Non-rigid registration of temporal live cell microscopy image sequences using deep learning

Yin, Wenzhe, Chagin, Vadim , Cardoso, M. Cristina, Rohr, Karl

Wenzhe Yin, Vadim O. Chagin, M. Cristina Cardoso, Karl Rohr, "Non-rigid registration of temporal live cell microscopy image sequences using deep learning," Proc. SPIE 12032, Medical Imaging 2022: Image Processing, 120321B (4 April 2022); doi: 10.1117/12.2611440

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Non-Rigid Registration of Temporal Live Cell Microscopy Image Sequences Using Deep Learning

Wenzhe Yin[a], Vadim O. Chagin[b,c], M. Cristina Cardoso[b], and Karl Rohr[a]

[a]Biomedical Computer Vision Group, BioQuant, IPMB, Heidelberg University, Germany
[b]Cell Biology and Epigenetics, Dept. of Biology, Technical University of Darmstadt, Germany
[c]Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

ABSTRACT

Registration of microscopy images is an important task in biomedical applications. We introduce a deep learning approach for non-rigid registration of cell nuclei in temporal microscopy image sequences. First, we present a segmentation-based registration method which combines different transformation models and can cope with strong image intensity changes. Second, as an extension, we propose a joint segmentation and registration method which includes a cycle consistency loss and automatically determines the segmentation. Both methods do not need labeled data for network training. The methods were applied to live cell microscopy images of cell nuclei and yield better results than baseline methods.

Keywords: Microscopy Images, Image Registration, Deep Learning

1. INTRODUCTION

Image registration is of central importance in biomedical applications and required for spatial normalization. For live cell microscopy images, a main task is to determine the relative motion of subcellular structures (e.g., DNA replication structures) with respect to a cell. To achieve this, each frame of a temporal image sequence needs to be registered to a reference frame, before performing tracking of subcellular structures. However, image registration is challenging due to complex cell deformation and strongly varying image intensities.

In previous work, different classical methods were introduced for non-rigid registration of *cell microscopy* images (e.g., Ref. 1–4). Yang *et al.*[1] proposed an intensity-based non-rigid registration approach, and Tektonidis *et al.*[2] introduced a non-rigid multi-frame approach. Sorokin *et al.*[3] presented a contour-based method using a dynamic non-linear elasticity model. Gao and Rohr[4] proposed a global optical flow model. However, these classical methods iteratively solve an optimization problem, which is time-consuming.

Recently, there is a trend using deep learning for image registration. Such methods were used for *natural* images (e.g., Ref. 5–7) or *medical* images (e.g., Ref. 8–14). Jaderberg *et al.*[5] introduced a differentiable deformer network with an image resampling layer for convolutional neural networks (CNN), which was applied to the MNIST handwritten digit dataset. Rocco *et al.*[6] proposed a deep learning approach for semantic matching of natural images from the PF-PASCAL dataset (e.g., persons, bikes, cars)[15] using affine transformations and Thin-Plate Splines, where the two transformations are trained separately. Chen *et al.*[7] improved this method for images from the same dataset by jointly training segmentation and matching, but the segmentation result is not used for matching. Balakrishnan *et al.*[8] proposed an unsupervised learning framework for registration of MR images that predicts a dense deformation field. Hu *et al.*[9] introduced a CNN to register ultrasound and MR images of the prostate using anatomical labels for training. Elmahdy *et al.*[10] employed adversarial learning for registration of prostate CT data using ground truth segmentations for training. Xu *et al.*[11] used a joint semi-supervised training scheme for registration and segmentation of MR images, where the two networks can benefit each other. de Vos *et al.*[12] proposed an unsupervised method based on a CNN and B-splines to register chest CT and cardiac MR data. Mok *et al.*[13] performed unsupervised diffeomorphic registration of MR images using

Further author information: (Send correspondence to Wenzhe Yin and Karl Rohr)
Wenzhe Yin: E-mail: logicyin@gmail.com
Karl Rohr: E-mail: k.rohr@uni-heidelberg.de

Medical Imaging 2022: Image Processing, edited by Olivier Colliot, Ivana Išgum, Proc. of SPIE Vol. 12032, 120321B · © 2022 SPIE
1605-7422 · doi: 10.1117/12.2611440

a deep Laplacian pyramid network and an intensity-based similarity metric. Gu *et al.*[14] used both pair-wise and group-wise consistency to improve the registration consistency of MR data. None of these methods was applied to live cell microscopy images which have different challenges compared to natural and medical images.

In this contribution, we introduce a novel deep learning method for non-rigid registration of cell nuclei in temporal live cell microscopy images. First, we present a segmentation-based method for registration which combines different transformation models and can cope with strong intensity changes by using segmented images. Compared to the previous method in Ref. 6, which was applied to natural images, we use optical flow as a basic transformer, combine different transformers, and train them jointly. Second, we extend this method and propose a joint segmentation and registration method, which has the advantage that the segmentation is determined automatically. The method directly exploits the image intensities and includes a cycle consistency constraint to improve the result. The two tasks of segmentation and registration are trained jointly, and registration is based on the segmentation result. Most closely related to our method are the methods in Ref. 7, 10, 11 which were applied to natural images and CT or MR data. However, these methods do not automatically determine the segmentation or registration is not based on the segmentation during inference. In addition, previous work did not consider live cell microscopy images, which have different challenges compared to natural and medical images. An advantage of our proposed methods is that labeled data of registration transformations is not required. To our best knowledge, we are the first that employ deep learning for non-rigid registration of live cell microscopy image sequences. We performed a quantitative performance evaluation, which showed that our methods yield better results than baseline methods.

2. METHODS

The proposed deep learning method for non-rigid registration of temporal microscopy images of cell nuclei does not require labeled data of registration transformations (correspondences) for training. Instead, we generate suitable synthetic data and exploit this data for training. Below, we first introduce a segmentation-based registration method. Then, as an extension, we present a joint segmentation and registration method. Finally, we describe the used technique for generating synthetic data.

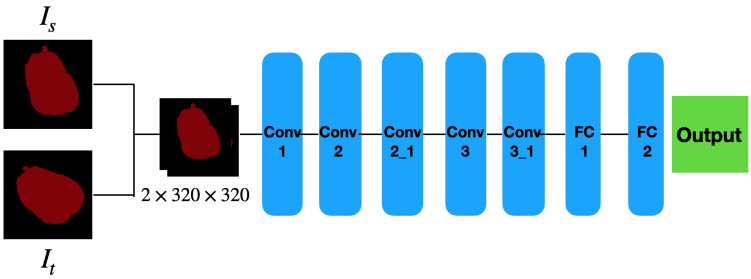

Figure 1. Segmentation-based deep learning network.

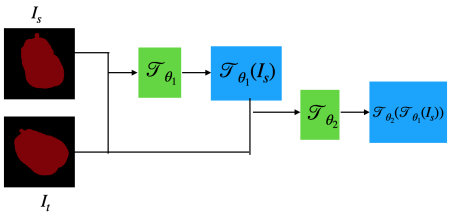

Figure 2. Model cascade.

2.1 Segmentation-Based Registration

The proposed segmentation-based deep learning method for non-rigid registration of temporal microscopy images uses different transformation models (affine, spline-based, optic flow-based). The network architecture is shown in Fig. 1. The network consists of 5 convolution layers and 2 fully connected layers. For affine transformations, the network output vector consists of 6 parameters. For spline-based transformations (Thin-Plate Splines[16,17] or Gaussian), we use regular grid points of the source image as the control points (parameters). Then, given a pair of segmented images and a set of n control points of the source image, the transformed control points for the target image are predicted by the network. The output dimension of the network is $2n$. Denoting the transformer by $\mathcal{T}$, the transformation parameters by θ, and all m points of the source image by G_s, then the transformed points for the target image can be represented by $G_t = \mathcal{T}_\theta(G_s)$, where the ground truth parameters are denoted by θ_{GT}. The aim is to estimate the parameters θ so that the transformation loss $\frac{1}{m}||\mathcal{T}_\theta(G_s) - \mathcal{T}_{\theta_{GT}}(G_s)||_2^2$ is minimal. For optic flow-based transformations (FlowNet[18]), the transformation loss is computed analogously.

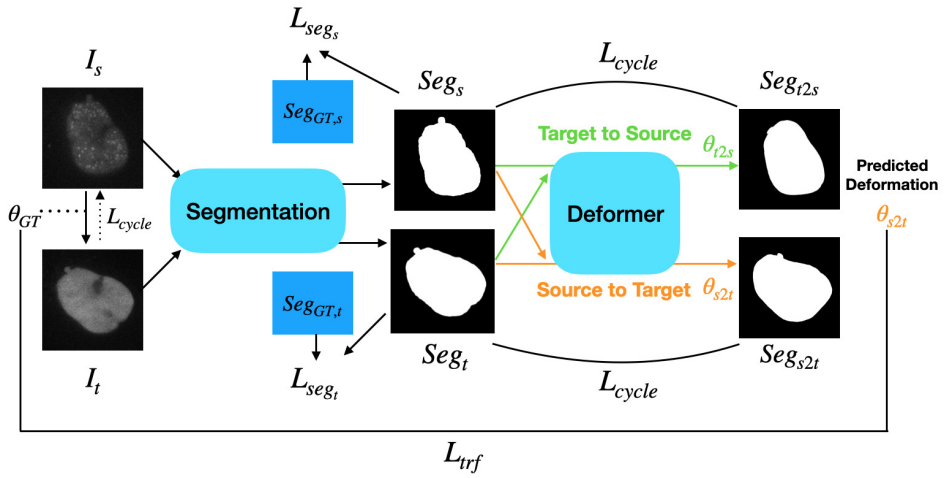

Figure 3. Overview of joint segmentation and registration method.

We combine the transformation models (affine, Thin-Plate Splines,[16,17] Gaussian, FlowNet[18]) in a model cascade and train them simultaneously. We use the following combinations: Affine-FlowNet, TPS-FlowNet, and Gaussian-FlowNet. For the combined models, we apply the second transformation $\mathcal{T}_{\theta_2}$ to the transformed source image from the first transformation $\mathcal{T}_{\theta_1}$, leading to the transformation loss for the model cascade: $\frac{1}{m}||\mathcal{T}_{\theta_2}(\mathcal{T}_{\theta_1}(G_s)) - \mathcal{T}_{\theta_{GT}}(G_s)||_2^2$, where $\mathcal{T}_{\theta_1}$ and $\mathcal{T}_{\theta_2}$ can be any of the used transformation models (Fig. 2).

2.2 Joint Segmentation and Registration

A disadvantage of the segmentation-based non-rigid registration method described in Sect. 2.1 is that prior segmentation of the images is necessary. Therefore, we extend this method and introduce a novel method which performs joint segmentation and registration. This method directly exploits the image intensities and has the advantage that the segmentation is determined automatically. The segmentation network and the registration network are trained jointly. We also integrate a cycle-consistency constraint for registration to improve the performance. An overview of the joint segmentation and registration method is provided in Fig. 3. For an image pair $\{I_s, I_t\}$, the segmentation network determines the segmentation heat maps $\{Seg_s, Seg_t\}$, which are used for registration and for computing the cycle consistency loss. For the *segmentation loss*, we use the binary cross entropy (BCE) between the segmentation heat maps of the source as well as the target image and the corresponding ground truth segmentations:

$$L_{seg} = BCE(Seg_s, Seg_{GT,s}) + BCE(Seg_t, Seg_{GT,t}). \tag{1}$$

Based on the automatic segmentation result determined by thresholding the heat map from the segmentation network, the transformer $\mathcal{T}$ is used to register the source image with the target image yielding the transformation parameters θ_{s2t}. The *transformation loss* is defined by:

$$L_{trf} = \frac{1}{m}||\mathcal{T}_{\theta_{s2t}}(G_s) - \mathcal{T}_{\theta_{GT}}(G_s)||_2^2. \tag{2}$$

In our method, we perform registration bidirectionally, i.e. we additionally transform the target to the source image, which yields the parameters θ_{t2s}. Based on this, we define a *cycle consistency loss* for the segmentation results:

$$\begin{aligned} L_{cycle} = \frac{1}{wh}(||Seg_t - Seg_{s2t}||_2^2 + ||Seg_s - Seg_{t2s}||_2^2) \\ + \frac{\alpha}{wh}(||I_t - I_{s2t}||_2^2 + ||I_s - I_{t2s}||_2^2), \end{aligned} \tag{3}$$

where w and h are the width and height of the image, and α is a weighting parameter. Seg_{s2t} and Seg_{t2s} are the transformed source and target segmentation heat maps, respectively. I_{s2t} and I_{t2s} denote the intensities of the transformed source and target images, respectively. Thus, the loss exploits both the segmentation heat maps and the image intensity information. The *overall loss* is then defined by:

$$L = L_{seg} + \lambda_1 L_{trf} + \lambda_2 L_{cycle}, \tag{4}$$

where λ_1 and λ_2 are used to weight the different terms.

For the segmentation network model, we use the U-Net.[19] An image is fed into 4 convolution layers with maxpooling layers. Then, four deconvolution layers are used to upsample the feature map and predict the final segmentation, where skip-connections are used to keep the low-level information. For registration, we employ the same network architecture as for the method in Sect. 2.1.

2.3 Generation of Training Data and Network Training

Training deep neural networks generally requires large amount of labeled data, which is not available in our application. Therefore, we use synthetic images which are generated by random geometric deformations.[6] This provides us a large number of image pairs with ground truth for training.

Given a source image I_s from the live cell microscopy image sequences, we randomly generate a set of transformation parameters and employ a transformer on I_s to obtain a target image I_t with the ground truth transformation parameters θ_{GT}. As transformer we use Thin-Plate Splines (TPS).[16,17] Given the coordinates of n control points in the source image (x_i, y_i), $i = 1, 2, ..., n$, we randomly generate the corresponding transformed coordinates $(x_i^{'}, y_i^{'})$, $i = 1, 2, ..., n$ (we used n=25). Then the parameters of TPS are determined by solving a linear system of equations, and the transformed target image is computed by interpolation. In total, we generated more than 100,000 image pairs.

The generated image data and transformations are used for network training. In all our experiments, we apply the Adam optimizer with a learning rate of 0.001. For TPS and Gaussian transformations we use 5×5 control points. The registration and segmentation networks are trained simultaneously.

3. RESULTS

We performed a quantitative evaluation of the proposed methods based on temporal live cell microscopy image sequences of cell nuclei acquired by a spinning disk confocal microscope.[20] The data is challenging since the image intensities change strongly over time, strong cell nucleus deformations occur, and the image noise is high. The temporal 2D microscopy image sequences consist of 7-13 frames. We trained our methods on seven sequences and used two sequences for testing. We resized each frame to 320×320 pixels (from originally 350×350 pixels) by bi-linear interpolation since the network requires the image size to be divisible by 16. Segmentations are obtained by thresholding the heat maps determined by the segmentation network (we used a threshold of 0.5). For the overall loss in (4) we used the weights $\lambda_1 = 1.0$ and $\lambda_2 = 0.1$. To quantify the performance, we registered all frames of an image sequence to the first frame (reference frame), and calculated the average Intersection over Union (IoU) score over all frames before and after registration.

The results of the *segmentation-based registration* method (Sect. 2.1) are given in Table 1. It can be seen that among the single transformation models, TPS achieves the best result and yields a relatively high accuracy. Among the combined transformation models (model cascade), TPS-FlowNet yields the best result. The registration accuracy of both TPS and TPS-FlowNet is much better compared to the unregistered case.

The results of the *joint segmentation and registration* method (Sect. 2.2) are provided in Table 2. Thin-Plate Splines were used as transformation model because of the superior performance in our previous experiment (Table 1). We investigated TPS without cycle consistency loss (TPS-Joint) and with cycle consistency loss (TPS-Joint-Cycle). The performance was quantified by the segmentation IoU between the predicted segmentation and the ground truth segmentation, and also by the registration IoU between the transformed segmentation and the target segmentation. It can be seen that using the cycle-consistency loss improves the result. It also turns out that the result of the joint segmentation and registration method is similar to that of the segmentation-based

Table 1. Results of segmentation-based registration methods.

Method	Registration IoU
Unregistered	0.865
Affine	0.910
Gaussian	0.929
TPS	**0.951**
Affine-FlowNet	0.950
Gaussian-FlowNet	0.958
TPS-FlowNet	**0.971**

registration method, but has the advantage that the segmentation is determined automatically. An example registration result for two different time points of a microscopy image sequence is displayed in Fig. 4. It can be seen that the automatically determined segmentation result is quite accurate (overlay of the original images in red transparent, see Fig. 4 left) and that the source image is well registered to the target image (represented by the red segmentation contour, see Fig. 4 right).

Table 2. Results of joint segmentation and registration methods.

Method	Segmentation IoU	Registration IoU
Unregistered	-	0.865
TPS-Joint	0.968	0.938
TPS-Joint-Cycle	**0.972**	**0.952**

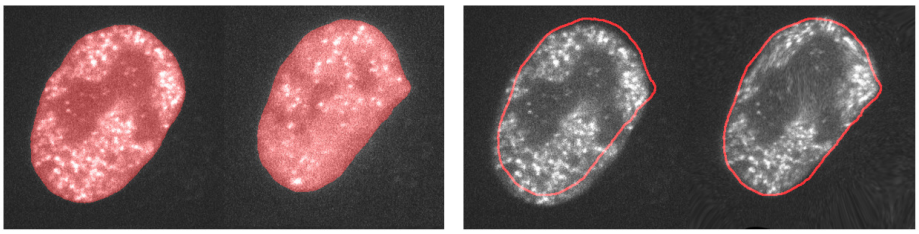

Figure 4. Result of joint segmentation and registration method. (Left) Original images (time points 1 and 5) with overlaid segmentations in transparent red. (Right) Unregistered and registered source image overlaid with the automatically determined segmentation contour of the target image.

4. CONCLUSION

We proposed a novel deep learning method for non-rigid registration of cell nuclei in temporal live cell microscopy image sequences. First, we described a segmentation-based registration method. Second, we extended this method and introduced a joint segmentation and registration method which automatically determines the segmentation and integrates a cycle consistency constraint. Both methods do not need labeled data due to utilizing synthetic data for training. The experiments demonstrate that the proposed methods yield better results than baseline methods, and that the cycle consistency constraint improves the registration result.

ACKNOWLEDGMENTS

Support of the DFG (German Research Foundation) within the SPP 2202 (RO 2471/10-1, CA 198/15-1) is gratefully acknowledged.

REFERENCES

[1] Yang, S., Köhler, D., Teller, K., et al., "Nonrigid registration of 3-d multichannel microscopy images of cell nuclei," *IEEE TIP* **17**(4), 493–499 (2008).

[2] Tektonidis, M. et al., "Non-rigid multi-frame registration of cell nuclei in live cell fluorescence microscopy image data," *Medical Image Analysis* **19**(1), 1–14 (2015).

[3] Sorokin, D. V. et al., "Non-rigid contour-based registration of cell nuclei in 2-d live cell microscopy images using a dynamic elasticity model," *IEEE TMI* **37**(1), 173–184 (2017).

[4] Gao, Q. and Rohr, K., "A global method for non-rigid registration of cell nuclei in live cell time-lapse images," *IEEE TMI* **38**(10), 2259–2270 (2019).

[5] Jaderberg, M., Simonyan, K., Zisserman, A., et al., "Spatial transformer networks," in [*Advances in Neural Information Processing Systems*], 2017–2025 (2015).

[6] Rocco, I. et al., "Convolutional neural network architecture for geometric matching," in [*Proc. CVPR*], 6148–6157 (2017).

[7] Chen, Y. C. et al., "Show, match and segment: Joint weakly supervised learning of semantic matching and object co-segmentation," *IEEE TPAMI* **43**, 3632–3647 (2021).

[8] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V., "Voxelmorph: a learning framework for deformable medical image registration," *TMI* **38**(8), 1788–1800 (2019).

[9] Hu, Y., Modat, M., et al., "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical Image Analysis* **49**, 1–13 (2018).

[10] Elmahdy, M. S. et al., "Adversarial optimization for joint registration and segmentation in prostate CT radiotherapy," in [*Proc. MICCAI*], 366–374, Springer (2019).

[11] Xu, Z. and Niethammer, M., "DeepAtlas: Joint semi-supervised learning of image registration and segmentation," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 420–429, Springer (2019).

[12] de Vos, B. D. et al., "A deep learning framework for unsupervised affine and deformable image registration," *Medical Image Analysis* **52**, 128–143 (2019).

[13] Mok, T. and Chung, A., "Large deformation diffeomorphic image registration with Laplacian pyramid networks," in [*Proc. MICCAI*], 211–221, Springer (2020).

[14] Gu, D. et al., "Pair-wise and group-wise deformation consistency in deep registration network," in [*Proc. MICCAI*], 171–180, Springer (2020).

[15] Ham, B. et al., "Proposal flow: Semantic correspondences from object proposals," *IEEE TPAMI* **40**(7), 1711–1725 (2018).

[16] Bookstein, F. L., "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE TPAMI* **11**(6), 567–585 (1989).

[17] Rohr, K. et al., "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Trans. on Medical Imaging* **20**(6), 526–534 (2001).

[18] Dosovitskiy, A. et al., "FlowNet: Learning optical flow with convolutional networks," in [*Proc. ICCV*], 2758–2766 (2015).

[19] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional networks for biomedical image segmentation," in [*Proc. MICCAI 2015*], 234–241.

[20] Chagin, V. O. et al., "4d visualization of replication foci in mammalian cells corresponding to individual replicons," *Nature Comm* **7**(1), 1–12 (2016).