2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-2923-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISBI52829.2022.9761507

DENOISEREG: UNSUPERVISED JOINT DENOISING AND REGISTRATION OF TIME-LAPSE LIVE CELL MICROSCOPY IMAGES USING DEEP LEARNING

Kerem Celikay[1], *Vadim O. Chagin*[2,3], *M. Cristina Cardoso*[2], *and Karl Rohr*[1]

[1]Biomedical Computer Vision Group, BioQuant, IPMB, Heidelberg University and DKFZ, Germany
[2]Cell Biology and Epigenetics, Department of Biology, Technical University of Darmstadt, Germany
[3]Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

ABSTRACT

Image registration is important for analysing time-lapse live cell microscopy images. However, this is challenging due to significant image noise and complex cell movement. We propose a novel end-to-end trainable deep neural network for joint denoising and affine registration of temporal live cell microscopy images. Our network is trained unsupervised, and only a single network is required for both tasks which reduces overfitting. Our experiments show that the proposed network performs better than deep affine registration without denoising, and better than sequential deep denoising and affine registration. In combination with deep non-rigid registration, we outperform state-of-the-art non-rigid registration methods.

Index Terms— Biomedical Imaging, Image Registration, Denoising, Fluorescence Microscopy Images, Deep Learning

1. INTRODUCTION

The analysis of time-lapse temporal microscopy images of cell nuclei plays a crucial role to better understand biological mechanisms such as DNA replication, DNA repair, nucleoli assembly or viral defence. In particular, the motion of particles within the cell nucleus is important for cellular processes. However, the motion of these particles is composed of their own local motion and a global movement of the cell nucleus. This superposition makes image analysis difficult. To compensate for the movement of the cell nucleus, image registration can be used, which aligns each frame of an image sequence with the first frame and allows determining the local motion of nuclear particles.

In previous work on the registration of temporal microscopy images, often classical rigid and affine registration methods were used (e.g., [1–5]). To cope with local deformations, classical non-rigid registration methods were proposed (e.g., [6], [7]). However, performing non-rigid registration directly on the original images often does not yield good results, since, for example, translation or rotation of cells generally lead to unrealistic deformations. Thus, pre-alignment by rigid or affine registration is necessary. In addition, classical methods often require high computation time.

Recently, deep learning approaches have been proposed, especially for medical images (MR, CT), but not for temporal live cell microscopy images. Since annotating data is difficult, [8] proposed a *supervised* affine registration network for MR images, which is trained using synthetic data. However, synthetic data generally differs from real data, which can reduce the registration accuracy. Alternatively, *unsupervised* methods were proposed for affine [9] and non-rigid registration (e.g., [9], [10]) of MR and CT images. However, these methods do not perform image denoising, which is important for live cell microscopy images since they are degraded by strong noise.

Denoising methods can be used as a preprocessing step to improve the image quality (e.g., [11]). In [12] a classical method for combined denoising and registration of MR images using graph cuts was described. Recently, unsupervised deep learning methods were proposed for denoising and image reconstruction of MR images, which use registration to exploit information from several images [13], [14]. However, the goal of these methods is denoising or image reconstruction, but not registration. In addition, separate networks for denoising and registration were used. Moreover, these methods were applied to MR images, but not to temporal microscopy images, which pose different challenges.

Contribution. In this work, we propose a novel end-to-end trainable deep neural network named *DenoiseReg* for joint denoising and affine registration of temporal microscopy image sequences. *DenoiseReg* is trained *unsupervised* and thus does not require annotated data. Another advantage is that only a single neural network needs to be trained compared to using separate networks for denoising and registration. This reduces overfitting and improves practical applicability. We applied our approach to temporal live cell microscopy images. Our experiments show that *DenoiseReg* can cope with the strong noise in microscopy images and achieves better results than deep affine registration without denoising, and better results than sequential deep denoising and registration. In addition, we demonstrate that combining *DenoiseReg* with deep non-rigid registration performs better than state-of-the-art non-rigid registration methods.

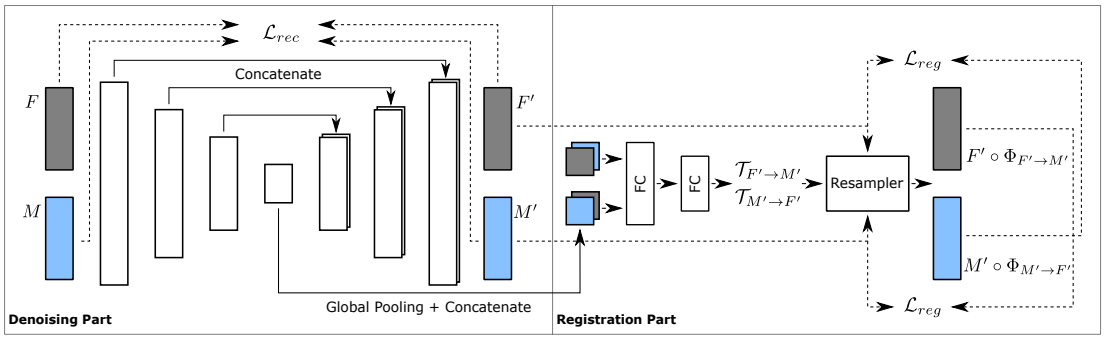


Fig. 1. *DenoiseReg* performs joint denoising and affine registration of a fixed image F and a moving image M. The parameters are optimized by jointly minimizing the reconstruction loss $\mathcal{L}_{rec}$ and the registration loss $\mathcal{L}_{reg}$. The white rectangles represent feature maps obtained by a convolutional block, and *FC* denotes a fully connected layer. The resampler transforms the denoised image F' and M' using the transformation matrix $\mathcal{T}$.

2. METHODS

The proposed network *DenoiseReg* performs joint denoising and registration of successive images of temporal microscopy image sequences. The network consists of a denoising part and a registration part which are trained simultaneously (Fig. 1).

For the *denoising* part of *DenoiseReg*, we use a blind-spot network [15]. Given a noisy image X defined over an n-dimensional domain $\Omega \subset \mathbb{R}^n$, randomly chosen pixels are removed using a masking scheme, generating the masked input image $\tilde{X}$. The network f predicts the intensity values of the masked pixels by only considering the remaining ones in the local neighbourhood. It is assumed that the noise is pixel-wise independent but the actual signal is not. This yields a denoised image $X' = f(\tilde{X})$. The used *reconstruction loss* is defined by the L1-norm for the masked pixels

$$\mathcal{L}_{rec}(X, X', m) = \frac{1}{|\Omega|} \sum_{\substack{p\in\Omega: \\ m(p)=1}} |X'(p) - X(p)|, \tag{1}$$

with m being the binary mask specifying which pixels have been excluded, and p denoting the position. The denoising network learns a feature representation of the local neighbourhood to predict the intensity value of a masked pixel. For *DenoiseReg*, we suggest exploiting the feature representations of two images of an image sequence (fixed image F and moving image M) for registration. For F and M the same denoising network is used. The denoising part of *DenoiseReg* consists of an encoding path and a decoding path with skip connections for better gradient flow (Fig. 1 left part). In contrast to the U-Net [16], we use a convolutional block which consists of two convolutional layers (kernel size 3), each followed by a normalization layer and a ReLU activation function. Afterwards, the feature maps are downsampled using max pooling (kernel size 2), while the number of feature maps are doubled. In the encoding path, we downsample three times using 16 up to 64 feature maps. Subsequently, in the decoding path, we use nearest neighbour interpolation for upsampling, followed by a convolutional block and the number of feature maps are halved.

The *registration* part builds upon [9], where the features of the last encoder block from the denoising network of the input images F and M are first downsampled to a spatial dimension of 1 using global average pooling (Fig. 1 right part). Afterwards, the downsampled features are concatenated and fed into fully connected layers, where each layer is followed by a normalization layer and a ReLU activation function. The order of concatenation determines which image is registered to which. The last fully connected layer outputs the affine transformation parameters τ that are added to the identity transformation matrix to align the denoised images F' and M'. In the 3D case, the transformation matrix $\mathcal{T}_{M'\to F'}$ is given by

$$\mathcal{T}_{M'\to F'} = \begin{pmatrix} 1+\tau_{11} & \tau_{12} & \tau_{13} & \tau_{14} \\ \tau_{21} & 1+\tau_{22} & \tau_{23} & \tau_{24} \\ \tau_{31} & \tau_{32} & 1+\tau_{33} & \tau_{34} \end{pmatrix}, \tag{2}$$

and comprises 12 parameters for rotation, scaling, shear, and translation. In the 2D case there are 6 parameters. To prevent the untrained network from predicting too large values for the transformation parameters, the predicted values are scaled (we used 0.1). Instead of using the negative normalized cross correlation as in [9], we define the *registration loss* as the mean-squared-difference of the intensities between the transformed image M' and the fixed image F'

$$\mathcal{L}_{reg}(F', M' \circ \Phi_{M'\to F'}) = \frac{1}{|\Omega|} \sum_{p\in\Omega} \left[F'(p) - \left[M' \circ \Phi_{M'\to F'}\right](p)\right]^2, \tag{3}$$

where $\Phi_{M'\to F'}$ is the displacement field generated by $\mathcal{T}_{M'\to F'}$. We block the gradients of $\mathcal{L}_{reg}$ with respect to the denoised images F' and M' to ensure that the weight update does not affect the denoising part which otherwise might lead to strong

smoothing of the image structures in order to minimize $\mathcal{L}_{reg}$.

Finally, the total loss is defined by the sum of the reconstruction losses in (1) for the input images F and M and the registration losses in (3) between the denoised images F' and M' computed bidirectionally (i.e. $M' \rightarrow F'$ and $F' \rightarrow M'$):

$$\mathcal{L} = \frac{1}{2}\left[\mathcal{L}_{rec}(F, F', m) + \mathcal{L}_{rec}(M, M', m)\right] + \frac{1}{2}\left[\mathcal{L}_{reg}(F', M' \circ \Phi_{M' \rightarrow F'}) + \mathcal{L}_{reg}(M', F' \circ \Phi_{F' \rightarrow M'})\right]. \tag{4}$$

DenoiseReg is trained for $300,000$ iterations using the Adam optimizer [17] ($\beta_1 = 0.9, \beta_2 = 0.999$) with a learning rate of 0.0001. For pre-processing, we standardize the images to have zero mean and unit variance, and use histogram matching to cope with intensity variations. To increase the variation of the considered small datasets, we perform data augmentation by applying random affine transformations in every iteration with an execution probability of 0.3.

3. EXPERIMENTAL RESULTS

Data. Dataset A consists of two 2D microscopy image sequences of U2OS cell nuclei with an image size of 532×532 pixels and a sequence length of 25 (A1) and 38 (A2) frames [18]. Dataset B comprises four 2D image sequences of HeLa cells with an image size of about 400×300 pixels and sequence lengths between 30 and 42 frames [18]. Dataset C consists of nine 3D image sequences which shows replication foci of HeLa cell nuclei during S-phase [19]. The sequences have a length between 5 and 13 frames and the image size is about 27×200×175 voxels. All datasets were acquired with a confocal microscope. Datasets A and B have been randomly split into 80% for training, 10% for validation and 10% for testing. Since Dataset C is small, we used all images for training. Note that we do not use ground truth transformations for training but only the original images. Thus, our approach is unsupervised. To select the best model from training, the geometric registration error using key points between corresponding image pairs in the validation set is employed.

Evaluation. For evaluation we use manually annotated key points of image structures in datasets A and B. We determine the geometric registration error as the Euclidean distance between each point $p_{i,t}$ at time point t and its initial position $p_{i,1}$ at the first time point:

$$e_{i,t} = ||p_{i,1} - p_{i,t}||_2, \tag{5}$$

where i denotes the i-th key point. For dataset C we show only qualitative results, since defining consistent key points for evaluation is difficult.

Quantitative results for Dataset A are presented in Table 1. *DenoiseReg* always yields much better results than deep affine registration without denoising [9] and outperforms sequential deep denoising and deep affine registration (Deep Denoise + Deep Affine), see the average values. For Dataset B (Table 2), *DenoiseReg* performs better than deep affine registration without denoising and better than sequential deep denoising and deep affine registration (according to the contour points and the interior points). This confirms that denosing is important and that using a single network for both tasks (i.e. using shared parameters) improves the generalizability. By combining *DenoiseReg* with deep non-rigid registration (VoxelMorph (VM) [10]), we outperform classical state-of-the-art non-rigid registration methods [6], [7] for both datasets (except for Dataset B with interior points, for which we obtain very similar results as [7]). Directly using VM (without prior denoising and affine registration) yields much worse results. This shows that denoising and pre-alignment are important for deep learning-based non-rigid registration of live cell microscopy images in order to cope with the strong noise and the movement of the cell nucleus. To further demonstrate the importance of denoising, we performed affine registration of the original images using the transformation parameters obtained by *DenoiseReg* and trained VM using the original images (denoted as *DenoiseReg** + VM). This improves the result compared to only using *DenoiseReg*, however, the best result is obtained by combining *DenoiseReg* with VM.

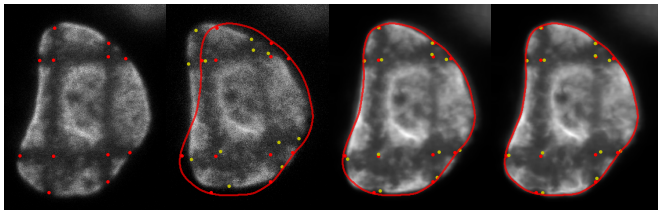
Sequence	A1				A2			
Annotator	H1	H2	H3	Avg.	H1	H2	H3	Avg.
Unregistered	30.17	30.75	31.29	30.74	49.42	50.20	49.05	49.55
Cont. based [6]	5.71	7.18	5.78	6.22	7.95	9.10	8.78	8.61
Opt. flow based [7]	5.69	6.47	**4.26**	5.47	6.45	**6.78**	7.86	7.03
Deep Affine [9]	35.84	34.74	36.25	35.61	27.14	26.94	26.99	27.03
VM [10]	30.32	30.92	31.52	30.92	49.22	50.04	48.95	49.40
Deep Denoise + Deep Affine	7.98	8.74	7.56	8.09	8.71	10.43	11.16	10.10
DenoiseReg (Ours)	6.90	7.86	7.22	7.33	9.53	10.10	10.47	10.03
DenoiseReg* + VM	6.79	7.81	7.08	7.23	9.39	9.98	10.52	9.96
DenoiseReg (Ours) + VM	**4.05**	**5.14**	4.45	**4.55**	**6.14**	6.85	**7.27**	**6.75**

Table 1. Registration error for Dataset A.

Contour Points					
Sequence	B1	B2	B3	B4	Avg.
Unregistered	18.15	6.40	9.30	8.13	10.50
Cont. based [6]	3.69	2.02	2.17	2.43	2.58
Opt. flow based [7]	3.36	**1.42**	1.82	1.86	2.12
Deep Affine [9]	4.68	3.15	3.11	2.07	3.25
VM [10]	10.02	2.73	4.60	3.21	5.14
Deep Denoise + Deep Affine	4.86	3.03	3.13	2.01	3.26
DenoiseReg (Ours)	4.76	3.07	3.05	2.05	3.23
DenoiseReg* + VM	3.18	2.11	2.31	1.62	2.31
DenoiseReg (Ours) + VM	**2.87**	1.85	**1.77**	**1.52**	**2.00**
Interior Points					
Sequence	B1	B2	B3	B4	Avg.
Unregistered	11.30	5.71	8.00	6.79	7.95
Cont. based [6]	7.92	5.22	2.76	3.04	4.73
Opt. flow based [7]	**6.71**	4.99	**2.35**	**2.85**	**4.22**
Deep Affine [9]	7.42	5.56	3.03	3.31	4.83
VM [10]	7.83	4.84	5.44	3.23	5.34
Deep Denoise + Deep Affine	7.32	5.55	3.02	3.33	4.81
DenoiseReg (Ours)	7.46	5.39	2.94	3.18	4.74
DenoiseReg* + VM	6.95	5.02	2.78	3.04	4.45
DenoiseReg (Ours) + VM	6.77	**4.74**	2.37	3.02	4.23

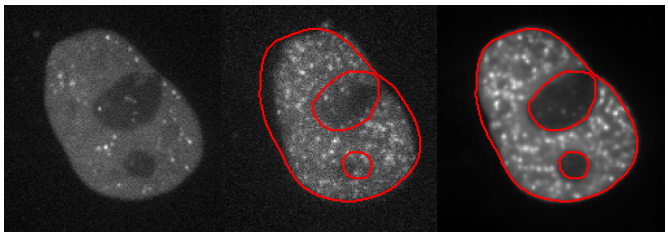
Table 2. Registration error for Dataset B.

Qualitative results for Dataset B are shown in Fig. 2. *DenoiseReg* reduces the noise while retaining the content of the image. The contour lines and the marked key points indicate that *DenoiseReg* registers the images reasonably well. The results for Dataset C in Fig. 3 show that *DenoiseReg* can successfully register 3D images.



(a) 1st frame (b) Unregistered last frame (c) *DenoiseReg* last frame (d) *DenoiseReg* + VM last frame

Fig. 2. Registration results of *DenoiseReg* for dataset B (red: contour and key points of first frame, yellow: key points of last frame).



(a) 1st frame (b) Unregistered last frame (c) *DenoiseReg* last frame

Fig. 3. Registration results of *DenoiseReg* for dataset C (slice of a 3D image). The red lines show contours of the first frame. Note that due to 3D registration, the slices in (b) and (c) do not correspond.

4. CONCLUSION

We introduced *DenoiseReg*, a novel deep learning method for joint denoising and affine image registration. We evaluated our method on 2D and 3D temporal microscopy images and showed that the proposed network performs better than deep affine registration without denoising and better than sequential deep denoising and deep registration. In combination with deep non-rigid registration, we outperformed state-of-the-art non-rigid registration methods.

Acknowledgements. Support of the DFG within the SPP 2202 (RO 2471/10-1, CA 198/15-1) and the BMBF project de.NBI (031A537C, HD-HuB) is gratefully acknowledged.

Compliance with Ethical Standards. This work is a study for which no ethical approval was required.

5. REFERENCES

[1] A. Goobic, J. Tang, and S. Acton, "Image stabilization and registration for tracking cells in the microvasculature," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 2, pp. 287–299, 2005.

[2] D. Sage, F. Neumann, F. Hediger, S. Gasser, and M. Unser, "Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1372–1383, 2005.

[3] P. Matula, M. Kozubek, and V. Dvorák, "Fast point-based 3-D alignment of live cells," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2388–2396, 2006.

[4] C. Wilson and J. Theriot, "A correlation-based approach to calculate rotation and translation of moving cells," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1939–1951, 2006.

[5] M. van de Giessen, A. van der Laan, E. Hendriks, M. Vidorreta, J. Reiber, C. Jost, H. Tanke, and B. Lelieveldt, "Fully automated attenuation measurement and motion correction in FLIP image sequences," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 461–473, 2011.

[6] D. Sorokin et al., "Non-rigid contour-based registration of cell nuclei in 2-D live cell microscopy images using a dynamic elasticity model," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 173–184, 2017.

[7] Q. Gao and K. Rohr, "A global method for non-rigid registration of cell nuclei in live cell time-lapse images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2259–2270, 2019.

[8] E. Chee and Z. Wu, "Airnet: Self-supervised affine registration for 3d medical images using neural networks," *arXiv:1810.02583*, 2018.

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

[10] G. Balakrishnan, A. Zhao, M. Sabuncu, J. Guttag, and A. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[11] W. Meiniel, J. Olivo-Marin, and E. Angelini, "Denoising of microscopy images: a review of the state-of-the-art, and a new sparsity-based method," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3842–3856, 2018.

[12] H. Lombaert and F. Cheriet, "Simultaneous image de-noising and registration using graph cuts: Application to corrupted medical images," in *Proc. ISSPA*. IEEE, 2012, pp. 264–268.

[13] J. Xu and E. Adalsteinsson, "Deformed2Self: Self-Supervised Denoising for Dynamic Medical Imaging," *arXiv:2106.12175*, 2021.

[14] W. Gan, Y. Sun, C. Eldeniz, J. Liu, H. An, and U. Kamilov, "MoDIR: Motion-Compensated Training for Deep Image Reconstruction without Ground Truth," *arXiv:2107.05533*, 2021.

[15] A. Krull, T. Buchholz, and F. Jug, "Noise2Void-Learning denoising from single noisy images," in *Proc. CVPR*, 2019, pp. 2129–2137.

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

[17] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[18] V. Foltánková, P. Matula, D. Sorokin, S. Kozubek, and E. Bártová, "Hybrid detectors improved time-lapse confocal microscopy of PML and 53BP1 nuclear body colocalization in DNA lesions," *Microscopy and Microanalysis*, vol. 19, no. 2, pp. 360–369, 2013.

[19] V. Chagin, C. Casas-Delucchi, and M. Reinhart et al., "4D Visualization of replication foci in mammalian cells corresponding to individual replicons," *Nature Communications*, vol. 7, no. 1, pp. 1–12, 2016.