morishima, M. Ito, and T. Tokoro, "Indocyanine green angiographic findings of lacquer cracks in pathologic myopia," *Jpn. J. Ophthalmol.* **42**(4), 293–299 (1998).
7. H. Jiang, X. Chen, F. Shi, Y. Ma, D. Xiang, L. Ye, J. Su, Z. Li, Q. Chen, and Y. Hua, "Improved cGAN based linear lesion segmentation in high myopia ICGA images," *Biomed. Opt. Express* **10**(5), 2355–2366 (2019).
8. M. Hope-Ross, L. A. Yannuzzi, E. S. Gragoudas, D. R. Guyer, J. S. Slakter, J. A. Sorenson, S. Krupsky, D. A. Orlock, and C. A. Puliafito, "Adverse reactions due to indocyanine green," *Ophthalmology* **101**(3), 529–533 (1994).
9. F. P. Oliveira and J. M. R. Tavares, "Medical image registration: a review," *Comput. Methods Biomech. Biomed. Eng.* **17**(2), 73–93 (2014).
10. M. A. Viergever, J. B. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. W. Pluim, "A survey of medical image registration - under review," *Med. Image Anal.* **33**, 140–144 (2016).
11. B. Keikhosravi, Y. Li, K. W. Liu, and Eliceiri, "Intensity-based registration of bright-field and second-harmonic generation images of histopathology tissue sections," *Biomed. Opt. Express* **11**(1), 160–173 (2020).
12. M. Chen, A. Lang, H. S. Ying, P. A. Calabresi, J. L. Prince, and A. Carass, "Analysis of macular OCT images using deformable registration," *Biomed. Opt. Express* **5**(7), 2196–2214 (2014).
13. B. A. Ardekani, S. Guckemus, A. Bachman, M. J. Hoptman, M. Wojtaszek, and J. Nierenberg, "Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans," *J. Neurosci. Methods* **142**(1), 67–76 (2005).
14. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging* **16**(2), 187–198 (1997).

15. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Proc. - Workshop Pattern Recognit. Appl. Oil Identif.* **32**(1), 71–86 (1999).
16. J. Ashburner, J. L. R. Andersson, and K. J. Friston, "High-dimensional image registration using symmetric priors," *NeuroImage* **9**(6), 619–628 (1999).
17. M. S. Miri, M. D. Abràmoff, Y. H. Kwon, and M. K. Garvin, "Multimodal registration of SD-OCT volumes and fundus photographs using histograms of oriented gradients," *Biomed. Opt. Express* **7**(12), 5252–5267 (2016).
18. H. Chen, Y. He, L. Wei, J. Yang, X. Li, G. Shi, and Y. Zhang, "Polynomial transformation model for frame-to-frame registration in an adaptive optics confocal scanning laser ophthalmoscope," *Biomed. Opt. Express* **10**(9), 4589–4606 (2019).
19. D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (Ieee, 1999), 1150–1157.
20. S. M. Smith and J. M. Brady, "SUSAN - A new approach to low level image processing," *Int. J. Comput. Vis.* **23**(1), 45–78 (1997).
21. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008).
22. F. Zana and J. C. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform," *IEEE Trans. Med. Imaging* **18**(5), 419–428 (1999).
23. G. Wang, Z. Wang, Y. Chen, and W. Zhao, "Robust point matching method for multimodal retinal image registration," *Biomed. Signal Process. Control* **19**, 68–76 (2015).
24. Z. Ghassabi, J. Shanbehzadeh, A. Sedaghat, and E. Fatemizadeh, "An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIFD descriptors," *J. Image Video Proc.* **2013**(1), 25 (2013).
25. C.-L. Tsai, H.-C. Hsu, X.-C. Wu, S.-J. Chen, and W.-Y. Lin, "Accurate Joint-Alignment of Indocyanine Green and Fluorescein Angiograph Sequences for Treatment of Subretinal Lesions," *IEEE J. Biomed. Health Inform.* **21**(3), 785–793 (2017).
26. I. Krizhevsky, G. E. Sutskever, and Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM* **60**(6), 84–90 (2017).
27. P. Ronneberger, T. Fischer, and Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), 234–241.
28. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017).
29. S. Miao, Z. J. Wang, and R. Liao, "A CNN Regression Approach for Real-Time 2D/3D Registration," *IEEE Trans. Med. Imaging* **35**(5), 1352–1363 (2016).
30. J. M. Sloan, K. A. Goatman, and J. P. Siebert, "Learning rigid image registration-utilizing convolutional neural networks for medical image registration," in *Proceedings of International Conference on Bioimaging*, (2018), 89–99.
31. J. Fan, X. Cao, P.-T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," *Med. Image Anal.* **54**, 193–206 (2019).
32. Y. Hu, M. Modat, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Label-driven weakly-supervised learning for multimodal deformable image registration," in *Proceedings of International Symposium on Biomedical Imaging*, (IEEE, 2018), 1070–1074.
33. Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, S. Ourselin, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Weakly-supervised convolutional neural networks for multimodal image registration," *Med. Image Anal.* **49**, 1–13 (2018).
34. G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," *IEEE Trans. Med. Imaging* **38**(8), 1788–1800 (2019).
35. B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (Springer, 2017), 204–212.
36. J. Krebs, H. e Delingette, B. Mailhe, N. Ayache, and T. Mansi, "Learning a Probabilistic Model for Diffeomorphic Registration," *IEEE Trans. Med. Imaging* **38**(9), 2165–2176 (2019).
37. G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Mach. Vis. Appl.* **31**(1-2), 8(2020).
38. Y. Hu, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, T. Vercauteren, J. A. Noble, and D. C. Barratt, "Adversarial deformation regularization for training image registration neural networks," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer, 2018), 774–782.
39. P. Yan, S. Xu, A. R. Rastinehad, and B. J. Wood, "Adversarial Image Registration with Application for MR and TRUS Image Fusion," in *Machine Learning in Medical Imaging*, *Machine Learning in Medical Imaging* (Springer International Publishing, 2018), 197–204.
40. J. Fan, X. Cao, Z. Xue, P.-T. Yap, and D. Shen, "Adversarial Similarity Network for Evaluating Image Alignment in Deep Learning based Registration," *IEEE Trans. Pattern Anal. Mach. Intell.* **11070**, 739–746 (2018).
41. J. Fan, X. Cao, Q. Wang, P.-T. Yap, and D. Shen, "Adversarial learning for mono- or multi-modal registration," *Med. Image Anal.* **58**, 101545 (2019).

42. T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao, "Hi-Net: Hybrid-fusion Network for Multi-modal MR Image Synthesis," *IEEE Trans. Med. Imaging*, DOI: 10.1109/TMI.2020.2975344 (2020).
43. B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Isgum, "A deep learning framework for unsupervised affine and deformable image registration," *Med. Image Anal.* **52**, 128–143 (2019).
44. Z. Shen, X. Han, Z. Xu, and M. Niethammer, "Networks for joint affine and non-parametric image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2019), 4224–4233.
45. S. Zhao, Y. Dong, E. I. Chang, and Y. Xu, "Recursive cascaded networks for unsupervised medical image registration," in *Proceedings of the IEEE International Conference on Computer Vision*, (2019), 10600–10610.
46. Z. Li, F. Huang, J. Zhang, B. Dashtbozorg, S. Abbasi-Sureshjani, Y. Sun, X. Long, and Q. Yu, "B. ter Haar Romeny, and T. Tan, "Multi-modal and multi-vendor retina image registration," *Biomed.* *Opt. Express* **9**(2), 410–422 (2018).
47. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), 770–778.
48. Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep Learning in Medical Image Registration: A Review," arXiv preprint arXiv:1912.12318 (2019).
49. K. Shinohara, M. Moriyama, N. Shimada, Y. Tanaka, and K. Ohno-Matsui, "Myopic stretch lines: Linear Lesions in Fundus of Eyes With Pathologic Myopia That Differ From Lacquer Cracks," *Retina* **34**(3), 461–469 (2014).
50. D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid Image Registration Using Conditional Mutual Information," *IEEE Trans. Med. Imaging* **29**(1), 19–29 (2010).