An unsupervised network for fast microscopic image registration

Chang Shu, Xi Chen, Qiwei Xie, Hua Han

Chang Shu, Xi Chen, Qiwei Xie, Hua Han, "An unsupervised network for fast microscopic image registration," Proc. SPIE 10581, Medical Imaging 2018: Digital Pathology, 105811D (6 March 2018); doi: 10.1117/12.2293264

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

An Unsupervised Network for Fast Microscopic Image Registration

Chang Shu[1,2], Xi Chen[2], Qiwei Xie[2], and Hua Han[2,3,4]

[1]University of Chinese Academy of Sciences, Beijing, China
[2]Institute of Automation, Chinese Academy of Sciences, Beijing, China
[3]School of Future Technology, University of Chinese Academy of Sciences, Beijing, China
[4]The Center for Excellence in Brain Science and Intelligence Technology, CAS, Beijing, China

ABSTRACT

At present, deep learning is widely used and has achieved excellent results in many fields except in the field of image registration, the reasons are two-fold: Firstly all the steps of deep learning should be derivable; nevertheless, the nonlinear deformation which is usually used in registration algorithms is hard to be depicted by explicit function. Secondly, success of deep learning is based on a large amount of labeled data, this is problematic for the application in real scenes. To address these concerns, we propose an unsupervised network for image registration. In order to integrate registration process into deep learning, image deformation is achieved by resampling, which can make deformation step derivable. The network optimizes its parameters directly by minimizing the loss between registered image and reference image without ground truth. To further improve algorithm's accuracy and speed, we incorporate coarse-to-fine multi-scale iterative scheme. We apply our method to register microscopic section images of neuron tissue. Compared with highly fine-tuning method sift flow, our method achieves similar accuracy with much less time.

Keywords: Microscopic image registration, deep learning, unsupervised learning, coarse-to-fine multi-scale iterative scheme.

1. INTRODUCTION

In order to figure out whether there is connection between neurons, we need to know whether synapses exist between them. Automated tape-collecting ultramicrotome scanning electron microscopy[1] (ATUM-SEM) provides a reliable method to observe synapses. Neuron tissue is cut into a series of sections, collected on the tape, and imaged under the scanning electron microscope (SEM). We recover the 3D structure of neuron tissue with the collected serial section images.

The advantages of this approach are that sections can be maintained and reused. However, due to stretching, distortion and folding during the sample preparation, large displacement exists between adjacent images, so this approach needs high-precision image registration method.

When applied on biological tissue microscopic images, traditional image registration methods have encountered following difficulties: Firstly, due to inadequate texture, the extracted features have low discrimination. Secondly, deformation occurred in these images is complicate and difficult to be depicted by explicit function. Thirdly, the computation will grow dramatically when dealing with large images.

Nowadays, deep learning is applied in many research areas rapidly. However its development in image registration is relatively slow, the main reasons are follows: Firstly, success of deep learning is based on a large amount of labeled data; nevertheless, annotating data is problematic in the field of image registration. Secondly, for a long time, deep learning methods can not directly warp the input image to register it, so they need assistance of other methods. Those hybrid approaches can't fully exploit out the potential of deep learning.

To address these concerns, a new deep learning method for image registration needs to be improved from the following aspects: Firstly, it should be able to incorporate image deformation and form an end-to-end architecture. Secondly, it should be trained in unsupervised manner and no longer rely on labeled data.

Medical Imaging 2018: Digital Pathology, edited by John E. Tomaszewski, Metin N. Gurcan,
Proc. of SPIE Vol. 10581, 105811D · © 2018 SPIE · CCC code:
1605-7422/18/$18 · doi: 10.1117/12.2293264

In this work, we design a novel unsupervised convolutional neural network to achieve fast registration for microscopic images. It includes a novel layer to conduct image deformation, and it directly optimizes its parameters by minimizing the loss between the registered image and the reference image without ground truth, which can save the trouble of annotating data. We utilize this unsupervised network to calculate correspondences between two images. According to those correspondences, we warped deformed image to achieve registration.

Compared with highly fine-tuning method siftflow,[2] our method achieves similar accuracy with much less computing time.

2. RELATED WORK

Image registration is very important to neuroscience and clinical studies for normalizing individual subjects to the reference space. Image registration can be mainly divided into three steps: feature extraction, correspondence searching and image deformation. Namely, extracted features are matched across images to get correspondences, and then image deformation is conducted according to those correspondences.

It is critical to select highly discriminative features that can accurately capture the morphological patterns presented in the image patch. Many algorithms like SIFT,[3] FAST[4] and HARRIS[5] are introduced to extract feature. However, features extracted from the region with inadequate texture are not discriminative.

After we extract features from images, we need to count on correspondence searching algorithms to find feature's corresponding relationship across images. Correspondence searching methods can be categorized into sparse correspondence searching and dense correspondence searching. Sparse correspondence searching algorithms[6,7] are mainly based on combination optimization algorithms. They aim at finding optimal combination to achieve least error between corresponding points. The main drawback of these approaches is when feature has low discrimination, there will exist many wrong matches. Dense correspondence searching methods such as opticflow[8] and siftflow[2] are devised to calculate correspondences for all the points on the image. They take into account neighborhood relationship; even though individual feature has low discrimination, its correspondence can be inferred from neighborhood relationship. In recent years, some dense correspondence searching methods based on deep learning were published,[9–11] their performance are similar to traditional methods. Nevertheless, they have difficulty in practical application, because they are trained in supervised manner. This means that they need tremendous annotated data, which is problematic in real scenes.

Image deformation is used to warp image according to correspondences obtained from above steps. Image deformation can be usually seen as two parts: linear transformation and nonlinear transformation. Linear transformation primarily includes rigid transformation, similarity transformation and affine transformation. Those transformation treat image as a whole, but they do not work well with local deformation. Nonlinear transformation can be realized in two ways. One way is: the whole image deformation is determined by the motion of a few pivots. Representative works are thin-plate-spline,[12] b-spline[13] and moving least square.[14] Their computation complexity is high, computation grows dramatically when pivot number increases. The other way is: image warping is conduct by resampling according to dense correspondences. This way is fast, but at the expense of losing local detail. For a long time, registration method[15,16] based on neural network can not achieve warping until spatial transformer[17] was proposed, which can be included into a standard neural network architecture to provide spatial transformation capabilities, making it possible for designing end-to-end unsupervised network tailored to image registration.

Our work utilizes deep learning to calculate spatial transformation between images, and introduces unsupervised learning to avoid the problem of annotating data of real scene. We further apply image meshing and coarse-to-fine multi-scale iterative framework to reduce computational burden and enhance the ability to deal with large displacement.

3. METHOD

We aim to achieve following goals: Firstly, we approximate reference image with a grid, and our goal is to obtain corresponding points of each vertex in the grid in the absence of annotated data. Secondly, in order to reduce amount of calculation, instead of calculating correspondences of grid vertices on the original images directly, it is desirable to do calculation on the image patches. Thirdly, the warp of images needs to be as fast as possible.

In order to achieve unsupervised learning, we add spatial transformer[17] as a layer to our network. It can register deformed image according to spatial transformation obtained by preceding layers. Then the network is able to optimize its parameters by minimizing the loss between registered image and reference image. This process is conducted without ground truth.

In order to achieve calculation on the image patches, it is important to make sure that selected image patches contain the target correspondences. To address this issue, we adopt coarse-to-fine multi-scale iterative method, the position of correspondences calculated at a small scale is used to estimate its position at a larger scale. In this way, even though the position of correspondences is unknown, we can guarantee that they are in the selected image patches.

In order to achieve fast image deformation, correspondences are required to be sparse, and then interpolation is adopted to make them dense. Those dense correspondences are used to do resampling to generate the final registered image. The work flow of our algorithm is shown in Fig. 1.

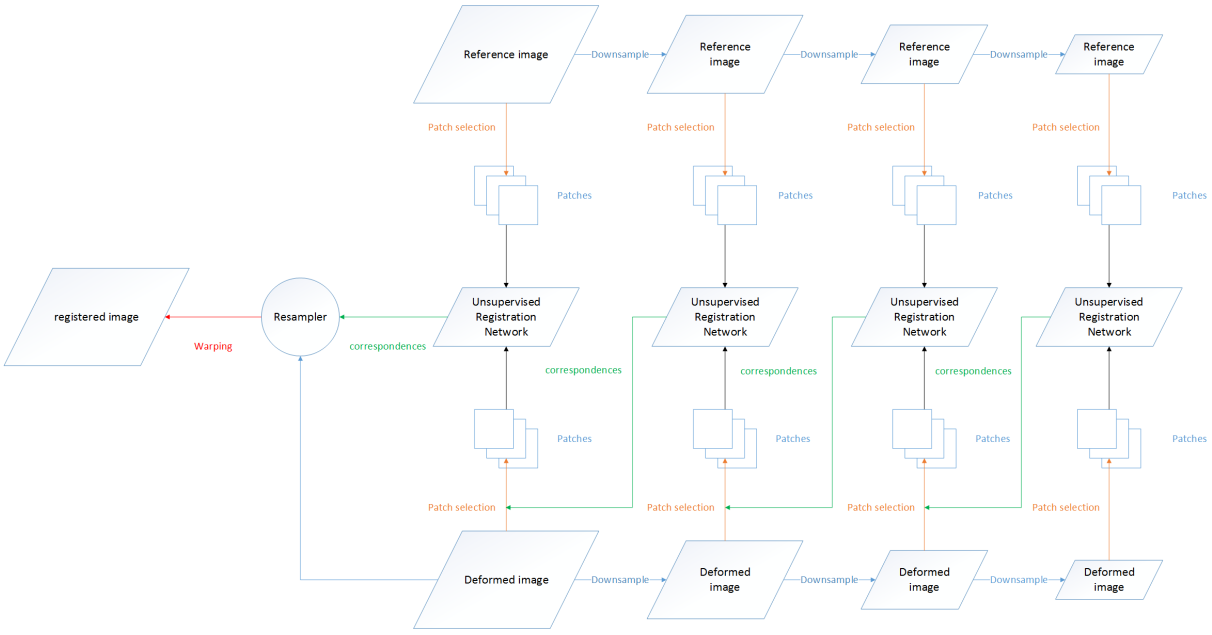


Figure 1. Work flow of proposed method.

To begin with, we approximate the reference image with a grid, then compute the corresponding points of grid vertices on the deformed image. We select an image patch centered at one of grid vertices in the reference image, and select another image patch with the same size in the deformed image. Then they are sent to proposed unsupervised network to compute correspondences. To ensure the image patches selected from the deformed image has the target correspondence, our method is designed under a multi-scale iterative framework. We downsample original images small enough to make sure that the corresponding points are included in the selected image patches at same location. The correspondences obtained from small scale are used to select image patches in the larger scale. When iterating to original scale, we use bicubic interpolation to get dense correspondences from sparse correspondences obtained from above steps. Finally, we use these dense correspondences to warp the deformed image to get final registration result.

Unsupervised network architecture, image patch selection rule, and image deformation approach will be discussed in detail in next subsections.

3.1 Unsupervised Network

The role of unsupervised network is getting correspondences of grid vertices rapidly. The architecture of this network is shown in Fig. 2.

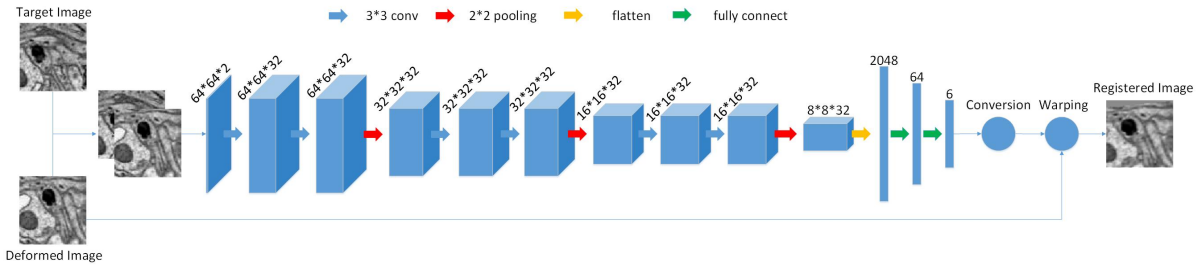


Figure 2. Network Architecture. The thin blue arrow represents input and output. The thick blue arrow denotes convolution layer with size of 3 * 3, and each convolution layer contains 32 convolution kernels and its stride is 1. The red arrow denotes max-pooling layer with size of 2 * 2, the yellow arrow denotes flatten layer, and the green arrow denotes fully connected layer. The blue cuboid denotes feature map, whose size is denoted by the number on its top. The blue circle represents the layer we design.

This network takes concatenated image patches, one is from reference image, the other is from deformed image. The size of input images are 64 * 64. Through the processing of convolution layers and max-pooling layers, we get 6 parameters of affine transformation, we get correspondence's position by (1).

$$\begin{bmatrix} \mathrm{x} \\ y \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

This process is tackled by a layer, which is shown in Fig. 2 and named by 'conversion'. Since equation in (1) is derivable, so this process is able to be integrated in deep learning framework. In order to maintain biological specimens' local morphological patterns, we choose to calculate not nonlinear transformation but linear transformation instead.

After we get correspondences of grid vertices, we need to use them to warp the input image patch. Inspired by spatial transformer,[17] we choose to achieve warping by resampling. Merit of this method is: all the steps are differentiable, only in this way they can be integrated in neural network. Resmapling is achieved by bilinear interpolation, shown in (2).

$$\mathrm{I}_{output}(p) = \sum_{q \in N(p+w)} \mathrm{I}_{input}(q)(1 - |p+\mathrm{w} - q|_2^2) \tag{2}$$

Where I_{output} is output image, I_{input} is input image, p and q are coordinates on the image, and w is the displacement of p, $N(p+w)$ is the set of 4-pixel neighbors (top-left, top-right, bottom-left, bottom-right) of $p+w$.

$$\frac{\partial \mathrm{I}_{output}(p)}{\partial p} = \sum_{q \in N(p+w)} \mathrm{I}_{input}(q)(1 - 2\,(p + w - q)) \tag{3}$$

We can see from (3), output image is derivative with respect to coordinates, guaranteeing that loss can back-propagates[18] to coordinates. This process is conducted by a layer, which is shown in Fig. 2 and named by 'warping'.

Now that we can get deformed image patch registered, we can evaluate loss between it and reference image patch, and utilize this loss to adjust network parameters. When training is done, correspondences will be output. Those correspondences of each grid vertex will used to warp the whole image.

3.2 Image Patch Selection Rule

To ensure the image patches selected from the deformed image has the target correspondence, we select according to a certain rule, Fig. 3 shows an illustration of this rule.

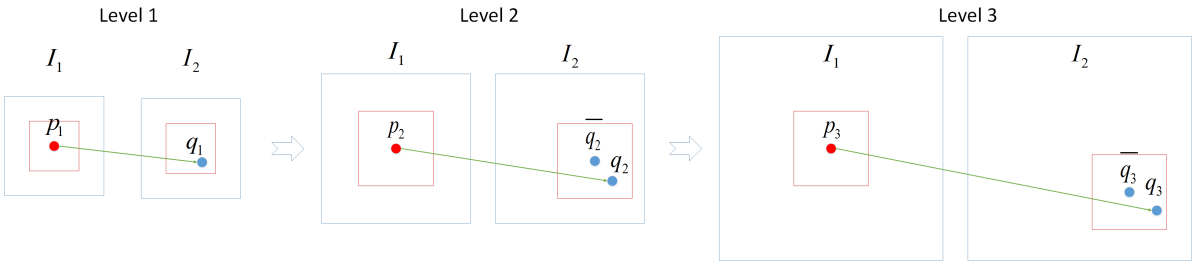


Figure 3. An illustration of coarse-to-fine multi-scale iterative scheme on pyramid. p_k is an grid vertex on the reference image, and q_k is its correspondence on the deformed image. p_{k-1} and p_k refer to the same position at different pyramid level, q_{k-1} and $\overline{q_k}$ also refer to the same position at different pyramid level. The red box is the patch we select.

At first, pyramids of reference image and deformed image are established. A certain pyramid level is smoothed and downsampled from lower pyramid level, and the last pyramid level maintains at original scale.

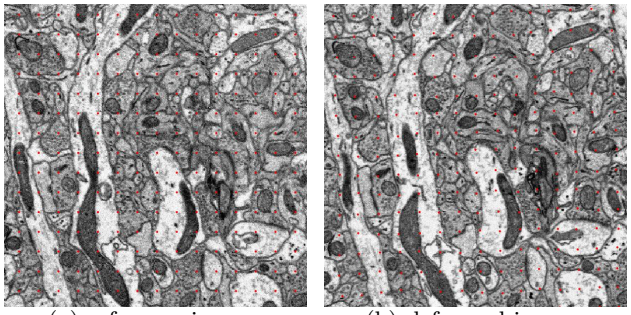
At pyramid level 1, namely the top level, for grid vertex p_1, we take a patch from reference image centered at p_1. If the images in the top level are small enough, the displacement between them will also be small. In this case, grid vertex will not be too far from its corresponding point. Taking image patches at the same location will guarantee target correspondences are included. We put these two patches into our network to get correspondence of p_1, denoted by q_1.

At pyramid level 2, by stretching the coordinates of p_1, we get corresponding position of p_1 at pyramid level 2, denoted by p_2; Similarly, we can get corresponding coordinate of q_1, denoted by $\overline{q_2}$. We take image patches respectively centered at p_2 and $\overline{q_2}$ into our network to get correspondence of p_2, denoted by q_2. Namely, result at previous level is used as initial estimate in the current level.

In a similar way, at pyramid level 3 or more, we repeat the steps at pyramid level 2. After iterating to original scale, we will know accurate correspondences of grid vertices. This procedure is called coarse-to-fine multi-scale iterative scheme, it can help deal with large displacement and decrease computation complexity. It is embedded in dominant dense correspondence searching paradigms like sift flow[2] and optic flow.[8]

3.3 Image deformation

Embedding unsupervised network into above framework, we get accurate correspondences to warp images, shown in Fig. 4.



(a) reference image (b) deformed image

Figure 4. An illustration exhibits the correspondences we find. The red points on the reference image denote grid vertices, and the red points on the deformed image denote correspondences.

After knowing correspondences of all the grid vertices, we use bicubic interpolation to get correspondences of points in between grid vertices. At last, we adopt bilinear interpolation in (2) to achieve registration by applying resampling to the deformed image.

3.4 Summary

We combine unsupervised network, coarse-to-fine multi-scale iterative framework and image deformation into our registration method. Unsupervised network saves the trouble to annotate data. Coarse-to fine multi-scale iterative framework enhances the ability to deal with large displacement. Image meshing and image blocking improve the speed of the method.

4. EXPERIMENTS AND RESULTS

4.1 Data

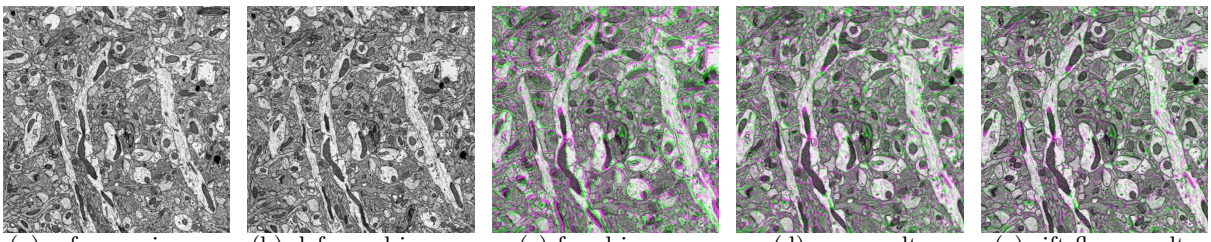
We test the proposed method with serial section images of mouse brain, which contains 71 microscopic images obtained with scanning electron microscope (SEM). The thickness of the section is 50nm, the resolution of the image is 6400 by 6400 pixels, and the pixel size of the image is 2nm. There exists different level of rotation, translation, scaling and nonlinear deformation between adjacent sections due to section and imaging.

4.2 Training

57 images are selected for training, the rest are for testing. Overlapped patch pairs are extracted from adjacent section images. To augment data for network training, we warp one of the patch pair with combination of rotation, translation, shearing and scaling. After data augmentation, our data size increases 100 times. The network is trained with Adam[19] for 2000 iteration, loss function is mean square error and learning rate is set to 0.00001.

4.3 Testing

With the trained network, we test the proposed method and illustrate the results in Fig. 5.



(a) reference image (b) deformed image (c) fused image (d) our result (e) sift flow result

Figure 5. An exhibition of our experimental result. Figure c is fused image of figure a and figure b, figure d is fused image of our registered image and reference image, figure e is fused image of sift flow's registered image and reference image.

We use fused image to show overlap ratio before and after the registration. Different color indicates those non-overlapping regions form different image. The more colorful the fused image is, the less overlap ratio it gets. we choose dice coefficient[20,21] to quantify overlap ratio. It is a statistic used for comparing the similarity of two samples. This metric is computed by comparing the pixel-wise agreement between reference image (Y) and image to be registered (X), illustrated by (4).

$$Dice\text{coefficient} = 2 \cdot \frac{|X \cdot Y|}{|X| + |Y|} \tag{4}$$

Tab. 1 exhibits our method and sift flow method's performances measured by dice coefficient on mouse brain neural tissue image. All the methods are realized on a server equipped with an Intel i7 CPU of 512 GB main memory and a Tesla K40 GPU.

We can see from above that our method performs similarly to sift flow method, but is much faster than it.

Table 1. Quantitative result on test data.

Method	Image Size	Time(s)	Dice Coefficient
sift flow	640 * 640	60.8	0.8952
our method	640 * 640	0.192	0.8951

5. CONCLUSION AND FUTURE WORK

In this article, we propose an unsupervised network for image registration which has seldom been studied, and demonstrate the effectiveness of our method by registering microscopic section images of neuron tissue. Compared with the highly fine-tuning method sift flow, the proposed method can achieve similar accuracy with much less time.

There is still plenty of room for improvement. Mean square error could be replaced with well-designed loss function to evaluate the loss between registered image and reference image. Furthermore, the network can go deeper and adopt more delicate structure.

6. ACKNOWLEDGMENTS

This paper is supported by Scientific Instrument Developing Project of Chinese Academy of Sciences (No.YZ201671), National Science Foundation of China (NO. 61201050) and Special Program of Beijing Municipal Science and Technology Commission (No.Z161100000216146).

REFERENCES

[1] Schalek, R., Kasthuri, N., Hayworth, K., Berger, D., Tapia, J., Morgan, J., Turaga, S., Fagerholm, E., Seung, H., and Lichtman, J., "Development of high-throughput, high-resolution 3d reconstruction of large-volume biological tissue using automated tape collection ultramicrotomy and scanning electron microscopy," *Microscopy and Microanalysis* **17**(S2), 966 (2011).

[2] Liu, C., Yuen, J., and Torralba, A., "Sift flow: Dense correspondence across scenes and its applications," *IEEE transactions on pattern analysis and machine intelligence* **33**(5), 978–994 (2011).

[3] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International journal of computer vision* **60**(2), 91–110 (2004).

[4] Rosten, E. and Drummond, T., "Machine learning for high-speed corner detection," *Computer Vision–ECCV 2006* , 430–443 (2006).

[5] Harris, C. and Stephens, M., "A combined corner and edge detector," in [*Alvey vision conference*], **15**(50), 10–5244, Manchester, UK (1988).

[6] Liu, Z.-Y. and Qiao, H., "Gnccp: Graduated nonconvexity and concavity procedure," *IEEE transactions on pattern analysis and machine intelligence* **36**(6), 1258–1267 (2014).

[7] Maciel, J. and Costeira, J. P., "A global solution to sparse correspondence problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(2), 187–199 (2003).

[8] Brox, T., Bruhn, A., Papenberg, N., and Weickert, J., "High accuracy optical flow estimation based on a theory for warping," *Computer Vision-ECCV 2004* , 25–36 (2004).

[9] Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D., and Brox, T., "Flownet: Learning optical flow with convolutional networks," in [*Proceedings of the IEEE International Conference on Computer Vision*], 2758–2766 (2015).

[10] Weinzaepfel, P., Revaud, J., Harchaoui, Z., and Schmid, C., "Deepflow: Large displacement optical flow with deep matching," in [*Proceedings of the IEEE International Conference on Computer Vision*], 1385–1392 (2013).

[11] Zhou, T., Krahenbuhl, P., Aubry, M., Huang, Q., and Efros, A. A., "Learning dense correspondence via 3d-guided cycle consistency," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 117–126 (2016).

[12] Bookstein, F. L., "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence* **11**(6), 567–585 (1989).

[13] De Boor, C., De Boor, C., Mathématicien, E.-U., De Boor, C., and De Boor, C., [*A practical guide to splines*], vol. 27, Springer-Verlag New York (1978).
[14] Schaefer, S., McPhail, T., and Warren, J., "Image deformation using moving least squares," in [*ACM transactions on graphics (TOG)*], **25**(3), 533–540, ACM (2006).
[15] Miao, S., Wang, Z. J., Zheng, Y., and Liao, R., "Real-time 2d/3d registration via cnn regression," in [*13th International Symposium on Biomedical Imaging (ISBI)*], 1430–1434, IEEE (2016).
[16] Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., and Comaniciu, D., "An artificial agent for robust image registration," in [*AAAI*], 4168–4175 (2017).
[17] Jaderberg, M., Simonyan, K., Zisserman, A., et al., "Spatial transformer networks," in [*Advances in Neural Information Processing Systems*], 2017–2025 (2015).
[18] Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al., "Learning representations by back-propagating errors," *Cognitive modeling* **5**(3), 1 (1988).
[19] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).
[20] Dice, L. R., "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).
[21] Sørensen, T., "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons," *Biol. Skr.* **5**, 1–34 (1948).