NeuroImage 235 (2021) 118004

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# Longitudinal diffusion MRI analysis using Segis-Net: A single-step deep-learning framework for simultaneous segmentation and registration ☆

Bo Li [a,*], Wiro J. Niessen [a,b], Stefan Klein [a], Marius de Groot [a,c], M. Arfan Ikram [a,c,d], Meike W. Vernooij [a,c], Esther E. Bron [a]

[a] Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands
[b] Imaging Physics, Applied Sciences, Delft University of Technology, the Netherlands
[c] Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
[d] Department of Neurology, Erasmus MC, Rotterdam, the Netherlands

ARTICLE INFO

*Keywords:*
Segmentation
Registration
Diffusion MRI
Deep learning
CNN
Longitudinal
White matter tract

ABSTRACT

This work presents a single-step deep-learning framework for longitudinal image analysis, coined Segis-Net. To optimally exploit information available in longitudinal data, this method concurrently learns a multi-class segmentation and nonlinear registration. Segmentation and registration are modeled using a convolutional neural network and optimized simultaneously for their mutual benefit. An objective function that optimizes spatial correspondence for the segmented structures across time-points is proposed. We applied Segis-Net to the analysis of white matter tracts from N=8045 longitudinal brain MRI datasets of 3249 elderly individuals. Segis-Net approach showed a significant increase in registration accuracy, spatio-temporal segmentation consistency, and reproducibility compared with two multistage pipelines. This also led to a significant reduction in the sample-size that would be required to achieve the same statistical power in analyzing tract-specific measures. Thus, we expect that Segis-Net can serve as a new reliable tool to support longitudinal imaging studies to investigate macro- and microstructural brain changes over time.

## 1. Introduction

The increasing availability of longitudinal imaging data is expanding our ability to capture and characterize progressive anatomical changes, ranging from normal changes in the life span, to responses along disease trajectories or therapeutic actions. Compared to cross-sectional studies, longitudinal imaging studies have the advantage of allowing to trace the order of events at the individual level and to correct for the confounding effect of time-invariant individual differences (van der Krieke et al., 2017). They are thus considered to be more accurate and sensitive in capturing subtle changes over time. To analyze spatio-temporal changes from longitudinal imaging data, a tailored framework that involves both segmentation and registration is required to segment the structures-of-interest and to register temporal frames. This can be achieved by directly combining two existing segmentation and registration tools, which are often designed for cross-sectional studies. However, the information offered in longitudinal data remains underutilized.

Various studies have shown that combining segmentation and registration at the stage of algorithm optimization can lead to improved performance. A popular combination strategy is to use the output of one task to optimize the other. Registration can be improved by using segmentation-level correspondences as input for deformation initialization (Dai and Khorram, 1999; Postelnicu et al., 2008) and optimization (Balakrishnan et al., 2019; Bastiaansen et al., 2020; De Groot et al., 2013b; Hu et al., 2018; Rohé et al., 2017; Zhu et al., 2020). Likewise, segmentation can benefit from registration by propagating anatomical information to subsequent frames, as has been shown in classical multi-atlas based segmentation methods (Fischl et al., 2002; Vakalopoulou et al., 2018) and in recent data-augmentation techniques which introduce labels to support unsupervised (Pathak et al., 2017) and weakly-supervised segmentation (Bortsova et al., 2019; Vlontzos and Mikolajczyk, 2018).

Other approaches combine the optimization of parameters from both tasks on a deeper level. Wyatt and Noble (2003) sub-grouped these methods into two types according to the way in which they update their parameters: (1) “simultaneous estimation” that updates both the class labels and the transformations in a single-step optimization, and (2) “joint estimation” that alternately updates (separate) models in a multi-step optimization. Although the initialization and robustness of

*Abbreviations:* MRI, Magnetic Resonance Imaging; DTI, Diffusion Tensor Imaging; FA, Fractional Anisotropy; MD, Mean Diffusivity; TE, Echo Time; TR, Repetition Time.

☆ Special Issue on Longitudinal Neuroimaging.

* Corresponding author.
*E-mail address:* b.li@erasmusmc.nl (B. Li).

https://doi.org/10.1016/j.neuroimage.2021.118004.
Received 20 December 2020; Received in revised form 12 March 2021; Accepted 19 March 2021
Available online 29 March 2021.
1053-8119/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

joint estimation can be influenced by the selection of the order to optimize and the criteria to switch tasks, this approach is preferred as it requires less computation power and allows to use task-specific training datasets (Ashburner and Friston, 2005; Cheng et al., 2017; Gooya et al., 2011; Parisot et al., 2014; Pohl et al., 2006; Wyatt and Noble, 2003; Xu and Niethammer, 2019; Yezzi et al., 2003). Simultaneous estimation is expected to be more accurate, as it fully exploits the conditional correlations between two tasks that can be discounted in sequential processing (Ashburner and Friston, 2005). In addition, simultaneous estimation can explicitly optimize performances that rely on both tasks. We expect that this advantage has a large potential in improving the reliability of analysis of longitudinal imaging data, for instance by optimizing the spatio-temporal consistency of the segmentation. With the growing capability of modeling and computation by deep learning techniques, several simultaneous methods have been proposed and coupled segmentation with deformable registration in different ways, either for 2D (Qin et al., 2018) or 3D images (Estienne et al., 2020; 2019; Li et al., 2019).

Diffusion magnetic resonance imaging (MRI) is a non-invasive imaging technique that measures the diffusion of water in-vivo and can be used to quantitatively characterize white matter (WM) microstructure. In addition, diffusion MRI derived measures, such as diffusion tensor imaging (DTI) metrics (Le Bihan et al., 2001), are likely to be more sensitive than structural measures in the early detection of changes in WM, and are therefore promising for the identification of subtle changes that relate to the early stages of the disease (Niessen, 2016), for instance in studying dementia subtypes (Meijboom et al., 2019). Longitudinal diffusion MRI has been widely studied at various levels, i.e., from regions-of-interest (Keihaninejad et al., 2013; Sullivan et al., 2010), to tract level (Dimond et al., 2020; Lebel and Beaulieu, 2011; Meijboom et al., 2019; Yendiki et al., 2016), and voxel level (Barrick et al., 2010; Farbota et al., 2012; De Groot et al., 2016). Since WM tracts are functionally grouped axonal fibers and thought to subserve particular brain functions, tract-specific investigation may highlight categorical differences in vulnerability to neurodegeneration and bridge the interpretation of imaging biomarkers with clinical symptoms.

Segmentation of WM tracts is however non-trivial because tracts cannot be identified directly from diffusion MRI, i.e., there is no in-vivo "gold standard" for tract (Crick and Jones, 1993), and because their anatomy can be complex. WM tracts are commonly segmented based on diffusion tractography by reconstruction of potential fiber pathways (Conturo et al., 1999). Recently, deep learning based methods, in particular using convolutional neural networks (CNN), have emerged and showed promising accuracy and efficiency in segmenting WM tracts (Li et al., 2020a; 2018; Wasserthal et al., 2018).

In the present work we focus on a CNN-based framework for longitudinal analysis of WM tracts, i.e., Segis-Net, and investigate the value of simultaneous optimization of segmentation and registration in this setting. In (Li et al., 2019), we introduced a generic framework for simultaneous optimization, in which increased accuracies of both tasks were observed in a pilot analysis of a single tract (forceps minor; FMI). In this paper, we extend the tract-specific method by enabling concurrent segmentation of multiple tracts, which is a non-trivial task as a voxel can belong to multiple tracts. This also solves the problem of inconsistencies in deformations because of tract-specific ROIs. The registration task within the framework is updated to learn only local deformations rather than an end-to-end composite including rigid transformation, as brain local changes over time is a focus in longitudinal imaging studies. In addition, we compare the performance of Segis-Net to two multistage pipelines based on both classical and deep learning algorithms, and two state-of-the-art methods. The segmentation accuracy, registration accuracy, spatio-temporal consistency of segmentation, and reproducibility of segmentation and tract-specific measures of the pipelines are quantitatively evaluated. Also, we evaluate the sample-size reduction that can be achieved in the imaging analysis of WM tracts to provide insight into the practical value of the methods in clinical applications.

## 2. Methods

In this section, we first describe how the segmentation and registration tasks are individually modeled using CNN-based approaches. Subsequently, we present the proposed Segis-Net that integrates both tasks in a single-step CNN framework.

### 2.1. CNN-based image segmentation

Given a $n$-D image $I$ which can be described by either intensity values, multi-channel features or directional tensors, the goal of CNN-based segmentation is to automatically infer, for each voxel $x \in \mathbb{R}^n$, its probability of belonging to the structure $k \in [1, K]$ with $K$ the number of structures, i.e., voxel-wise classification. The CNN model can be interpreted as a parameterized mapping function $\mathcal{F}_{\boldsymbol{\Theta}}$ such that the segmented structures spatially correspond to a segmentation ground truth with multiple channels $S = \{S_1, \ldots, S_K\}$. The estimation of the segmentation is formulated as:

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

$\mathcal{F}_{\boldsymbol{\Theta}}$ is commonly modeled by a nested series of convolutions, non-linearity, normalization, and re-sampling operations embedded in the network architecture. $\boldsymbol{\Theta}$ indicate trainable parameters.

The procedure of estimating parameter $\boldsymbol{\Theta}$ is then defined as an optimization with respect to a loss function $\mathcal{L}_{seg}$, aiming at minimizing the classification error over all the $N$ pairs of training samples $\{(S^i, I^i)\}_{i=1}^N$, i.e.,

$$\boldsymbol{\Theta} \leftarrow \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^{N} \mathcal{L}_{seg}\left(S^i, \mathcal{F}_{\boldsymbol{\Theta}}(I^i)\right). \tag{2}$$

The loss function comprises metrics that quantify the difference between the prediction and the ground truth. In this study, $\mathcal{L}_{seg}$ is the average Dice coefficient (Crum et al., 2006; Dice, 1945) over all $K$ structures:

$$\mathcal{L}_{seg}(S, \hat{S}) = -\frac{2}{K} \sum_{k=1}^{K} \frac{\sum_x S_k \hat{S}_k}{\sum_x (S_k)^2 + \sum_x (\hat{S}_k)^2}. \tag{3}$$

After estimation of the map function $\mathcal{F}_{\boldsymbol{\Theta}}$, a probabilistic prediction for the structures of interest $\hat{S}$ in a given image can be inferred (Eq. 1).

### 2.2. CNN-based deformable registration

Let us consider a pair of $n$-D images, $I_s$ in the source space $\Omega_s \subset \mathbb{R}^n$, and $I_t$ in the target space $\Omega_t \subset \mathbb{R}^n$, which contain a common structure to be aligned. The spatial correspondence between images can be established by estimating a dense displacement field $\boldsymbol{\phi}$, such that $I_s \circ \boldsymbol{\phi}$ and $I_t$ correspond spatially.

In line with the hierarchical optimization scheme of classical registration algorithms, most existing learning-based registration methods use affine alignment as a prepossessing step, in which case the displacement field denotes the composition of affine and deformable transform, i.e., $\boldsymbol{\phi} = \boldsymbol{\phi}_A \circ \boldsymbol{\phi}_D$. To estimate $\boldsymbol{\phi}_D$, the CNN model can be interpreted as a shared domain-invariant mapping function $\mathcal{G}_{\boldsymbol{\Psi}}$ such that for any unseen pair of images a most likely transformation between them can be inferred without pair-specific optimization, i.e.,

$$\hat{\boldsymbol{\phi}}_D = \mathcal{G}_{\boldsymbol{\Psi}}(I_t, I_s \circ \boldsymbol{\phi}_A) \tag{4}$$

$$\Rightarrow \quad I_s \circ \hat{\boldsymbol{\phi}} \leftarrow \mathcal{G}_{\boldsymbol{\Psi}}(I_t, I_s, \boldsymbol{\phi}_A). \tag{5}$$

The parameters $\boldsymbol{\Psi}$ of the mapping function are optimized based on a registration dissimilarity loss $\mathcal{L}_{reg}$, aimed at minimizing the registration error. Meanwhile, to penalize large deviations of deformation and preserve anatomical topology during transformations, a deformation smoothness term $\mathcal{L}_{def}$ is commonly included in the loss function. In this work, we use the mean squared error based on intensities for $\mathcal{L}_{reg}$

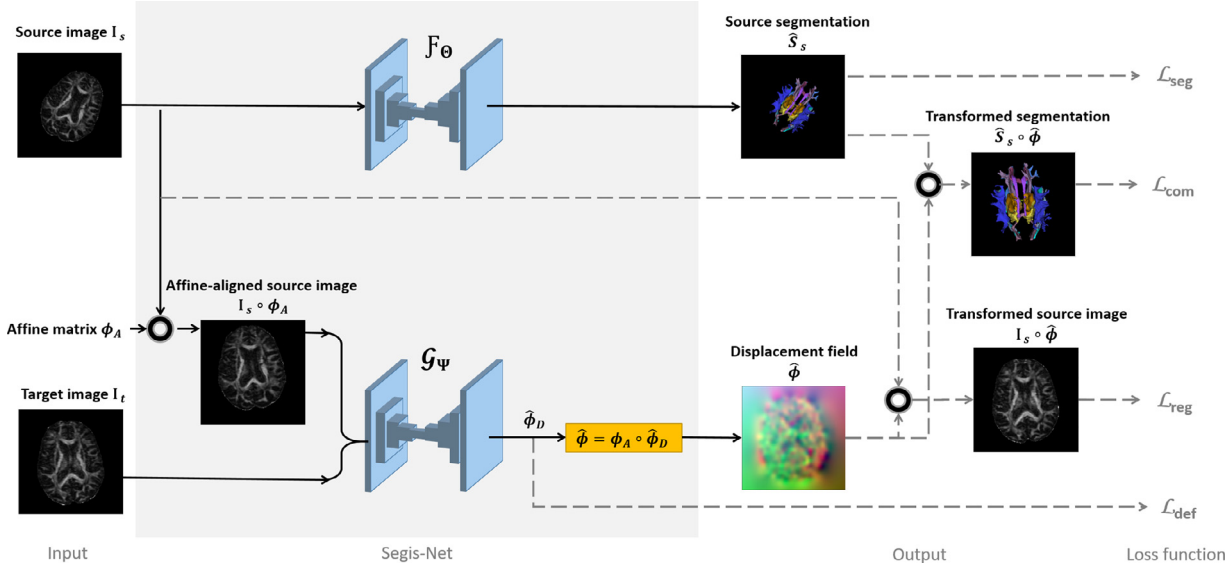


**Fig. 1.** Overview of the Segis-Net framework. $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ denote the parameters of the segmentation ($\mathcal{F}_\Theta$) and registration ($\mathcal{G}_\Psi$) function, respectively. Black circle indicates spatial warp with affine matrix ($\phi_A$) or the composite displacement field ($\hat{\phi}$). The concatenation of the affine-aligned images is used as the input for $\mathcal{G}_\Psi$. Loss function consists of $\mathcal{L}_{seg}$, $\mathcal{L}_{com}$, $\mathcal{L}_{reg}$ and $\mathcal{L}_{def}$ terms. Solid lines indicate the primary workflow of the method; dashed lines indicate the operations that are only implemented during training and could be adapted for applications.

and the average spatial gradients of the displacement field for $\mathcal{L}_{def}$, i.e.,

$$\mathcal{L}_{reg}(I_t, I_s \circ \hat{\phi}) = \frac{1}{|\Omega_t|} \|I_t - I_s \circ \hat{\phi}\|_2^2, \tag{6}$$

$$\mathcal{L}_{def}(\hat{\phi}_D) = \frac{1}{|\Omega_t|} \|\nabla \hat{\phi}_D\|_2^2. \tag{7}$$

Combining Eqs. (4), (5), (6) and (7), the estimation of $\boldsymbol{\Psi}$ over all the $N$ training samples $\{(I_t^i, I_s^i, \phi_A^i)\}_{i=1}^N$ can be formulated as:

$$\boldsymbol{\Psi} \leftarrow \underset{\Psi}{\operatorname{argmin}} \sum_{i=1}^{N} \mathcal{L}_{reg}\left(I_t^i, \mathcal{G}_\Psi(I_t^i, I_s^i, \phi_A^i)\right) + \mathcal{L}_{def}\left(\mathcal{G}_\Psi(I_t^i, I_s^i \circ \phi_A^i)\right). \tag{8}$$

### 2.3. Simultaneous estimation of segmentation and registration

In this work, we aim to simultaneously estimate the parameters for segmentation ($\boldsymbol{\Theta}$) and for registration ($\boldsymbol{\Psi}$) in a single-step optimization. For this purpose, we integrate the segmentation and registration function $\mathcal{F}_\Theta$ and $\mathcal{G}_\Psi$ using an end-to-end optimization with the Segis-Net. The loss function of the Segis-Net is designed to meet the joint objective of both tasks and meanwhile to optimize the spatio-temporal consistency of segmentation which rely on both tasks. The overview of the proposed framework is illustrated in Fig. 1. We describe the framework architecture and loss function in the following paragraphs.

#### 2.3.1. Segis-Net framework

In the present study, we focus on the analysis of 3D images and utilize 3D convolutions for the Segis-Net framework. The framework involves function $\mathcal{F}_\Theta$ and $\mathcal{G}_\Phi$ as two parallel streams that interact on their outputs. In order to eliminate the loss in image quality caused by multiple interpolations, Segis-Net warps source images with only the composite displacement fields ($\phi$) by taking as input the original source image ($I_s$) and pre-estimated affine matrix ($\phi_A$). This design has additional advantages over existing methods that prepare all ordered pairs of affine-aligned images in disk storage, as only up to half the storage is needed and as it can be flexibly applied to related images in the same space such as the DTI metrics.

$\mathcal{F}_\Theta$ outputs a set of probabilistic segmentations ($\hat{S}_s$) of the source image. $\mathcal{G}_\Psi$ outputs a dense local displacement $\hat{\phi}_D$ along the x, y, and z axes. The source image and its segmentations are subsequently warped into the target space using the composed displacement field. The warp operation is implemented by a computational layer with differentiable trilinear interpolation (Balakrishnan et al., 2019; Jaderberg et al., 2015). The segmentation and registration streams have independent network architectures which are only connected by the output, i.e., the transformed source-segmentation to the target space ($\hat{S}_s \circ \hat{\phi}$). Thus, they can be applied separately after taking advantage of the simultaneous optimization. The Segis-Net framework gives four outputs during training:

1. The segmentation of the structures of interest from the source image ($\hat{S}_s$),
2. A local displacement field between the source and target images ($\hat{\phi}_D$),
3. The warped source image in the target space ($I_s \circ \hat{\phi}$),
4. The warped source segmentations in the target space ($\hat{S}_s \circ \hat{\phi}$).

We propose a generic framework where the architecture of each stream can be adapted based on specific applications. For the particular network used in this study, we encoded two streams with a U-Net architecture, that was modified as detailed below (Ronneberger et al., 2015). In short, each stream was composed of an encoder and decoder path with skip connections of feature pyramid at multiple scales in order to merge coarse- and fine-convolved features, similar to the multi-resolution strategy used in classical algorithms to increase robustness. The encoder paths with max-pooling operation between convolution layers gradually extract abstract features for the target anatomy ($\mathcal{F}_\Theta$) and global transformation between images ($\mathcal{G}_\Psi$). Subsequently, the decoder paths restore the details in segmentations ($\mathcal{F}_\Theta$) and refine local deformations ($\mathcal{G}_\Psi$) by linear up-sampling the feature maps and concatenating them with the coarse counterpart at the same scale. The convolution layers produce a set of feature maps by individually convolving inputs with 3D kernels of size (3,3,3), followed by batch normalization (Ioffe and Szegedy, 2015) and a leaky ReLu layer ($a = 0.2$) for modeling non-linearity (Maas et al., 2013). For the segmentation stream ($\mathcal{F}_\Theta$), we split the output layer into sub-branches to facilitate multi-class classification for voxels with multiple labels. The final layer of the sub-branches consisted of a (1,1,1) convolution and a sigmoid activation. For the registration stream, the output layer was a convolution layer with three kernels that yielded the local displacement $\hat{\phi}_D$. We provide detailed im-

plementation of the network architecture in the supplementary material (Figs. 8 and 9).

#### 2.3.2. Segis-Net loss function

The loss function of Segis-Net is composed of four terms that measure segmentation accuracy ($\mathcal{L}_{seg}$, Eq. (3)), intensity similarity between registered images ($\mathcal{L}_{reg}$, Eq. (6)), deformation field smoothness ($\mathcal{L}_{def}$, Eq. (7)), and longitudinal composite of registration and segmentation ($\mathcal{L}_{com}$, Eq. (11)). It is formulated as:

$$\begin{aligned}\mathcal{L} \quad &= \mathcal{L}_{seg}\left(S_s,\, \hat{S}_s\right) \\ &+\alpha\mathcal{L}_{reg}\left(I_t,\, I_s\circ\hat{\boldsymbol{\phi}}\right) + \beta\mathcal{L}_{def}\left(\hat{\boldsymbol{\phi}}_D\right) \\ &+\gamma\mathcal{L}_{com}\left(S_t,\, \hat{S}_s\circ\hat{\boldsymbol{\phi}}\right),\end{aligned} \tag{9}$$

and optimized for $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ over all $N$ training samples $\{(S_t^i, S_s^i, I_t^i, I_s^i, \boldsymbol{\phi}_A^i)\}_{i=1}^N$:

$$\begin{aligned}\boldsymbol{\Theta}, \boldsymbol{\Psi} \quad &\leftarrow \operatorname{argmin}_{\boldsymbol{\Theta},\boldsymbol{\Psi}} \textstyle\sum_{i=1}^N \mathcal{L}_{seg}\left(S_s^i, \mathcal{F}_{\boldsymbol{\Theta}}(I_s^i)\right) \\ &+\alpha\mathcal{L}_{reg}\left(I_t^i,\, \mathcal{G}_{\boldsymbol{\Psi}}(I_t^i,\, I_s^i,\, \boldsymbol{\phi}_A^i)\right) + \beta\mathcal{L}_{def}\left(\mathcal{G}_{\boldsymbol{\Psi}}(I_t^i,\, I_s^i\circ\boldsymbol{\phi}_A^i)\right) \\ &+\gamma\mathcal{L}_{com}\left(S_t^i, \mathcal{F}_{\boldsymbol{\Theta}}(I_s^i), \mathcal{G}_{\boldsymbol{\Psi}}(I_t^i,\, I_s^i,\, \boldsymbol{\phi}_A^i)\right).\end{aligned} \tag{10}$$

We quantify the longitudinal composite loss term using the average Dice coefficient over all $K$ structures:

$$\mathcal{L}_{com}\left(S_t,\, \hat{S}_s\circ\hat{\boldsymbol{\phi}}\right) = -\frac{2}{K}\sum_{k=1}^{K}\frac{\sum_{x\in\Omega_t} S_t^k(\hat{S}_s^k\circ\hat{\boldsymbol{\phi}})}{\sum_{x\in\Omega_t}(S_t^k)^2 + \sum_{x\in\Omega_t}(\hat{S}_s^k\circ\hat{\boldsymbol{\phi}})^2}. \tag{11}$$

In longitudinal imaging studies, the spatial correspondence in segmentation depends on the performance of both the segmentation and the registration procedure. Besides an explicit optimization of correspondence, the $\mathcal{L}_{com}$ term also exploits longitudinal information to boost both tasks, which introduces some degree of augmentation and regularization for registration and on the other hand constraints and prior knowledge for segmentation.

The hyperparameters $\alpha$ and $\gamma$ balance the loss magnitude of segmentation, registration, and their interdependent composite. The degree of regularization on the deformation is described by $\beta$. The procedure of simultaneous optimization is summarized with pseudo code in supplementary material (Algorithm 1).

## 3. Application to diffusion MRI

The performance of Segis-Net is demonstrated by analyzing white matter tracts in a large diffusion MRI dataset, and compared to that of two multi-stage pipelines, in which segmentation and registration are independently optimized. Performance is evaluated in a longitudinal setting where multiple time-points from the same individual are available.

### 3.1. Dataset

The Rotterdam Study is a prospective and population-based study targeting causes and consequences of age-related diseases (Ikram et al., 2020). For the present analysis, we included 3249 individuals who underwent diffusion MRI scanning twice or more often, resulting in $N = 8045$ scans. The mean age at first scan was $61.2 \pm 9.4$ years (range: $45.7 - 91.1$ years). The number of female participants was 1780 (54.8%). A flowchart for the inclusion, exclusion, and split of the datasets is shown in Fig. 2. We split the data into two subsets. The larger subset was repeatedly acquired in a time interval of 1–5 years ($N = 7770$ scans from 3166 individuals). In these long time-interval scans, it is expected that brain microstructure changes due to aging exist. By matching any two time-points from the same individual regardless of the visiting order, these long time-interval scans can be grouped into 6043 pairs. We used 5175 pairs of scans as training data, 200 pairs as validation data to tune the hyperparameters, monitor the decay of learning rate and select the optimal epoch, and used an independent cohort of 668 pairs for testing. The remaining scans from the smaller subset were from 97 individuals who were scanned twice within a month. No changes in brain macro- and microstructure were expected within such a short time-interval. We used these scans for evaluation of reproducibility of the algorithm. The data split was based on the participants, namely, we made sure that scans from the same participant ended up in either training, validation, or test dataset.

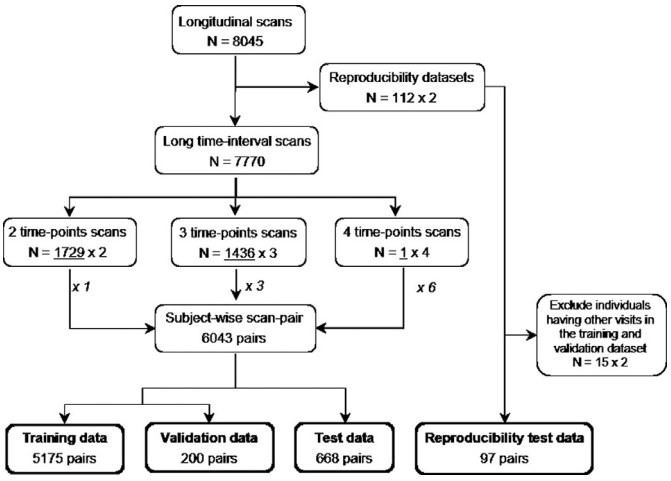


**Fig. 2.** A flowchart for the inclusion, exclusion, and split of the datasets.

### 3.2. MRI acquisition

Scans were acquired on a 1.5T MRI scanner (GE Signa Excite). The acquisition parameters for structural and diffusion MRI can be found in Ikram et al. (2011). Specifically, diffusion MRI was scanned with the following parameters: TR/TE $=8575ms/82.6ms$, imaging matrix of $64 \times 96$, FOV=$21 \times 21cm^2$, 35 contiguous slices with slice thickness 3.5 mm, 25 diffusion weighted volumes with a b-value of $1000s/mm^2$ and 3 non-weighted volumes (b-value $= 0s/mm^2$). The voxel size was resampled from $3.3 \times 2.2 \times 3.5mm^3$ to $1mm^3$ as required for probabilistic tractography (Behrens et al., 2007).

### 3.3. Image preprocessing

Diffusion data were preprocessed using a standardized pipeline (Koppelmans et al., 2014). In short, motion and eddy currents were corrected by affine co-registration of all diffusion weighted volumes to the averaged b0 volumes, including correction of gradient vector directions using Elastix software (Klein et al., 2010). Diffusion tensors were estimated with a Levenberg–Marquardt non-linear least-squares optimization algorithm (Leemans et al., 2009). We subsequently computed DTI measures: fractional anisotropy (FA) and mean diffusivity (MD). Due to noise, tensor estimation failed in a small proportion of voxels, resulting in significant outliers. Outlier voxels with a tensor norm (Frobenius norm) larger than $0.1mm^2/s$ were set to zero (Zhang et al., 2007). Brain tissue masks including WM and gray matter segmentations were obtained based on structural imaging (Vrooman et al., 2007) and applied to the diffusion tensor images. In this study, we used a ROI of $112 \times 208 \times 112$ voxels to analyze six WM tracts, including left and right cingulate gyrus part of cingulum (CGC), left and right parahippocampal part of cingulum (CGH), forceps major (FMA) and forceps minor (FMI). Diffusion tensor images were image-wise normalized by setting the union of the six components to zero-mean and standard deviation of one. The affine matrix ($\boldsymbol{\phi}_A$) of each image pair was estimated by optimizing the mutual information of FA images using Elastix software.

### 3.4. Reference segmentations

The segmentation labels for model training and evaluation were generated using a probabilistic tractography and atlas-based segmentation

method by De Groot et al. (2015). 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 on a subset of 30 participants (De Groot et al., 2013b). We did not exclude this subset from the reproducibility test data (Fig. 2), as it remains unseen to the proposed method and other baseline methods.

### 3.5. Baseline multi-stage pipelines

We compared the performance of the proposed Segis-Net with two multi-stage pipelines that consist of either non-learning-based or learning-based algorithms to investigate the added value of simultaneous optimization. To assess whether the performance difference between approaches was statistically significant, paired t-tests with P-value threshold $< 0.05$ and Bonferroni correction for controlling the family-wise error of multiple testings were performed.

First, a non-learning-based *Classical* pipeline was built using an existing tractography-based segmentation algorithm (Section 3.4) and a deformable registration algorithm Elastix (Klein et al., 2010). Elastix was adopted as a competing classical registration method since it has been widely used on our dataset and thereby an optimal parameter setting can be applied for performance comparison. Elastix is designed to run in a cascade of resolutions, and offers the choice between multiple objective functions and multiple optimizers including an efficient adaptive stochastic gradient descent optimizer (Klein et al., 2009). For Elastix (version 4.8), we used a rigid, affine, and B-spline transformation model consecutively by maximizing mutual information between images. The B-spline transformation of spline order 3 was implemented using a multi-resolution framework with isotropic control-point spacing of 24, 12, and 6 *mm* in three-level resolutions. The maximum number of iterations was 1024.

Second, we built a learning-based *CNN* pipeline using components from the proposed Segis-Net to evaluate the sole contribution of simultaneous optimization. In this pipeline, we split the integrated segmentation $\mathcal{F}_{\boldsymbol{\Theta}}$ and registration stream $\mathcal{G}_{\boldsymbol{\Phi}}$ into two separate neural networks for independent optimization. Subsequently, the segmented images and estimated transformations were combined. The segmentation network had the same architecture as that for $\mathcal{F}_{\boldsymbol{\Theta}}$, except being independently optimized using the segmentation accuracy $\mathcal{L}_{seg}$ term. As this is a typical setting for CNN-based segmentation approaches (Li et al., 2018; Ronneberger et al., 2015), we denote it as *Seg-Net*. Similarly, the registration network denoted as *Reg-Net* had the same setting as that for $\mathcal{G}_{\boldsymbol{\Psi}}$, except being independently optimized using registration similarity $\mathcal{L}_{reg}$ and regularization $\mathcal{L}_{def}$ terms (Balakrishnan et al., 2019). We ensured that the training dataset for the *Seg-Net* and *Reg-Net* was the same as that used for the Segis-Net framework.

### 3.6. Related methods involving segmentation and registration

When it comes to the combination of segmentation and registration, there are various integration strategies (Section 1). To investigate the benefit of the proposed simultaneous optimization strategy, we additionally compared Segis-Net with two previously published methods:

- U-ReSNet for simultaneous segmentation and registration that used a shared feature encoder and separate decoders (Estienne et al., 2019).
- VoxelMorph for image registration alone that used correspondence in existing segmentation labels to boost registration (Balakrishnan et al., 2019).

### 3.7. Implementation

For this diffusion MRI application, the segmentation ($\mathcal{F}_{\boldsymbol{\Theta}}$) and registration ($\mathcal{G}_{\boldsymbol{\Psi}}$) components of Segis-Net used different input images. Specifically, segmentation was based on the diffusion tensor image, as it contains directional information of fiber populations and was shown to be optimal in the present setting of clinical-quality resolution (Li et al., 2020a). For spatial alignment, we adopted the input the commonly used scalar-value FA map derived from diffusion tensor imaging.

To mitigate class imbalance and to improve computational efficiency, we combined the reference segmentation for the six tracts (Section 3.3) into a three-channel map for using it as the segmentation ground truth $S$. This combination was possible since only few crossing fibers are expected between codirectional WM tracts (e.g., FMI and FMA). To evaluate performance on individual tracts after training, we extracted the two largest components from each of the three channels of the probabilistic prediction and subsequently identified the left and right (for CGC and CGH) or the anterior and posterior (for FMI and FMA) tract based on coordinates.

The experiments of model training and evaluation were performed on an NVIDIA 1080Ti GPU and an AMD 1920X CPU. CNN-based methods were implemented using Keras-2.2.0 with a Tensorflow-1.4.0 backend and the Adam optimizer (Kingma and Ba, 2014). For *Reg-Net, Seg-Net* and Segis-Net, weights of convolution kernels were initialized with the Glorot uniform distribution (Glorot and Bengio, 2010). In each training epoch, input images were fed in random batches (size = 1). Loss function hyperparameters were optimized based on segmentation and registration performance on the validation dataset (search range: $[10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}]$); we set to $\alpha = 10$, $\beta = 0.01 \times \alpha$ for *Reg-Net;* for Segis-Net we linearly increased $\alpha$ from 10 to 100 by 4 per epoch (with $\beta$ increased accordingly), and set the additional parameter $\gamma = 1$. The initial learning rates were experimentally optimized on the validation dataset and set to $1e^{-4}$, $1e^{-3}$ and $1e^{-3}$ for *Reg-Net, Seg-Net* and Segis-Net, which were decayed with a factor of 0.8 if the validation loss stopped decreasing for 10 epochs (decay condition, Algorithm 1). We stopped the training procedure at the point that the validation loss showed consecutive increases, i.e., early stopping (Bishop, 2006). The parameters of the model with the smallest error with respect to the validation dataset were used.

For VoxelMorph, the implementation as detailed by Balakrishnan et al. (2019) was used directly. For U-ResNet, in contrast to the other tensor-based segmentation methods, we used the FA map as input for both segmentation and registration since the shared feature-encoder required the same input for both tasks. Affine registration was applied as a pre-processing step. Hyperparameters were tuned on the validation dataset; and we obtained improved performance by using an initial learning rate of 0.0005 and by clipping the warped segmentation predictions into the range of $[10^{-7}, 1 - 10^{-7}]$.

## 4. Experiments and results

We applied the methods to analyze six WM tracts. The performance of the proposed Segis-Net was compared with the two baseline multi-stage pipelines on segmentation accuracy, registration accuracy, spatio-temporal consistency of segmentation, reproducibility of segmentation and measurements, and sample-size reduction; and compared with the two related methods in terms of the segmentation and registration accuracy.

### 4.1. Segmentation accuracy

Segmentation accuracy was quantified with respect to the reference segmentation (Section 3.4) using the Dice coefficient metric.

The proposed method yielded similar segmentation accuracy as the baseline multistage *CNN* pipeline (*Seg-Net*) for all six tracts (Fig. 3). Both methods achieved relatively high accuracy in segmenting cingulum, i.e., the accuracy of left and right CGC and CGH tracts was around $0.76 \pm 0.07$. The accuracy was lowest for FMI (*CNN*: $0.68 \pm 0.09$; Segis-Net: $0.67 \pm 0.09$), which is a thin and arch-shaped tract that is known to be more difficult to segment. Correcting for 6 tests resulted in an

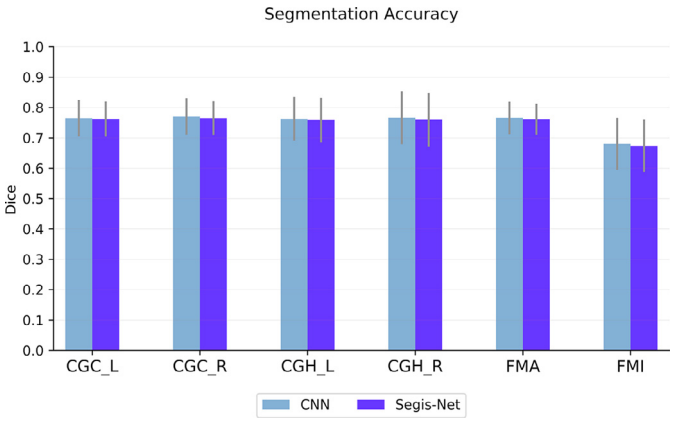


**Fig. 3.** Segmentation accuracies of the *CNN* pipeline and Segis-Net for different tracts. Error bars indicate standard deviations.

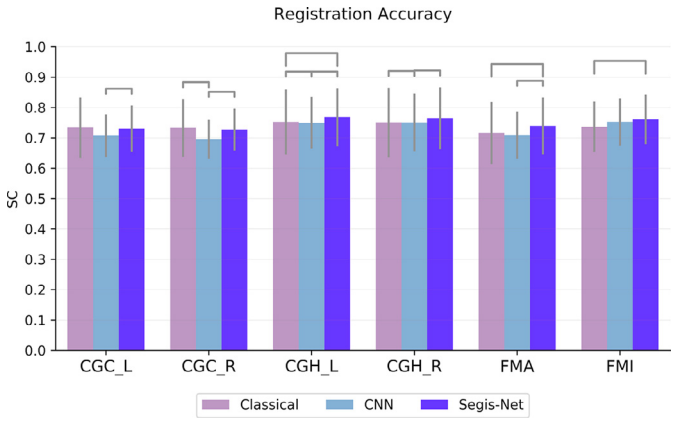


**Fig. 4.** Registration accuracies of the *Classical, CNN*, and Segis-Net pipeline as quantified by spatial correlation (SC) of the registered tract density maps. Error bars indicate standard deviations. The bracket hat indicates a significant difference between two methods (*t*-test, $p < 2.8 \times 10^{-3}$).

**Table 1**
Segmentation Dice coefficient of U-ReSNet and Segis-Net. The bold value indicates a better performance in each row.

| | U-ReSNet | Segis-Net |
|---|---|---|
| CGC_L | 0.69 ± 0.06 | **0.76 ± 0.06** |
| CGC_R | 0.69 ± 0.07 | **0.76 ± 0.06** |
| CGH_L | 0.67 ± 0.08 | **0.76 ± 0.07** |
| CGH_R | 0.67 ± 0.09 | **0.76 ± 0.09** |
| FMA | 0.69 ± 0.06 | **0.76 ± 0.05** |
| FMI | 0.60 ± 0.08 | **0.67 ± 0.09** |

**Table 2**
Registration performance of U-ReSNet, VoxelMorph and Segis-Net, as quantified by the spatial correlation (SC) similarity, the Dice coefficient (DC), and the mean squared error (MSE). The bold value indicates the best performance in each row.

| | | U-ReSNet | VoxelMorph | Segis-Net |
|---|---|---|---|---|
| SC | CGC_L | **0.77 ± 0.11** | 0.72 ± 0.09 | 0.73 ± 0.08 |
| | CGC_R | **0.77 ± 0.11** | 0.71 ± 0.09 | 0.73 ± 0.07 |
| | CGH_L | **0.77 ± 0.11** | 0.75 ± 0.10 | **0.77 ± 0.10** |
| | CGH_R | **0.77 ± 0.12** | 0.75 ± 0.11 | 0.76 ± 0.10 |
| | FMA | 0.73 ± 0.11 | 0.73 ± 0.10 | **0.74 ± 0.09** |
| | FMI | 0.75 ± 0.09 | 0.74 ± 0.08 | **0.76 ± 0.08** |
| DC | CGC_L | 0.69 ± 0.07 | 0.65 ± 0.06 | **0.74 ± 0.06** |
| | CGC_R | 0.70 ± 0.07 | 0.65 ± 0.06 | **0.74 ± 0.05** |
| | CGH_L | 0.67 ± 0.08 | 0.64 ± 0.07 | **0.71 ± 0.08** |
| | CGH_R | 0.67 ± 0.10 | 0.64 ± 0.08 | **0.71 ± 0.09** |
| | FMA | 0.70 ± 0.06 | 0.68 ± 0.06 | **0.72 ± 0.06** |
| | FMI | 0.57 ± 0.10 | 0.56 ± 0.07 | **0.60 ± 0.09** |
| MSE ($\times 10^{-2}$) | | 0.47 ± 0.38 | 0.19 ± 0.88 | **0.13 ± 0.10** |

adjusted P-value threshold of $8.3 \times 10^{-3}$. There was no significant differences in segmentation accuracy between two methods.

The proposed method showed higher segmentation accuracy than U-ReSNet for all six tracts with a margin of around 10% and smaller standard deviation (Table 1).

### 4.2. Registration accuracy

Registration accuracy of the approaches was evaluated with the spatial correlation (SC) similarity on the test dataset. According to the procedure in De Groot et al. (2013b), the estimated transformation was applied to the continuous density maps of individual tracts obtained from probabilistic tractography, subsequently, the SC similarity between warped density maps was computed as follow:

$$SC_k = \frac{\sum_{x \in \Omega_t} J_t^k (J_s^k \circ \hat{\boldsymbol{\phi}})}{\left( \sum_{x \in \Omega_t} \sqrt{(J_t^k)^2} \right) \left( \sum_{x \in \Omega_t} \sqrt{(J_s^k \circ \hat{\boldsymbol{\phi}})^2} \right)}, \tag{12}$$

where $J_t^k$ and $J_s^k$ indicate intensity of the target and source density image of the tract $k$. Despite a lot of intensity variation in the tract density maps across scans due to the probabilistic nature of tractography, higher intensity in general indicates more support for the tract while lower intensity conversely indicates increased uncertainty. Therefore, we assume that SC reflects the spatial correspondence of tracts.

Fig. 4 presents the registration accuracy (SC) of Segis-Net and the baseline multistage pipelines. The SC in all six tracts was overall highest for the Segis-Net, followed by the *Classical* pipeline. Correcting for 18 tests resulted in Bonferroni adjusted P-value threshold of $2.8 \times 10^{-3}$. Segis-Net results yielded a significantly better spatial correspondence than the *Classical* pipeline in the left CGH (Segis-Net vs *Classical* = $0.77 \pm 0.09$ vs $0.75 \pm 0.11$), FMA ($0.74 \pm 0.09$ vs $0.72 \pm 0.10$), and FMI ($0.76 \pm 0.08$ vs $0.74 \pm 0.08$) tract. Statistically significant difference in the registration accuracy of Segis-Net and *CNN* pipeline were observed in the left CGC (Segis-Net vs *CNN* = $0.73 \pm 0.08$ vs $0.71 \pm 0.07$), right CGC ($0.73 \pm 0.07$ vs $0.69 \pm 0.06$), left CGH ($0.77 \pm 0.09$ vs $0.75 \pm 0.08$), right CGH ($0.76 \pm 0.10$ vs $0.75 \pm 0.10$), and FMA ($0.74 \pm 0.09$ vs $0.71 \pm 0.08$) tract. In general, the proposed Segis-Net approach achieved a better spatial correspondence than the two independently optimized registration algorithms using classical and learning-based techniques.

For comparing the proposed method with U-ReSNet and VoxelMorph, we added the performance metric that was used in their original papers (Balakrishnan et al., 2019; Estienne et al., 2019), i.e., the Dice coefficient (DC) of registered reference segmentation of tracts, and added the common loss metric, i.e., the mean squared error (MSE) between registered FA maps (Table 2). Generally, the SC similarity of the proposed method and U-ReSNet were better than that of VoxelMorph. Segis-Net led to the highest similarity in FMA and FMI tract; U-ReSNet was the highest for the left and right CGC, and the right of CGH tract; for the left CGH tract, a similar SC was observed for U-ReSNet and Segis-Net, although the variations were smaller in Segis-Net; Segis-Net achieved the best DC and MSE. For all six tracts, the DC of Segis-Net were higher than that of U-ReSNet, followed by VoxelMorph; the standard deviation of Segis-Net was overall smallest for all three metrics, except that of DC in three tracts (left and right CGH, and FMI) which were smallest for VoxelMorph.

### 4.3. Spatio-temporal consistency of segmentation

To evaluate the spatio-temporal consistency of segmentation (STCS) for Segis-Net and the baseline multistage pipelines, we measured the correspondence between warped segmentation results across time-points using the Dice coefficient. The consistency of each tract was av-

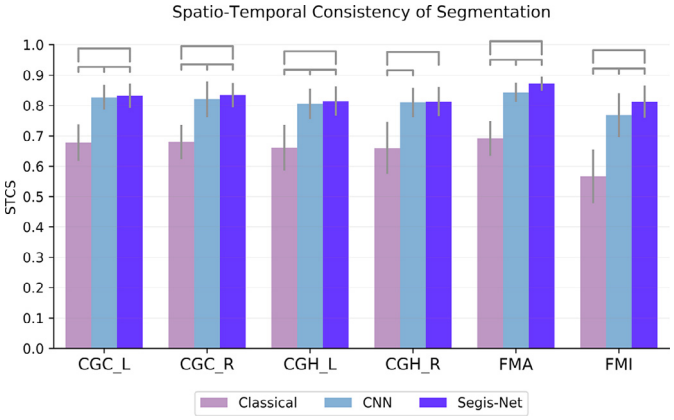


**Fig. 5.** Spatio-temporal consistency of segmentation (STCS) with the *Classical, CNN*, and Segis-Net pipeline. Error bars indicate standard deviations. The bracket hat indicates a significant difference between two methods (*t*-test, $p < 2.8 \times 10^{-3}$).

eraged over two directions by reversing the target and source image, which for the tract $k$ can be formulated as:

$$STCS_k = \frac{1}{2}\left(\frac{2|\hat{S}_t^k \cap \hat{S}_s^k \circ \hat{\boldsymbol{\phi}}|}{|\hat{S}_t^k| + |\hat{S}_s^k \circ \hat{\boldsymbol{\phi}}|} + \frac{2|\hat{S}_s^k \cap \hat{S}_t^k \circ \hat{\boldsymbol{\phi}}^{-1}|}{|\hat{S}_s^k| + |\hat{S}_t^k \circ \hat{\boldsymbol{\phi}}^{-1}|}\right). \tag{13}$$

Each pipeline was evaluated as a whole, that is, (1) in *Classical* pipeline the reference segmentation was warped by Elastix algorithm, (2) in *CNN* pipeline the prediction of *Seg-Net* was warped by the predicted transformation of the *Reg-Net*, and (3) in Segis-Net framework the segmentation prediction in native space and the segmentation warped from another time-point were available after a bidirectional test.

The proposed Segis-Net overall showed higher segmentation consistency than the *CNN* and the *Classical* pipeline (Fig. 5). Correcting for 18 tests resulted in an adjusted P-value threshold of $2.8 \times 10^{-3}$. In comparison with the *CNN* pipeline, Segis-Net results yielded significantly higher spatio-temporal consistency in left CGC (Segis-Net vs *CNN* $= 0.83 \pm 0.04$ vs $0.82 \pm 0.04$), right CGC ($0.83 \pm 0.04$ vs $0.82 \pm 0.06$), left CGH ($0.82 \pm 0.05$ vs $0.81 \pm 0.05$), FMA ($0.87 \pm 0.02$ vs $0.84 \pm 0.03$), and FMI ($0.81 \pm 0.05$ vs $0.77 \pm 0.07$) tract. In all six tracts, Segis-Net significantly outperformed the *Classical* pipeline, i.e., in left CGC (Segis-Net vs *Classical* $= 0.83 \pm 0.04$ vs $0.68 \pm 0.06$), right CGC ($0.83 \pm 0.04$ vs $0.68 \pm 0.06$), left CGH ($0.82 \pm 0.05$ vs $0.66 \pm 0.08$), right CGH ($0.81 \pm 0.05$ vs $0.66 \pm 0.09$), FMA ($0.87 \pm 0.02$ vs $0.69 \pm 0.06$), and FMI ($0.81 \pm 0.05$ vs $0.57 \pm 0.09$) tract.

### 4.4. Reproducibility of segmentation and measurements

Reproducibility of tract-specific segmentations, volumes, and diffusion metrics of the pipelines was evaluated using the reproducibility dataset. We quantified voxel-wise agreement between segmentations of repeated scans using Cohen's kappa coefficient ($\kappa$). The segmentations $(\hat{S}_t, \hat{S}_s)$ were obtained in the native space, and subsequently aligned $(\hat{S}_s \circ \hat{\boldsymbol{\phi}})$. Kappa $\kappa$ of the tract $k$ is defined as:

$$\kappa_k = \frac{p_o(\hat{S}_t^k, \hat{S}_s^k \circ \hat{\boldsymbol{\phi}}) - p_e(\hat{S}_t^k, \hat{S}_s^k \circ \hat{\boldsymbol{\phi}})}{1 - p_e(\hat{S}_t^k, \hat{S}_s^k \circ \hat{\boldsymbol{\phi}})}, \tag{14}$$

in which $p_o(\hat{S}_t^k, \hat{S}_s^k \circ \hat{\boldsymbol{\phi}})$ is the observed agreement between $\hat{S}_t^k$ and $\hat{S}_s^k \circ \hat{\boldsymbol{\phi}}$, $p_e$ is the hypothetical probability of the agreement. Given $|\Omega_t|$ being the total number of voxels in the target image, $|S|$ and $|\Omega_t| - |S|$ being the number of tract and non-tract voxels, the observed agreement (i.e., accuracy) is computed as:

$$p_o(\hat{S}_t^k, \hat{S}_s^k \circ \hat{\boldsymbol{\phi}}) = \frac{|\hat{S}_t^k \cap (\hat{S}_s^k \circ \hat{\boldsymbol{\phi}})| + |(1 - \hat{S}_t^k) \cap \left(1 - (\hat{S}_s^k \circ \hat{\boldsymbol{\phi}})\right)|}{|\Omega_t|}, \tag{15}$$

the hypothetical probability of the agreement can be formulated as:

$$p_e(\hat{S}_t^k, \hat{S}_s^k \circ \hat{\boldsymbol{\phi}}) = \frac{1}{|\Omega_t|^2}\left(|\hat{S}_t^k| \times |\hat{S}_s^k \circ \hat{\boldsymbol{\phi}}| + (|\Omega_t| - |\hat{S}_t^k|) \times (|\Omega_t| - |\hat{S}_s^k \circ \hat{\boldsymbol{\phi}}|)\right). \tag{16}$$

Typically, a $\kappa > 0.60$ indicates "substantial" agreement, and a $\kappa > 0.80$ indicates "almost perfect" agreement (Landis and Koch, 1977).

Similarly, to evaluate the reproducibility of tract-specific measurements, we computed the FA, MD and volume in image native space, and subsequently assessed relative differences in paired scan-rescan measures $(m_t, m_s)$ as an indicator of measurement error ($\epsilon$), i.e.,

$$\epsilon = \frac{2|m_s - m_t|}{(m_s + m_t)} \times 100\%. \tag{17}$$

For FA and MD, the tract-specific measures were quantified as the median of non-zero values within the segmented images. A lower $\epsilon$ indicates a better reproducibility.

Fig. 6 presents the reproducibility of tract-specific segmentation and measures determined with the baseline multi-stage pipelines and Segis-Net. The proposed Segis-Net achieved the best segmentation reproducibility, followed by the *CNN* pipeline (Fig. 6 (a)); in all six tracts, $\kappa$ was around 0.80 or higher, indicating "almost perfect" agreements between segmentations of repeated scans. Correcting for 18 tests for each metric resulted in an adjusted P-value threshold of $2.8 \times 10^{-3}$, resulting in overall statistically significant improvement by Segis-Net over the *Classical* pipeline. For two tracts, voxel-wise agreement of Segis-Net was significantly higher than that of the *CNN* pipeline, i.e., FMA (Segis-Net vs *CNN* $= 0.87 \pm 0.03$ vs $0.85 \pm 0.03$) and FMI ($0.82 \pm 0.06$ vs $0.79 \pm 0.08$).

Additionally, in the evaluation of the reproducibility in tract-specific volume measures, Segis-Net showed the smallest error in all six tracts (Fig. 6 (b)). The error of Segis-Net was significantly smaller than the *Classical* pipeline in left CGC (Segis-Net vs *Classical* $= 4.8 \pm 4.1\%$ vs $11 \pm 8.8\%$), right CGC ($4.5 \pm 3.9\%$ vs $11 \pm 8.9\%$), left CGH ($7.3 \pm 5.6\%$ vs $11 \pm 9.8\%$), and FMA ($3.4 \pm 2.6\%$ vs $6.6 \pm 5.9\%$) tract. This outperformed the *CNN* pipeline significantly in the FMA tract (Segis-Net vs *CNN* $= 3.4 \pm 2.6\%$ vs $4.9 \pm 3.6\%$). Reproducibility of FA and MD measurements was similar for the three methods (Fig. 6 (c, d)). For the CGH and left CGC tracts, the reproducibility of FA using the *CNN* pipeline was significantly higher than that of the *Classical* pipeline. Segis-Net outperformed the FA reproducibility of the *Classical* pipeline only in the left CGC tract. For MD, no significant improvement over the *Classical* pipeline was observed. A table (Table 3) with the results of Fig. 3–6 is provided in the supplementary files.

### 4.5. Sample-size reduction

An implication of the reduced measurement error ($\epsilon$) is that fewer participants or time-points would be required to achieve the same statistical power, i.e., a smaller sample size. We followed Diggle et al. (2002) and Reuter et al. (2012) to estimate the percentage of the sample sizes ($P$) that would be required for each of the pipelines:

$$P_{ij} = \frac{\sigma_i^2 \times (1 - \rho_i)}{\sigma_j^2 \times (1 - \rho_j)} \times 100\%, \tag{18}$$

where $\sigma_i$ and $\sigma_j$ are standard deviations in the measurements determined with the pipeline $i$ and $j$, and $\rho_i$ and $\rho_j$ are the correlation coefficients between the repeated measurements determined with the two pipelines.

Fig. 7 presents the percentage of sample-size reduction that could be achieved by the *CNN* and the proposed Segis-Net compared to the *Classical* pipeline. In line with the reproducibility results, the data analyzed with Segis-Net would overall require the least sample-size to achieve the same statistical power. The percentage of reduction was especially remarkable in volume measures, in which on average only 33.0% of data would be required. The average percentage of reduction was 60.5% for

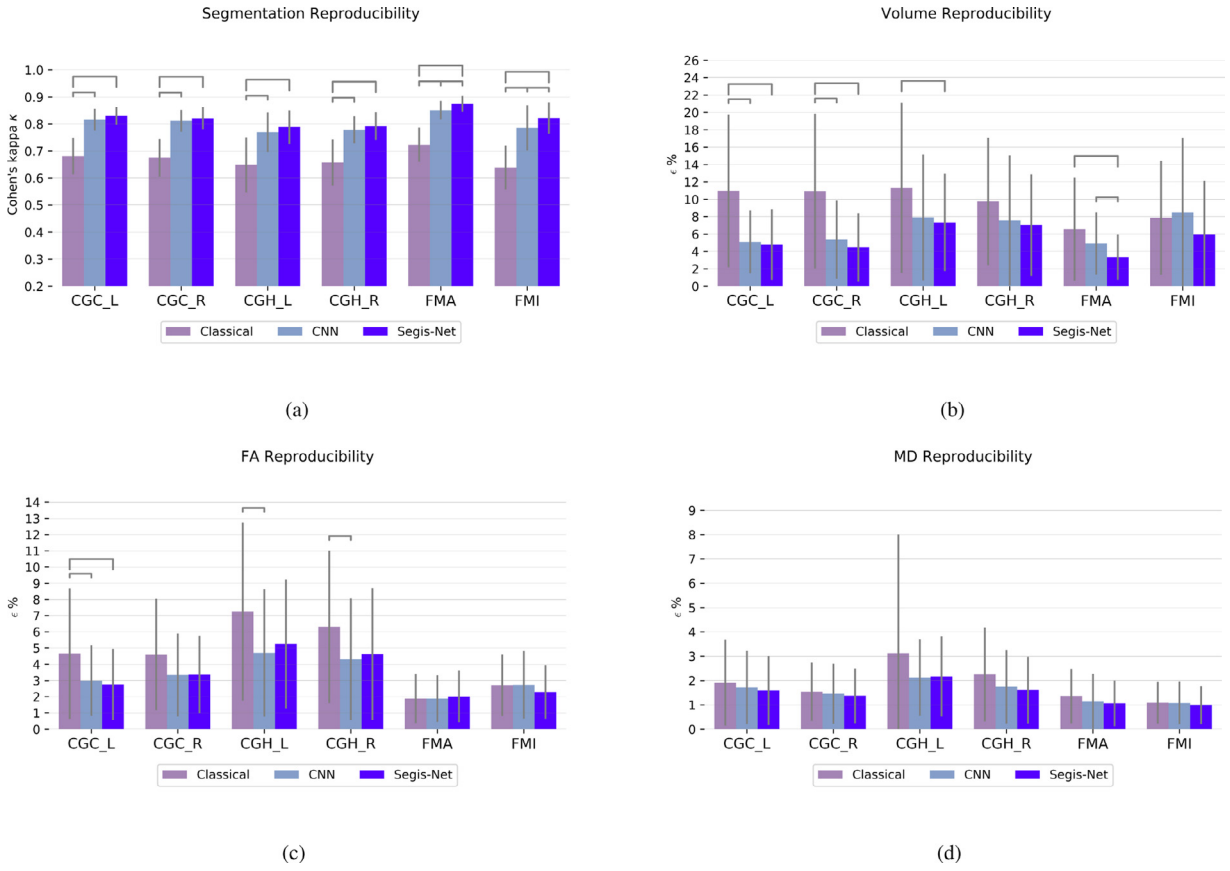


**Fig. 6.** Reproducibility of tract-specific measures with the *Classical, CNN*, and Segis-Net pipeline. Error bars indicate standard deviations. The bracket hat indicates a significant difference between two methods ($t$-test, $p < 2.8 \times 10^{-3}$). In figure (a), a higher Cohen's kappa coefficient ($\kappa$) indicates a better reproducibility. In figure (b-d), a lower error ($\epsilon\%$) indicates a better reproducibility. Volume: tract-specific volume (ml), FA: fractional anisotropy, MD: mean diffusivity ($10^{-3} mm^2/s$).

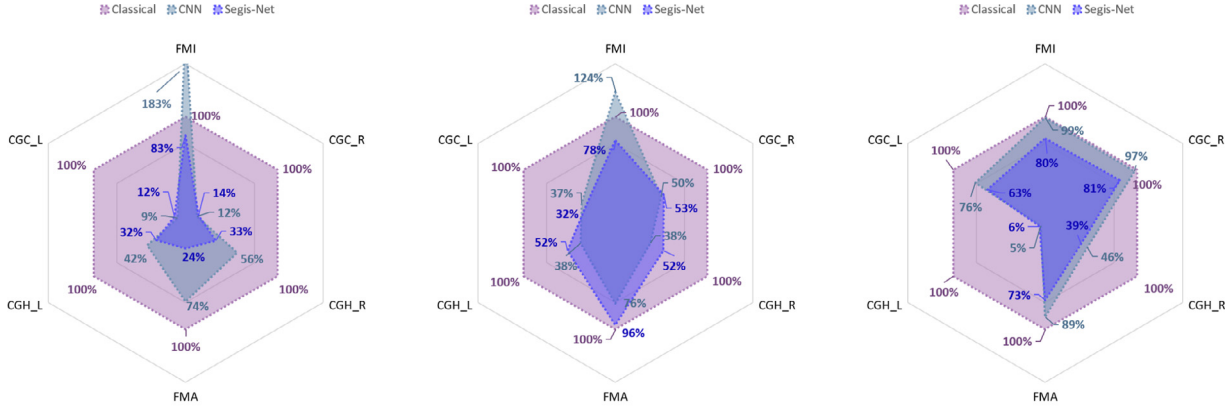


**Fig. 7.** The percentage of sample-size that would be required in tract measures of volume, FA, and MD with the *CNN* pipeline and Segis-Net. The sample-size required for the *Classical* pipeline is used as the reference (100%).

FA and 57.0% for MD. Several percentages of the *CNN* pipeline were smaller than those of the Segis-Net, e.g., in FA measures of CGH and FMA tract (Fig. 7 (b)), but its performance showed to be less stable across tracts than the Segis-Net, which in all settings consistently decreased in the required samples over the *Classical* pipeline. The percentage of reduction was generally similar for left/right homologous tracts except for the MD measure in the left of CGH (Fig. 7 (c)). This large reduction could be related to the MD reproducibility of the *Classical* pipeline, in which the left CGH tract had a much higher variation in errors comparing with that of the other tracts (Fig. 6 (d)).

## 5. Discussion

We developed a single-step deep learning framework, coined Segis-Net, for simultaneous optimization of segmentation and registration. The method was applied to analyze changes in WM tracts from a large set of longitudinal diffusion MRI images. To evaluate the performance of the method, we compared it with two state-of-the-art methods, and two multistage pipelines consisting of independent segmentation and registration components, i.e., the *Classical* and *CNN* pipeline. Segis-Net advanced the state-of-the-art by a higher segmentation and registration

accuracy, and led to improved performances in registration accuracy, spatio-temporal consistency of segmentation, and reproducibility of segmentation and tract-specific measures comparing with the multi-stage pipelines. We evaluated the practical value of the improved performance in terms of sample-size reduction that could be achieved when employing the method. The tract-specific mesures analyzed with Segis-Net would only require 33.0% – 60.5% sample-size of the data for achieving the same effect size as the *Classical* pipeline.

To date most developments in longitudinal analysis frameworks have focused on unbiased ways of registering image time series (Keihaninejad et al., 2013; Metz et al., 2011), in which a multistage approach combing independent segmentation and registration components is often used (De Groot et al., 2013a; Yendiki et al., 2016). In this paper, we aimed to investigate a different way to improve the performance of the longitudinal framework by using a single-step CNN that optimizes both tasks simultaneously. The sole value of simultaneous optimization was demonstrated by the comparison with the *CNN* pipeline. There was no benefit observed for segmentation alone, but for registration, spatio-temporal consistency of segmentation, and reproducibility, simultaneous optimization led to significantly improved performance.

In the evaluation of segmentation performance, similar accuracies for the *CNN* and Segis-Net framework was observed for the six tracts (Fig. 3). Relative segmentation accuracy between individual tracts were in line with those reported in literature (Li et al., 2020a; Wasserthal et al., 2018). For instance, a small and curved object like the FMI tract tended to have a lower Dice coefficient than the larger FMA tracts. For all six tracts, the proposed Segis-Net showed a better segmentation performance than U-ReSNet, an existing simultaneous method (Table 1). We expect the added value of Segis-Net to be related to two factors: (1) the method allows the use of diffusion tensor images for tract segmentation, as we use parallel network modules and only align the predicted segmentation; this circumvents the need to interpolate tensor images. In other words, task-specific inputs can be used; and (2) the sub-branches in the segmentation stream (Fig. 8) are designed for the prediction of white matter tracts which can overlap with each other, unlike the exclusive tissue labels focused by other works.

In the task of registration alone, Segis-Net overall yielded the best accuracy among the methods. It significantly outperformed the *Classical* pipeline for three tracts and the *CNN* pipeline for five tracts (Fig. 4). This is an important observation as (1) it showed that simultaneous optimization was beneficial to one of the individual tasks, and (2) it is non-trivial to improve registration accuracy over a classical algorithm, in which the transformation is pair-wise optimized on the test images. During the comparison with the state-of-the-art methods, we observed two interesting results (Table 2). First, VoxelMorph was the only method that directly optimized on the DC metric, but it led to a lowest DC score. This can be due to the fact that the segmentation labels used in diffusion imaging studies are often independently obtained for each image, which is much less correlated to the registration performance than is the case for atlas-based segmentation (Balakrishnan et al., 2019). As a result, the alignment of "imperfect" segmentation labels can be an obstructive loss term instead. Second, although the MSE of U-ReSNet was almost four times that of the Segis-Net, it achieved a good SC similarity, especially in the thin structures (CGC and CGH). This can be attributed to the formulation of their registration loss as the sum of local cross-correlation and MSE.

In all six tracts, we observed substantially higher spatio-temporal consistency of segmentation and reproducibility of segmentation with Segis-Net than with the two multistage pipelines (Figs. 5, 6). The spatio-temporal consistency of segmentation as quantified by the Dice coefficient ranged 0.81 – 0.87 for Segis-Net, significantly outperforming the *Classical* pipeline for all the six tracts (range: 0.57 – 0.69) and the *CNN* pipeline for five tracts (range: 0.77 – 0.84). The segmentation reproducibility as quantified by Cohen's kappa ranged 0.79 – 0.87 for Segis-Net, significantly higher than the *Classical* pipeline for all the six tracts (range: 0.64 – 0.72) and the *CNN* pipeline for two tracts (range: 0.77 – 0.85). These results indicate that Segis-Net can serve as a reliable alternative to the *Classical* pipeline in spatially capturing macrostructural brain changes over time.

In addition, more significant improvements were observed for the reproducibility of tract-specific volume assessment, but not for the FA and MD measures. For volume reproducibility, Segis-Net yielded the least error in the measurements of scan and re-scan, followed by the *CNN* pipeline (Fig. 6). For the FA and MD measures, we observed relatively similar reproducibility for the three methods, in which significant difference was only observed in FA reproducibility of CGH and left CGC tract. This suggests that diffusion measures are quite robust to variations in the geometry of the segmented tract. It's worth noting that the FA reproducibility of the *Classical* pipeline could be higher than the benchmark of tractography-based segmentation methods, since it is optimized on the FA reproducibility on a subset of the data.

These improved performances have practical values in a power analysis, where both the *CNN* pipeline and Segis-Net showed to be able to reduce the required sample-size to achieve the same statistical power as the *Classical* pipeline. The data processed with Segis-Net would require on average 33.0% of the sample-size for volume measures, 60.5% for FA, and 57.0% for MD measures, requiring consistently a decreased sample-size for all the settings. The averaged percentages for the *CNN* pipeline were 62.7%, 60.5% and 68.7%. For FMI tract, it would, however, require 183% and 124% of the sample-size for the volume and FA measures. The observed dispersion of sample-size reduction with the *CNN* pipeline may suggest that simultaneous optimization was beneficial to the robustness of the method across the concurrently segmented tracts.

Whereas the method is generic, we specifically implemented and optimized it for longitudinal study in diffusion MRI data. In diffusion MRI application, we adopted the commonly used scalar-value FA map as the input for registration. Deformable registration of diffusion tensor images is known to be challenging due to the directional components contained in voxels. Despite developments in classical methods for tensor reorientation during the optimization (Cao et al., 2006; Zhang et al., 2007), for learning-based registration it still largely remains unexplored. With the promising results of diffusion tensor interpolation as shown by Grigorescu et al. (2020), Segis-Net based on solely tensor images would be an interesting direction to explore.

The Segis-Net framework presented in the current study is limited to two time-points. This is because learning-based registration algorithms currently only support pairwise transformations (Balakrishnan et al., 2019). One limitation of our method is therefore that it does not allow for analysis of arbitrary number of time-points. In the present study, we grouped the available triple time-points from the same participant into orderless image-pairs for bidirectional analysis. A future possible improvement of the method could be extending the registration component of Segis-Net to enable learning-based group-wise analysis of a set of time-points (Li et al., 2020b).

Beyond the current application, we expect that this work could be extended to other imaging sequences and for example for segmentation of lesion images. For future work, we plan to adapt the proposed method to analyze brain diseases with large and progressive changes. For instance, registration of brains with lesions due to cortical infarct may benefit from a simultaneous segmentation of infarct regions.

## 6. Conclusion

We proposed a single-step deep learning framework for longitudinal diffusion MRI analysis, in which segmentation and deformable registration were integrated for simultaneous optimization. The comparison with baseline multistage approaches and state-of-the-art methods showed that the proposed Segis-Net can be applied as a reliable tool to support spatio-temporal analysis of WM tracts from longitudinal diffusion MRI imaging. Besides the improved performances, a two-in-one framework for concurrent segmentation and registration also enables a light-weight way of fast quantification of brain changes overtime. This

may lead to a more prominent role for tract-specific biomarkers in applications where tract segmentation and registration are subject to time constraints. With the increasing availability of longitudinal diffusion data, we expect future studies investigating progressive neurodegeneration can greatly benefit from the improved reliability and efficiency of Segis-Net.

## Data availability

The datasets analyzed during the current study are not publicly available. Due to the sensitive nature of the data used in this study, participants were assured raw data would remain confidential and would not be shared.

## Code availability

The code for Segis-Net, the *CNN* pipeline, as well as the implementation of Elastix is available at https://gitlab.com/blibli/segis-net.

## Ethics statement

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians (Ikram et al., 2020).

## Credit authorship contribution statement

**Bo Li:** Conceptualization, Methodology, Software, Formal analysis, Validation, Writing - original draft, Writing - review & editing, Visualization. **Wiro J. Niessen:** Conceptualization, Validation, Resources, Supervision, Writing - review & editing, Funding acquisition. **Stefan Klein:** Methodology, Software, Writing - review & editing. **Marius de Groot:** Software, Writing - review & editing. **M. Arfan Ikram:** Resources, Data curation, Writing - review & editing. **Meike W. Vernooij:** Resources, Investigation, Data curation, Writing - review & editing. **Esther E. Bron:** Formal analysis, Investigation, Validation, Writing - review & editing, Supervision, Project administration.

## Acknowlgedgments

This work was sponsored through grants of the Medical Delta Diagnostics 3.0: Dementia and Stroke, the EU Horizon 2020 project EuroPOND (666992), the Netherlands CardioVascular Research Initiative (Heart-Brain Connection: CVON2012-06, CVON2018-28), and the Dutch Heart Foundation (PPP Allowance, 2018B011).

## Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118004.

## References

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26 (3), 839–851.
Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imag..
Barrick, T.R., Charlton, R.A., Clark, C.A., Markus, H.S., 2010. White matter structural decline in normal ageing: a prospective longitudinal study using tract-based spatial statistics. Neuroimage 51 (2), 565–577.
Bastiaansen, W.A., Rousian, M., Steegers-Theunissen, R.P., Niessen, W.J., Koning, A., Klein, S., 2020. Towards segmentation and spatial alignment of the human embryonic brain using deep learning for atlas-based registration. In: Proceedings of the International Workshop on Biomedical Image Registration. Springer, pp. 34–43.
Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage 34 (1), 144–155.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.
Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., de Bruijne, M., 2019. Semi-supervised medical image segmentation via learning consistency under transformations. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 810–818.
Cao, Y., Miller, M.I., Mori, S., Winslow, R.L., Younes, L., 2006. Diffeomorphic matching of diffusion tensor images. In: Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). IEEE, p. 67.
Cheng, J., Tsai, Y.-H., Wang, S., Yang, M.-H., 2017. Segflow: Joint learning for video object segmentation and optical flow. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 686–695.
Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E., 1999. Tracking neuronal fiber pathways in the living human brain. Proc. Natl. Acad. Sci. 96 (18), 10422–10427.
Crick, F., Jones, E., 1993. Backwardness of human neuroanatomy.. Nature 361 (6408), 109–110.
Crum, W.R., Camara, O., Hill, D.L., 2006. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans. Med. Imaging 25 (11), 1451–1461.
Dai, X., Khorram, S., 1999. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments. IEEE Trans. Geosci. Remote Sens. 37 (5), 2351–2362.
Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.
Diggle, P., Diggle, P.J., Heagerty, P., Liang, K.-Y., Heagerty, P.J., Zeger, S., et al., 2002. Analysis of Longitudinal Data. Oxford University Press.
Dimond, D., Rohr, C.S., Smith, R.E., Dhollander, T., Cho, I., Lebel, C., Dewey, D., Connelly, A., Bray, S., 2020. Early childhood development of white matter fiber density and morphology. Neuroimage 210, 116552.
Estienne, T., Lerousseau, M., Vakalopoulou, M., Alvarez Andres, E., Battistella, E., Carré, A., Chandra, S., Christodoulidis, S., Sahasrabudhe, M., Sun, R., et al., 2020. Deep learning-based concurrent brain registration and tumor segmentation. Front Comput. Neurosci. 14, 17.
Estienne, T., Vakalopoulou, M., Christodoulidis, S., Battistela, E., Lerousseau, M., Carre, A., Klausner, G., Sun, R., Robert, C., Mougiakakou, S., et al., 2019. U-ReSNet: ultimate coupling of registration and segmentation with deep nets. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 310–319.
Farbota, K.D., Bendlin, B.B., Alexander, A.L., Rowley, H.A., Dempsey, R.J., Johnson, S.C., 2012. Longitudinal diffusion tensor imaging and neuropsychological correlates in traumatic brain injury patients. Front Hum. Neurosci. 6, 160.
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.
Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256.
Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C., 2011. Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 532–540.
Grigorescu, I., Uus, A., Christiaens, D., Cordero-Grande, L., Hutter, J., Edwards, A.D., Hajnal, J.V., Modat, M., Deprez, M., 2020. Diffusion tensor driven image registration: a deep learning approach. In: Proceedings of the International Workshop on Biomedical Image Registration. Springer, pp. 131–140.
De Groot, M., Cremers, L.G., Ikram, M.A., Hofman, A., Krestin, G.P., van der Lugt, A., Niessen, W.J., Vernooij, M.W., 2016. White matter degeneration with aging: longitudinal diffusion MR imaging analysis. Radiology 279 (2), 532–541.
De Groot, M., Ikram, M.A., Akoudad, S., Krestin, G.P., Hofman, A., van der Lugt, A., Niessen, W.J., Vernooij, M.W., 2015. Tract-specific white matter degeneration in aging: the rotterdam study. Alzheimer's & Dementia 11 (3), 321–330.
De Groot, M., Verhaaren, B.F., De Boer, R., Klein, S., Hofman, A., van der Lugt, A., Ikram, M.A., Niessen, W.J., Vernooij, M.W., 2013a. Changes in normal-appearing white matter precede development of white matter lesions. Stroke 44 (4), 1037–1042.
De Groot, M., Vernooij, M.W., Klein, S., Ikram, M.A., Vos, F.M., Smith, S.M., Niessen, W.J., Andersson, J.L., 2013b. Improving alignment in tract-based spatial statistics: evaluation and optimization of image registration. Neuroimage 76, 400–411.
Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Noble, J.A., Barratt, D.C., Vercauteren, T., 2018. Label-driven weakly-supervised learning for multimodal deformable image registration. In: Proceedings of the 15th ISBI. IEEE, pp. 1070–1074.
Ikram, M.A., Brusselle, G., Ghanbari, M., Goedegebure, A., Ikram, M.K., Kavousi, M., Kieboom, B.C., Klaver, C.C., de Knegt, R.J., Luik, A.I., et al., 2020. Objectives, design and main findings until 2020 from the rotterdam study. Eur. J. Epidemiol. 1–35.
Ikram, M.A., van der Lugt, A., Niessen, W.J., Krestin, G.P., Koudstaal, P.J., Hofman, A., Breteler, M.M., Vernooij, M.W., 2011. The rotterdam scan study: design and update up to 2012. Eur. J. Epidemiol. 26 (10), 811–824.
Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167.
Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. In: Advances in Neural Information Processing Systems, pp. 2017–2025.
Keihaninejad, S., Zhang, H., Ryan, N.S., Malone, I.B., Modat, M., Cardoso, M.J., Cash, D.M., Fox, N.C., Ourselin, S., 2013. An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer's disease. Neuroimage 72, 153–163.
Kingma, D. P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv:1412.6980.

Klein, S., Pluim, J.P., Staring, M., Viergever, M.A., 2009. Adaptive stochastic gradient descent optimisation for image registration. Int. J. Comput. Vis. 81 (3), 227.
Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2010. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imag. 29 (1), 196–205.
Koppelmans, V., de Groot, M., de Ruiter, M.B., Boogerd, W., Seynaeve, C., Vernooij, M.W., Niessen, W.J., Schagen, S.B., Breteler, M.M., 2014. Global and focal white matter integrity in breast cancer survivors 20 years after adjuvant chemotherapy. Hum. Brain Mapp. 35 (3), 889–899.
van der Krieke, L., Blaauw, F.J., Emerencia, A.C., Schenk, H.M., Slaets, J.P., Bos, E.H., de Jonge, P., Jeronimus, B.F., 2017. Temporal dynamics of health and well-being: a crowdsourcing approach to momentary assessments and automated generation of personalized feedback. Psychosom. Med. 79 (2), 213–223.
Landis, J.R., Koch, G.G., 1977. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 363–374.
Le Bihan, D., Mangin, J.-F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., Chabriat, H., 2001. Diffusion tensor imaging: concepts and applications. J. Magn. Reson. Imaging: Off. J. Int. Soc. Magn. Reson. Med. 13 (4), 534–546.
Lebel, C., Beaulieu, C., 2011. Longitudinal development of human brain wiring continues from childhood into adulthood. J. Neurosci. 31 (30), 10937–10947.
Leemans, A., Jeurissen, B., Sijbers, J., Jones, D., 2009. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In: Proceedings of the 17th Annual Meeting of International Society for Magnetic Resonance in Medicine, 209, p. 3537.
Li, B., de Groot, M., Steketee, R.M., Meijboom, R., Smits, M., Vernooij, M.W., Ikram, M.A., Liu, J., Niessen, W.J., Bron, E.E., 2020a. Neuro4neuro: a neural network approach for neural tract segmentation using large-scale population-based diffusion imaging. Neuroimage 116993.
Li, B., de Groot, M., Vernooij, M.W., Ikram, M.A., Niessen, W.J., Bron, E.E., 2018. Reproducible white matter tract segmentation using 3D U-Net on a large-scale DTI dataset. In: Proceedings of the International Workshop on MLMI. Springer, pp. 205–213.
Li, B., Niessen, W.J., Klein, S., de Groot, M., Ikram, M.A., Vernooij, M.W., Bron, E.E., 2019. A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 645–653.
Li, B., Niessen, W.J., Klein, S., Ikram, M.A., Vernooij, M.W., Bron, E.E., 2020b. Learning unbiased registration and joint segmentation: evaluation on longitudinal diffusion MRI. arXiv:2011.01869.
Maas, A.L., Hannun, A.Y., Ng, A.Y., 2013. Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the ICML, 30, p. 3.
Meijboom, R., Steketee, R., Ham, L., Mantini, D., Bron, E., van der Lugt, A., van Swieten, J., Smits, M., 2019. Exploring quantitative group-wise differentiation of Alzheimers disease and behavioural variant frontotemporal dementia using tract-specific microstructural white matter and functional connectivity measures at multiple time points. Eur Radiol 29 (10), 5148–5159.
Metz, C., Klein, S., Schaap, M., van Walsum, T., Niessen, W.J., 2011. Nonrigid registration of dynamic medical imaging data using nD+ t B-splines and a groupwise optimization approach. Med Image Anal. 15 (2), 238–249.
Niessen, W. J., 2016. MR brain image analysis in dementia: From quantitative imaging biomarkers to ageing brain models and imaging genetics.
Parisot, S., Wells III, W., Chemouny, S., Duffau, H., Paragios, N., 2014. Concurrent tumor segmentation and registration with uncertainty-based sparse non-uniform graphs. Med Image Anal. 18 (4).
Pathak, D., Girshick, R., Dollár, P., Darrell, T., Hariharan, B., 2017. Learning features by watching objects move. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2701–2710.
Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M., 2006. A Bayesian model for joint segmentation and registration. Neuroimage 31 (1), 228–239.
Postelnicu, G., Zollei, L., Fischl, B., 2008. Combined volumetric and surface registration. IEEE Trans. Med. Imaging 28 (4), 508–522.
Qin, C., Bai, W., Schlemper, J., Petersen, S.E., Piechnik, S.K., Neubauer, S., Rueckert, D., 2018. Joint learning of motion estimation and segmentation for cardiac mr image sequences. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 472–480.
Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61 (4), 1402–1418.
Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X., 2017. SVF-Net: Learning deformable image registration using shape matching. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 266–274.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on MICCAI. Springer, pp. 234–241.
Sullivan, E.V., Rohlfing, T., Pfefferbaum, A., 2010. Longitudinal study of callosal microstructure in the normal adult aging brain using quantitative DTI fiber tracking. Dev. Neuropsychol. 35 (3), 233–256.
Vakalopoulou, M., Chassagnon, G., Bus, N., Marini, R., Zacharaki, E.I., Revel, M.-P., Paragios, N., 2018. Atlasnet: multi-atlas non-linear deep networks for medical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 658–666.
Vlontzos, A., Mikolajczyk, K., 2018. Deep segmentation and registration in X-Ray angiography video. arXiv:1805.06406.
Vrooman, H.A., Cocosco, C.A., van der Lijn, F., Stokking, R., Ikram, M.A., Vernooij, M.W., Breteler, M.M., Niessen, W.J., 2007. Multi-spectral brain tissue segmentation using automatically trained k-nearest-neighbor classification. Neuroimage 37 (1), 71–81.
Wasserthal, J., Neher, P., Maier-Hein, K.H., 2018. Tractseg: fast and accurate white matter tract segmentation. Neuroimage 183, 239–253.
Wyatt, P.P., Noble, J.A., 2003. MAP MRF Joint segmentation and registration of medical images. Med. Image Anal. 7 (4), 539–552.
Xu, Z., Niethammer, M., 2019. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 420–429.
Yendiki, A., Reuter, M., Wilkens, P., Rosas, H.D., Fischl, B., 2016. Joint reconstruction of white-matter pathways from longitudinal diffusion MRI data with anatomical priors. Neuroimage 127, 277–286.
Yezzi, A., Zöllei, L., Kapur, T., 2003. A variational framework for integrating segmentation and registration through active contours. Med. Image Anal. 7 (2), 171–185.
Zhang, H., Avants, B.B., Yushkevich, P.A., Woo, J.H., Wang, S., McCluskey, L.F., Elman, L.B., Melhem, E.R., Gee, J.C., 2007. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. IEEE Trans. Med. Imaging 26 (11), 1585–1597.
Zhu, W., Myronenko, A., Xu, Z., Li, W., Roth, H., Huang, Y., Milletari, F., Xu, D., 2020. NeurReg: neural registration and its application to image segmentation. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, pp. 3617–3626.