A Hybrid Deep Learning Framework for Integrated Segmentation and Registration: Evaluation on Longitudinal White Matter Tract Changes

Bo Li[1,2](✉), Wiro J. Niessen[2,3], Stefan Klein[2], Marius de Groot[2], M. Arfan Ikram[2], Meike W. Vernooij[2], and Esther E. Bron[2]

[1] Northeastern University, Shenyang, China
[2] Erasmus MC, Rotterdam, The Netherlands
b.li@erasmusmc.nl
[3] Delft University of Technology, Delft, The Netherlands

Abstract. To accurately analyze changes of anatomical structures in longitudinal imaging studies, consistent segmentation across multiple time-points is required. Existing solutions often involve independent registration and segmentation components. Registration between time-points is used either as a prior for segmentation in a subsequent time point or to perform segmentation in a common space. In this work, we propose a novel hybrid convolutional neural network (CNN) that integrates segmentation and registration into a single procedure. We hypothesize that the joint optimization leads to increased performance on both tasks. The hybrid CNN is trained by minimizing an integrated loss function composed of four different terms, measuring segmentation accuracy, similarity between registered images, deformation field smoothness, and segmentation consistency. We applied this method to the segmentation of white matter tracts, describing functionally grouped axonal fibers, using N = 8045 longitudinal brain MRI data of 3249 individuals. The proposed method was compared with two multistage pipelines using two existing segmentation methods combined with a conventional deformable registration algorithm. In addition, we assessed the added value of the joint optimization for segmentation and registration separately. The hybrid CNN yielded significantly higher accuracy, consistency and reproducibility of segmentation than the multistage pipelines, and was orders of magnitude faster. Therefore, we expect it can serve as a novel tool to support clinical and epidemiological analyses on understanding microstructural brain changes over time.

Keywords: Simultaneous · Segmentation · Deformable registration · Diffusion MRI · White matter tract · CNN · Longitudinal

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11766, pp. 645–653, 2019.
https://doi.org/10.1007/978-3-030-32248-9_72

1 Introduction

In longitudinal imaging studies, the consistency of segmentations can be improved by using methods tailored to longitudinal data [11]. Existing solutions often involve independent registration and segmentation components, which are performed sequentially or iteratively in a multi-stage pipeline. Spatial correspondence established with deformable registration is used either to introduce a prior for segmentation in a subsequent time-point, or to perform segmentation in a common space. We here propose a novel hybrid convolutional neural network (CNN) that optimizes segmentation and registration in a single procedure. We hypothesize that the joint optimization leads to increased performance on both tasks.

This work is one of the first learning-based frameworks for joint optimization of segmentation and registration. Existing methods for joint optimization, e.g., using a Markov Random Field [7] or Expectation Maximization [8] framework, rely on non-learning based registrations, and therefore need to be optimized on test data. In addition, there are two types of work that are closely related but are different from joint optimization. The first type of methods focus on registration-based segmentation [12], e.g., atlas-based segmentation and contour propagation. These methods label the images by registering atlas images to the data to be segmented. An example of segmentation-based registration is [4]. The second type of methods aim at improving segmentation with predefined registration. For instance, deep learning-based segmentation methods apply pre-estimated transformations to introduce labels for a weakly supervised task [10]. An example of improving registration with pre-segmented images can be found in [1]. In contrast to the above methods, our hybrid method does not require online-optimization or any prepared label and transformation. Segmentation and registration are performed at the same time resulting in a segmented structure, a transformation between images, and a deformed image.

We propose the hybrid CNN in a general and cross-sectional manner in Sect. 2. The method is demonstrated on the segmentation of white matter tracts, describing functionally grouped axonal fibers, using a large diffusion-weighted MRI (DWI) dataset. We evaluate its performance in a longitudinal setting with multiple time-points per individual. In Sect. 3, we compare the hybrid method with two multistage pipelines, and assess the added value of joint optimization for segmentation and registration separately. A discussion of our method and the results can be found in Sect. 4.

2 Hybrid Method

Let $I(x)$ be an input image, $S(x)$ be its segmentation label and $x \in R^3$ denote the spatial coordinates. In structure segmentation, parameters $\boldsymbol{\Theta}$ for a function $\mathcal{F}_{\boldsymbol{\Theta}}$ are estimated such that

$$S = \mathcal{F}_{\boldsymbol{\Theta}}\big(I\big). \tag{1}$$

For registration of images, a transformation $\mathcal{T}$ is applied to an image (source image, I_s) to optimally fit another image (target image, I_t), i.e., for $\forall x \in R^3$, $I_t(x)$ and $I_s\big(\mathcal{T}(x)\big)$ correspond to a same anatomical location:

$$I_t(x) \approx I_s\big(\mathcal{T}(x)\big). \tag{2}$$

The transformation can be written as $\mathcal{T}_{t,s}(x)$, or in short as $\mathcal{T}(x)$. $\mathcal{T}(x)$ includes both global (rigid, affine) and local (non-rigid) transformations. For any pair of inputs, $\mathcal{T}(x)$ is estimated by a shared function $\mathcal{G}_{\boldsymbol{\Phi}}$, i.e., $\mathcal{T}_{t,s}(x) = \mathcal{G}_{\boldsymbol{\Phi}}(I_t, I_s)$. Parameters $\boldsymbol{\Phi}$ for $\mathcal{G}_{\boldsymbol{\Phi}}$ are globally optimized on all training data.

In this work, we integrate the parameters $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ in a single hybrid CNN. The overview of our method is illustrated in Fig. 1. We describe the loss function and network architecture in following paragraphs.

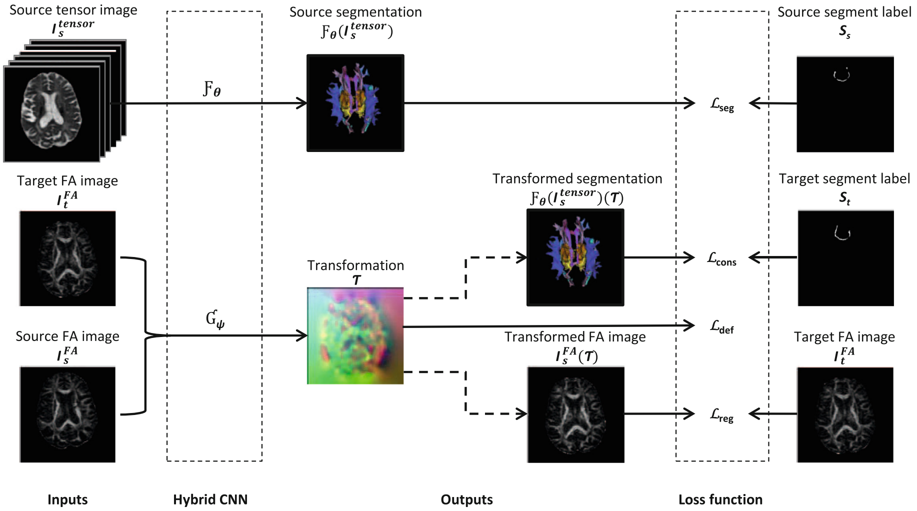

Fig. 1. Overview of the method. $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ denote the parameters of the segmentation and registration task, respectively. Both the source tensor and FA images were estimated from the same diffusion-weighted MRI scan. The loss function consists of $\mathcal{L}_{seg}$, $\mathcal{L}_{cons}$, $\mathcal{L}_{def}$ and $\mathcal{L}_{reg}$ terms.

Loss Function. The loss function for optimization of $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ is composed of four terms that measure segmentation accuracy ($\mathcal{L}_{seg}$), similarity between registered images ($\mathcal{L}_{reg}$), deformation field smoothness ($\mathcal{L}_{def}$) and segmentation consistency ($\mathcal{L}_{cons}$), respectively, i.e.,

$$\begin{aligned}\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Phi},\boldsymbol{\Theta}}{\operatorname{argmin}}\{&\mathcal{L}_{seg}\big(S_s, \mathcal{F}_{\boldsymbol{\Theta}}(I_s)\big) + \alpha\mathcal{L}_{reg}\big(I_t, I_s(\mathcal{T})\big) + \beta\mathcal{L}_{def}(\mathcal{T}) \\ &+ \gamma\mathcal{L}_{cons}\Big(S_t, \mathcal{F}_{\boldsymbol{\Theta}}(I_s)(\mathcal{T})\Big)\},\end{aligned} \tag{3}$$

where $\boldsymbol{\mathcal{T}}$ depends on $\boldsymbol{\Phi}$. Segmentation accuracy was quantified by the agreement between the predicted segmentation of source images $(\mathcal{F}_{\boldsymbol{\Theta}}(I_s))$ and the segment labels (S_s). Consistency was quantified by the correspondence between the transformed segmentation of source images $(\mathcal{F}_{\boldsymbol{\Theta}}(I_s)\big(\boldsymbol{\mathcal{T}}\big))$ and the segment labels in target space (S_t). The segmentation accuracy and consistency terms were computed using weighted inner product metric [6]. Similarity between the registered images was quantified using mean squared error. Smoothness of the deformation field was encouraged by minimizing a diffusion regularization term on the estimated transformations, which defined as a mean of squares of first derivatives. We set the hyperparameters to $\alpha = 10, \beta = 0.1$ and $\gamma = 1$.

Network Architecture. The hybrid 3D CNN models $\mathcal{F}_{\boldsymbol{\Theta}}$ and $\mathcal{G}_{\boldsymbol{\Phi}}$ in parallel by a series of convolutions and non-linearity operations, using similar U-Net architectures [9] with skip connections. The encoder paths were gradual compression processes of extracting abstract features of the structure for $\mathcal{F}_{\boldsymbol{\Theta}}$, and of estimating global deformations between images for $\mathcal{G}_{\boldsymbol{\Phi}}$. The decoder paths restored the details in segmentation $(\mathcal{F}_{\boldsymbol{\Theta}})$ and refined local deformations $(\mathcal{G}_{\boldsymbol{\Phi}})$ by decompressing features and combining them with the shallow information at the same scales. The convolution layers produced a set of k feature maps by individually convolving the input with k kernels. In this work, we used $k = [16, 32, 64, 128, 256, 128, 64, 32, 16]$ for both $\mathcal{F}_{\boldsymbol{\Theta}}$ and $\mathcal{G}_{\boldsymbol{\Phi}}$. Each convolution layer was of kernel size $(3, 3, 3)$, and followed by a batch normalization and a leaky ReLu layer $(a = 0.2)$ for non-linerities. Resamplings were performed using max-pooling and up-sampling operations.

The last layer of $\mathcal{F}_{\boldsymbol{\Theta}}$ was an softmax function resulting in a posterior probability $P\big(S(x)|\boldsymbol{\Theta}, I(x)\big)$. During performance evaluation, the probabilistic map was binarized with a threshold of $P > 0.5$. The last layer of $\mathcal{G}_{\boldsymbol{\Phi}}$ was a convolution layer with 3 kernels that yielded the transformation $\boldsymbol{\mathcal{T}}$ in x, y and z axes. To apply the estimated $\boldsymbol{\mathcal{T}}$ for deforming images and probabilistic segmentations, we adopted the spatial transformation function used in [1].

Application to DWI. Different DWI-derived metrics were used for the segmentation and registration component of the network (Fig. 1), i.e., the diffusion tensor image (with six components) and fractional anisotropy (FA) image. For other applications, such as structure segmentation based on T1-weighted MRI, $\mathcal{F}_{\boldsymbol{\Theta}}$ and $\mathcal{G}_{\boldsymbol{\Phi}}$ could take the same inputs directly.

3 Experiments and Results

3.1 Material and Preprocessing

Material. The Rotterdam Study (RS) is a prospective and population-based study targeting causes and consequences of age-related diseases [3]. For this study, we used 8045 longitudinal DWI scans of 3249 individuals. The majority of these scans were repeatedly acquired in a time interval of 1–5 years (7770 scans of 3166 individuals), in which changes in brain microstructure are expected owing to aging. These long time-interval scans were matched into 6043 pairs by grouping any two time-points of the same individual. We used 5175 pairs for training (D_{train}), 200 pairs for validation (D_{vali}), and an independent cohort of 668 pairs for testing (D_{test}). The remaining scans were from 112 individuals who were scanned twice within a month, in which no changes in microstructure are expected. After exclusion of 15 individuals who had other visits in D_{train}, we used 97 pairs of short time-interval scans from 97 individuals for a reproducibility dataset (D_{repro}).

MRI Acquisition. Scans were acquired on a 1.5T MRI scanner (GE Signa Excite). DWI was scanned with the following parameters: $TR/TE = 8575\,\text{ms}/82.6\,\text{ms}$; imaging matrix of 64×96 (zero-padded in k-space to 256×256) in a field of view of $210 \times 210\,\text{mm}^2$; 25 diffusion weighted volumes with a b-value of $1000\,\text{s/mm}^2$ and 3 non-weighted volumes. The voxel size was resampled from $3.3 \times 2.2 \times 3.5\,\text{mm}^3$ to $1\,\text{mm}^3$, resulting in an image of $211 \times 210 \times 123$ voxels.

Image Preprocessing. DWI preprocessing [6] included motion and eddy currents correction, diffusion tensors estimation using Levenberg–Marquard optimization, and computation of diffusion tensor imaging (DTI) measures such as FA. For each tract, an ROI was defined by taking the maximum bounding box based on the reference segmentation (section below). For the experiments, the forceps minor (FMI) tract is evaluated because it is functionally significant, i.e., related to aging and dementia, and not easy to be segmented. The ROI size for the FMI tract was $96 \times 64 \times 64$ voxels.

Reference Segmentations. The segmentation labels for model training and evaluation were generated using a published method [2] consisting of probabilistic tractography and atlas information. The resulting tract-density images for each tract were normalized by division with the total number of tracts in the tractography run. Finally, tract-specific thresholds for the normalized density images were established by maximizing the reproducibility of FA measures. In the remainder of the paper, we denote the reference segmentations as S^R.

3.2 Experiments and Evaluation Metrics

Multistage Pipelines. We compared the proposed hybrid method ($*^H$) with two multistage pipelines ($*^{R,E}$ and $*^{N,E}$). First, $*^{R,E}$ denotes the reference segmentation method (S^R) in combination with a conventional registration algorithm using Elastix software ($\boldsymbol{\mathcal{T}}^E$) [5]. Second, $*^{N,E}$ denotes a recent CNN-based segmentation method, Neuro4Neuro (S^N) [6], combined with Elastix. In Elastix (version 4.8), we used the B-spline based non-linear deformation, mutual information as similarity metric and a stochastic gradient descent optimizer with adaptive step size estimation. The paired samples t-test was used to test the statistical significance of below metrics.

Segmentation accuracy was quantified using the Dice coefficient (DC). For Neuro4Neuro, $accuracy^N = DC(S^R, S^N)$. For the hybrid method:

$$accuracy^H = DC\big(S^R, \mathcal{F}^H_{\boldsymbol{\Theta}}(I)\big). \tag{4}$$

For each image pair, the consistency of segmentations was computed using the DC in two directions, i.e., both transforming the baseline to follow-up and transforming the follow-up to the baseline. For multistage pipelines, the transformation $\boldsymbol{\mathcal{T}}$ was estimated by aligning FA images using Elastix, e.g., $consistency^{N,E} = DC\big(S^N_t, S^N_s(\boldsymbol{\mathcal{T}}^E)\big)$. For the hybrid method, we tested the dataset bidirectionally since the segmentation in native space and that transformed from another time-point were available:

$$consistency^H = DC\Big(\mathcal{F}^H_{\boldsymbol{\Theta}}(I_t), \mathcal{F}^H_{\boldsymbol{\Theta}}(I_s)\big(\boldsymbol{\mathcal{T}}^H\big)\Big). \tag{5}$$

Using the dataset D_{repro}, we evaluated the scan-rescan reproducibility on segmentations and on tract-specific FA measures. The agreement of segmentations was quantified using the Cohen's kappa coefficient (κ):

$$\kappa^H = \kappa\Big(\mathcal{F}^H_{\boldsymbol{\Theta}}(I_t), \mathcal{F}^H_{\boldsymbol{\Theta}}(I_s)\big(\boldsymbol{\mathcal{T}}^H\big)\Big). \tag{6}$$

Median of FA measures was computed for voxels inside the tract segmentation and used as the tract-specific measure. For each individual (i), the scan-rescan reproducibility of FA measures (FA_{i1},FA_{i2}) was quantified by computing the error (ϵ, Eq. 7). A lower error indicates a higher reproducibility.

$$\epsilon = \frac{1}{\frac{1}{2}} \frac{|FA_{i1} - FA_{i2}|}{|FA_{i1} + FA_{i2}|} 100\%. \tag{7}$$

Registration and Segmentation Components. To assess the added value of the hybrid approach for registration and segmentation components, we designed separate step-wise experiments.

First, we assessed whether registration improves by joint optimization. We built a pure registration CNN (RegNet) using the same architecture as $\mathcal{G}_{\boldsymbol{\Phi}}$, optimized it with only $\mathcal{L}_{reg}$ and $\mathcal{L}_{def}$ terms [1]. Registration performance was compared between RegNet (independent optimization) and the registration component of hybrid method (joint optimization), and evaluated by registering the same segmentation image S^R, i.e., S^R+RegNet vs S^R+Hybrid. Registration accuracy was quantified by the spatial correspondence of the registered segmentations, i.e., $\mathrm{DC}(S^R_t, S^R_s(\boldsymbol{\mathcal{T}}))$.

Second, we assessed whether segmentation improves by joint optimization. We built a pure segmentation CNN (SegNet) using the same architecture as $\mathcal{F}_{\boldsymbol{\Theta}}$, optimized it with only $\mathcal{L}_{seg}$ term. Segmentation performance was compared between SegNet (independent optimization) and the segmentation component of hybrid method (joint optimization), i.e., SegNet+Elastix vs Hybrid+Elastix.

Segmentation accuracy was measured using the DC with reference segmentation (S^R). The consistency was evaluated by the spatial correspondence of the segmentations registered using Elastix, i.e., DC($S_t, S_s(\boldsymbol{\mathcal{T}}^E)$).

Implementation. The experiments of model training and evaluation were performed on an NVIDA 1080Ti GPU and an AMD 1920X CPU. CNN-based methods were implemented using Keras-2.2.0 with a Tensorflow-1.4.0 backend. Models were trained using the Adam optimizer with an initial learning rate of $1e^{-4}$.

3.3 Results

The segmentation accuracy, consistency and reproducibility of the proposed hybrid method were significantly higher than those of the multistage pipelines (Table 1). The reproducibility of FA measures ($\epsilon = 2\%$) was higher than that reported in the literature [11] ($\epsilon = 11\%$), and higher than their tailored longitudinal version ($\epsilon = 5\%$), in which tracts were jointly reconstructed from both test-retest images. Figure 2 shows example results of scans with an average performance of the proposed method. Although all three methods showed reasonable results, the overlap of the transformed and target segmentations ($S_s(\boldsymbol{\mathcal{T}})$, S_t) was visually more consistent for the hybrid method than for the multistage pipelines.

Table 2 shows that both the registration and segmentation components of the hybrid CNN benefit from the joint optimization. For the registration task (red),

Table 1. Comparisons with the multistage pipelines. S^R, reference segmentation. κ, Cohen's kappa coefficient. ϵ, scan-rescan error in FA measures. **Bold font** indicates a statistically significant improvement over the multistage pipelines ($p < 0.01$).

Method	Accuracy$_{seg}$	Consistency$_{seg}$	κ	ϵ(%)
S^R before registration	–	0.31 ± 0.16	0.39 ± 0.18	–
S^R+Elastix	–	0.55 ± 0.09	0.64 ± 0.08	2.7%
Neuro4Neuro+Elastix	0.67 ± 0.08	0.70 ± 0.06	0.76 ± 0.06	2.5%
Hybrid	$\mathbf{0.70} \pm 0.07$	$\mathbf{0.73} \pm 0.11$	$\mathbf{0.78} \pm 0.08$	2.3%

Table 2. The added value of the hybrid method. Values in red cell indicate the mean ± SD of registration accuracy (DC($S_t^R, S_s^R(\boldsymbol{\mathcal{T}})$)) for the registration component. Those in blue cell indicate segmentation accuracy and consistency (DC($S_t, S_s(\boldsymbol{\mathcal{T}}^E)$)) for the segmentation component. **Bold font** indicates a statistically significant improvement over the other method within the colored group ($p < 0.01$).

Method	Accuracy$_{seg}$	Consistency$_{seg}$	Accuracy$_{reg}$
S^R + RegNet	-	-	0.53 ± 0.10
S^R + Hybrid	-	-	$\mathbf{0.55 \pm 0.11}$
SegNet + Elastix	0.69 ± 0.08	0.72 ± 0.06	-
Hybrid + Elastix	$\mathbf{0.70 \pm 0.07}$	0.72 ± 0.06	-

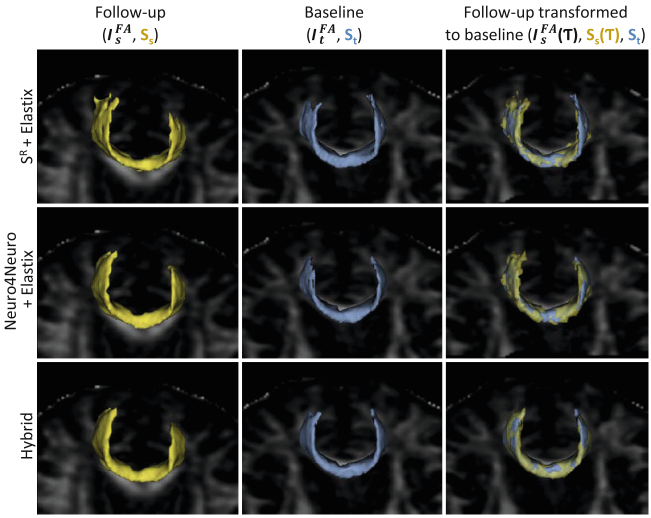

Fig. 2. Test results of the segmentation and registration. Colored structures indicate segmentations on the baseline (at 70 years old) and follow-up (26 months later) scans. The segmentation and FA image of the follow-up were transformed to the baseline. (Color figure online)

the jointly optimized registration component of the hybrid method yielded a significantly higher accuracy than the independently optimized RegNet on registering S^R. Accordingly, for the segmentation task (blue), the jointly optimized segmentation component of the hybrid method yielded a significantly higher segmentation accuracy than the independently optimized SegNet.

4 Discussion and Conclusion

We propose a novel hybrid deep learning framework for integrated segmentation and deformable registration in a single fast procedure. The framework was evaluated on longitudinal white matter tracts analysis using a large-scale diffusion MRI dataset. We show that the hybrid method leads to significantly higher accuracy, consistency and reproducibility of segmentation than multistage pipelines, and was orders of magnitude faster. Also, concurrent segmentation of structures and spatial alignment of time-points enables direct and consistent quantification of brain changes. Therefore, we expect the proposed method can open a novel way to support clinical and epidemiological analyses on understanding brain imaging changes over time.

References

1. Balakrishnan, G., et al.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)

2. de Groot, M., et al.: Tract-specific white matter degeneration in aging: the Rotterdam Study. Alzheimer's Dement. **11**(3), 321–330 (2015)
3. Hofman, A., et al.: The Rotterdam Study: 2016 objectives and design update. Eur. J. Epidemiol. **30**(8), 661–708 (2015)
4. Hu, Y., et al.: Label-driven weakly-supervised learning for multimodal deformable image registration. In: 15th ISBI, pp. 1070–1074. IEEE (2018)
5. Klein, S., et al.: Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imag. **29**(1), 196–205 (2010)
6. Li, B., de Groot, M., Vernooij, M.W., Ikram, M.A., Niessen, W.J., Bron, E.E.: Reproducible white matter tract segmentation using 3D U-Net on a large-scale DTI dataset. In: Shi, Y., Suk, H.-I., Liu, M. (eds.) MLMI 2018. LNCS, vol. 11046, pp. 205–213. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00919-9_24
7. Parisot, S., et al.: Concurrent tumor segmentation and registration with uncertainty-based sparse non-uniform graphs. Med. Image Anal. **18**(4), 647–659 (2014)
8. Pohl, K.M., et al.: An expectation maximization approach for integrated registration, segmentation, and intensity correction (2005)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
10. Vlontzos, A., Mikolajczyk, K.: Deep segmentation and registration in x-ray angiography video. arXiv preprint arXiv:1805.06406 (2018)
11. Yendiki, A., et al.: Joint reconstruction of white-matter pathways from longitudinal diffusion MRI data with anatomical priors. Neuroimage **127**, 277–286 (2016)
12. Yezzi, A., et al.: A variational framework for integrating segmentation and registration through active contours. Med. Image Anal. **7**(2), 171–185 (2003)