A Deep Learning Approach to MR-less Spatial Normalization for Tau PET Images

Jennifer Alvén[1]✉, Kerstin Heurling[2], Ruben Smith[3], Olof Strandberg[3], Michael Schöll[2,3,4], Oskar Hansson[3], and Fredrik Kahl[1]

[1] Chalmers University of Technology, Gothenburg, Sweden
alven@chalmers.se
[2] Wallenberg Centre for Molecular and Translational Medicine and the Department of Psychiatry and Neurochemistry, University of Gothenburg, Gothenburg, Sweden
[3] Clinical Memory Research Unit, Lund University, Malmö, Sweden
[4] Dementia Research Centre, Institute of Neurology, University College London, London, UK

Abstract. The procedure of aligning a positron emission tomography (PET) image with a common coordinate system, *spatial normalization*, typically demands a corresponding structural magnetic resonance (MR) image. However, MR imaging is not always available or feasible for the subject, which calls for enabling spatial normalization without MR, *MR-less* spatial normalization. In this work, we propose a template-free approach to MR-less spatial normalization for [18F]flortaucipir tau PET images. We use a deep neural network that estimates an aligning transformation from the PET input image, and outputs the spatially normalized image as well as the parameterized transformation. In order to do so, the proposed network iteratively estimates a set of rigid and affine transformations by means of convolutional neural network regressors as well as spatial transformer layers. The network is trained and validated on 199 tau PET volumes with corresponding ground truth transformations, and tested on two different datasets. The proposed method shows competitive performance in terms of registration accuracy as well as speed, and compares favourably to previously published results.

1 Introduction

One of the necessary criteria for Alzheimer's disease (AD) diagnosis is the presence of abnormal tau protein in the brain. Normal tau protein is expressed ubiquitously in the central nervous system, but starts to aggregate and forms neurofibrillary tangles in AD. The positron emission tomography (PET) tracer [18F]flortaucipir binds to the aggregates of tau formed in AD, which makes it possible to visualize the distribution of tau pathology in the living human brain by means of tau PET [18].

In research settings, tau deposition is typically quantified in a common template space, e.g. defined by a magnetic resonance (MR) template brain image. This common template space provides anatomical information, which is insufficiently available in the PET scan alone, and enables region/voxel-wise analysis

and comparisons. Consequently, this methodology requires spatial alignment of the acquired tau PET with this common coordinate system. This registration procedure, referred to as *spatial normalization*, is the focus of this paper.

Most approaches to spatial normalization utilize MR imaging, that is, the tau PET image is aligned with the template space via a MR image of the same subject. The alignment is typically carried out with image registration tools such as ANTs [1] or SPM [19]. One major disadvantage with this procedure is the need for the subject's MR; MR imaging is not necessarily available, and might be contraindicated or inconvenient for the subject. Enabling tau PET spatial normalization without MR, *MR-less spatial normalization*, would most definitely benefit large scale AD studies.

Common for all previous attempts on MR-less spatial normalization is the use of standard image registration techniques with an explicit PET template as target. The most simple approach uses one fix PET template, either computed as the mean over a set of training PET scans [6, 15], or generated synthetically from MR template images [11]. More refined approaches use a learned PET template model to which the native PET scan is iteratively fitted and aligned. Variants of this approach are linear PET templates [3, 16, 17] and principal component analysis (PCA) models [7, 8]. A deep learning based method for generating a synthetic PET template is proposed in [13], where a subject-specific template is predicted by a generative adversarial network (GAN) for each PET image. Note that a standard, non-learning, registration method is used when aligning the PET scan to the GAN-generated template.

None of the above-mentioned methods concern spatial normalization of PET scans using the tau tracer. These approaches are instead applied to β-amyloid PET or raclopride PET, i.e. PET scans using tracers that bind to other proteins than tau and showing less complex and more predictable uptake patterns. The only work on MR-less spatial normalization for tau PET that we are aware of is presented in [4], where the authors evaluate the three previously published template-based approaches [3, 6, 8] on tau PET scans. However, the paper notes that none of the compared template models (using a mean, linear and PCA template, respectively) capture the full variation of tau PETs. More specifically, a (linear) template model using the 1st PCA eigenvector only explained 28% of the variability within the tau PET dataset (consisting mainly of healthy subjects), while a PCA model, using the 1st and 2nd eigenvector, explained 42% of the total variance. For comparison, [7] reports a PCA model for β-amyloid PET where the 1st eigenvector accounts for 80% of the variability. This makes the template-based methods less robust and unreliable for general tau PET images.

In this work, we propose a template-free approach to MR-less spatial normalization for tau PET. In contrast to all previous attempts on MR-less spatial normalization, our method aligns the native PET scan directly without needing an explicit PET template. We use a deep neural network that takes a PET scan as input and outputs a parameterized transformation as well as the warped image. The implemented network includes convolutional neural network (CNN) regressors as well as spatial transformer layers, and is trainable end-to-end. Our

solution is general, rather than being developed specifically for tau tracers, and can without alterations be applied to any application involving PET spatial normalization, e.g. PET scans of other body parts and/or with other PET tracers. We compare our method to a MR-dependent baseline, and to the template-based methods in [4]. Our approach outperforms the MR-dependent baseline in terms of registration accuracy as well as speed, and compares favourably to the template-based approaches.

2   A Deep Learning Approach to Spatial Normalization

Let $\mathcal{I}$ be the PET image in its original space and $\mathcal{I}_{\text{aligned}}$ be the same image aligned with the common template space via the coordinate transformation T, $\mathcal{I}_{\text{aligned}} = \mathcal{I} \circ T$. We restrict T to be a composition of a rigid transformation T_R and an affine transformation T_A, where T_R and T_A are parameterized as

$$T_R\left(\boldsymbol{r}, \boldsymbol{t}^{\text{rot}}\right) = \begin{bmatrix} R & \boldsymbol{t}^{\text{rot}} \\ \boldsymbol{0} & 1 \end{bmatrix}, \quad R = \exp \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix}, \quad \boldsymbol{t}^{\text{rot}} = \begin{bmatrix} t_x^{\text{rot}} \\ t_y^{\text{rot}} \\ t_z^{\text{rot}} \end{bmatrix}, \quad (1)$$

$$T_A\left(\boldsymbol{a}, \boldsymbol{t}^{\text{aff}}\right) = \begin{bmatrix} A & \boldsymbol{t}^{\text{aff}} \\ \boldsymbol{0} & 1 \end{bmatrix}, \quad A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix}, \quad \boldsymbol{t}^{\text{aff}} = \begin{bmatrix} t_x^{\text{aff}} \\ t_y^{\text{aff}} \\ t_z^{\text{aff}} \end{bmatrix}. \quad (2)$$

We describe the problem of spatially normalizing the input image $\mathcal{I}$ as the task of finding the estimates $\hat{T}_R$ and $\hat{T}_A$ that minimize a loss

$$\mathcal{L} = \mathcal{L}_{\text{MAE}}\left(\mathcal{I}_{\text{aligned}}, \mathcal{I} \circ (\hat{T}_A \cdot \hat{T}_R)\right) = \mathcal{L}_{\text{MAE}}\left(\mathcal{I} \circ (T_A \cdot T_R), \mathcal{I} \circ (\hat{T}_A \cdot \hat{T}_R)\right), \quad (3)$$

where $\mathcal{L}_{\text{MAE}}$ is the mean absolute intensity difference over all pixels.

We propose an inference method that uses a deep network consisting of CNN regressors and spatial transformer layers. We learn the network parameters in a supervised fashion, i.e. with access to the ground truth transformations T_R and T_A. Below, we describe the network in detail.

2.1   Network Design

We design two CNN regressors, $\mathcal{F}_R$ and $\mathcal{F}_A$, that each takes an image as input and outputs a rigid or an affine transformation respectively. Both networks include four consecutive convolutional layers followed by a global average pooling layer and dense layers. All convolutional layers use a $3 \times 3 \times 3$ kernel with a stride of two, ReLU activations and 20, 30, 40 and 50 filters respectively. The rigid regressor $\mathcal{F}_R$ is terminated with six parallel dense layers that each outputs one of the parameters in (1), while the affine regressor $\mathcal{F}_A$ is terminated with twelve parallel dense layers that each outputs one of the parameters in (2).

We pair each CNN with a spatial transformer that warps the input image according to the predicted transformation. Spatial transformer networks, introduced in [12], are designed to make the warping operation differentiable, and

have successfully been used in previously published methods for deep medical image registration, e.g. [2, 10, 14, 20].

We reason that small transformations are easier to predict than large ones, and we therefore let $\hat{T}_R$ and $\hat{T}_A$ be estimated in an iterative manner, more specifically as a composition of smaller transformations, i.e.

$$\hat{T}_R = \hat{T}_R^{(N)} \cdot \ldots \cdot \hat{T}_R^{(2)} \cdot \hat{T}_R^{(1)}, \quad \hat{T}_A = \hat{T}_A^{(M)} \cdot \ldots \cdot \hat{T}_A^{(2)} \cdot \hat{T}_A^{(1)}, \quad (4)$$

where N and M equal the number of rigid and affine transformations. Note that this scheme of estimating several consecutive transformations, instead of a single one, mirrors the scheme that is often present in standard registration methods.

The first rigid transformation $\hat{T}_R^{(1)}$ is estimated by feeding the original image $\mathcal{I}$ to the rigid regressor. Succeeding rigid transformations are computed with the warped image as input, i.e.

$$\hat{T}_R^{(i)} = \mathcal{F}_R \left[\mathcal{I} \circ \left(\hat{T}_R^{(i-1)} \cdot \ldots \cdot \hat{T}_R^{(1)} \right) \right]. \quad (5)$$

Similarly, the first affine transformation $\hat{T}_A^{(1)}$ is computed by feeding a rigidly warped image $\mathcal{I} \circ \hat{T}_R$ to the affine regressor. Succeeding affine transformations are computed with the warped image as input, i.e.

$$\hat{T}_A^{(i)} = \mathcal{F}_A \left[\mathcal{I} \circ \left(\hat{T}_A^{(i-1)} \cdot \ldots \cdot \hat{T}_A^{(1)} \cdot \hat{T}_R \right) \right]. \quad (6)$$

Note that the warped input image is always computed by feeding the spatial transformer with the original image $\mathcal{I}$ and a composition of all so far computed transformations. Thanks to the parametrization in (1) and (2), compositions of transformations can be implemented in a differentiable manner, and the network can be trained end-to-end.

To speed up learning, and to ensure that the rigid regressor outputs reasonable rigid transformations, we use an intermediate loss on the rigidly warped image. Thus, the full loss is given by

$$\mathcal{L} = \mathcal{L}_{\text{MAE}} \left(\mathcal{I} \circ T_R, \mathcal{I} \circ \hat{T}_R \right) + \mathcal{L}_{\text{MAE}} \left(\mathcal{I} \circ (T_A \cdot T_R), \mathcal{I} \circ (\hat{T}_A \cdot \hat{T}_R) \right). \quad (7)$$

In summary, the neural network takes the original PET image as input, computes rigid and affine transformations in an iterative manner and outputs the spatially normalized PET image. Since all the steps are designed to be differentiable, we learn the network parameters using stochastic gradient descent based methods on the loss (7). The framework is exemplified in Figure 1.

3   Experimental Evaluation

Training and validation data. We use a dataset consisting of 199 subjects from the BioFINDER study, see http://biofinder.se/. This dataset contains subjects with a large variability in tau load and diagnosis [18]. The dataset is split into a
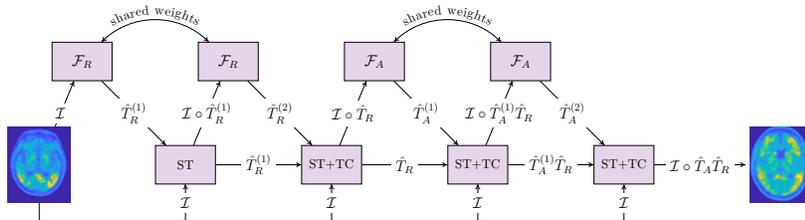
Fig. 1. Schematic illustration of the implemented network when $\hat{T}_R = \hat{T}_R^{(2)} \cdot \hat{T}_R^{(1)}$ and $\hat{T}_A = \hat{T}_A^{(2)} \cdot \hat{T}_A^{(1)}$. The network takes an image $\mathcal{I}$ as input and outputs a composed transformation $\hat{T}_A \cdot \hat{T}_R$ as well as the warped image. The network consists of two CNN regressors, $\mathcal{F}_R$ and $\mathcal{F}_A$, as well as spatial transformer (ST) layers and transformation composition (TC) layers.

training set (160 subjects) and a validation set (39 subjects). For each subject, there is a [18F]flortaucipir standardized uptake value (SUV) PET scan (non-smoothed), as well as ground truth transformations for spatial normalization pre-computed with a MR-dependent pipeline [9]. This pipeline includes preprocessing of the MR images (tailor-made for the BioFINDER data) as well as manual inspection, and when necessary, corrections.

Implementation details. Each PET image is downsampled to a voxel size of $2 \times 2 \times 2$ mm and voxel intensities are standardized image-wise to the range [-1, 1]. We train and validate on full-size images, $91 \times 109 \times 91$ voxels, to learn large-scale translations, as well as on smaller patches, $33 \times 33 \times 33$ voxels, to learn rotations, scaling and shearing. Note that the network allows for different input sizes due to the global average pooling layer. The training patches are augmented by a random brightness shift. Training is done with the Nadam optimizer [5] and with l_2 weight regularization (regularization parameter is set to 10^{-3}). We use a batch size of 10 and 500 training steps per epoch. The validation loss monitors early stopping (training stops after 250 epochs without improvement) and learning rate decay (learning rate is decreased with a factor of 0.95 after 10 epochs without improvement).

Test data. We use two different datasets for evaluation, (i) 37 additional subjects from the BioFINDER study and (ii) 111 subjects from the ADNI database, see http://adni.loni.usc.edu/. Compared to the BioFINDER data, the ADNI test set shows less variability in diagnosis, and contains subjects with a smaller average tau load. The difference in difficulty can be demonstrated by constructing a similar PCA model as in [4] for the both test sets. For the BioFINDER test set, such a model explains 36% of the variability, while the corresponding model explains 51% for the ADNI test set. Thus, we expect the BioFINDER test set to be more challenging for template-based approaches than the ADNI test set. Affine and non-rigid deformations computed by the MR-dependent methods in [4] and [9] are used as ground truth for the ADNI and the BioFINDER data respectively.

Compared baselines. We evaluate our approach on the BioFINDER test set for three setups of rigid and affine transformations estimated in the iterative scheme:

$$(a) \ \hat{T}_R = \hat{T}_R^{(1)}, \ \hat{T}_A = \hat{T}_A^{(1)}, \ (b) \ \hat{T}_R = \hat{T}_R^{(2)} \cdot \hat{T}_R^{(1)}, \ \hat{T}_A = \hat{T}_A^{(1)},$$
$$(c) \ \hat{T}_R = \hat{T}_R^{(2)} \cdot \hat{T}_R^{(1)}, \ \hat{T}_A = \hat{T}_A^{(2)} \cdot \hat{T}_A^{(1)}. \tag{8}$$

We also compare our results on the BioFINDER test set with a standard MR-dependent baseline: the PET image is aligned with the subject's skull-stripped T1 MR image via a rigid transformation, and thereafter, this MR image is aligned with the MR template space via a rigid, an affine and a deformable transformation. We use the ANTs software [1] with default parameters.

We evaluate the top-performing setup from (a)-(c) on the ADNI test set. We compare to the three template-based approaches in [4] (using mean, linear and PCA templates respectively), i.e. all MR-less methods previously evaluated on tau PET that we know of. Unfortunately, the code for these methods is not publicly available, but we have got access to paper's result on the ADNI test set.

Evaluation metrics. To facilitate comparisons between subjects, we evaluate on SUV ratio (SUVR) images. The SUVR images are computed by normalizing the images with respect to the tau uptake in the cerebellum. As performance metrics, we use (i) the absolute percentage error over tau load in the *temporal meta-ROI, defined in* [18], *which is a composite region including parts of the temporal cortex closely connected to tau deposition in AD subjects* (entorhinal cortex, parahippocampal and fusiform gyrus, amygdala, inferior and middle temporal cortices), and (ii) the absolute percentage error over tau load in the whole brain. All ROIs are extracted from the population atlas used in [4]. Note that the same (atlas) ROIs are used when computing the SUVR tau load for the ground truth.

We use the mean residual error over the deformation fields (in mm) as an additional performance metric for the BioFINDER test set. We do not evaluate this metric on the ADNI test set since we do not have access to the competing method's deformation fields.

Experimental results. See Tables 1 and 2 for the experimental results on the BioFINDER and the ADNI test set respectively. The proposed method clearly quantifies tau with less errors, and with more reliability, than the MR-dependent baseline and the compared MR-less approaches. The improvement in tau load quantification is statistically significant for both the temporal meta-ROI (the p-value equals 0.8% for a paired t-test of 'PCA templ.' vs. 'Ours, setup (c)' in Table 2) as well as for all ROIs. Figure 2 shows Bland-Altman plots of the tau deposition in the temporal meta-ROI for our top-performing approach (c) and the methods in [4] evaluated on the ADNI test set. The Bland-Altman plots confirm that the proposed method handles the most challenging cases, i.e. where the tau load error is larger than ±0.05, better than the template-based approaches. Our approach runs for 1.4 seconds per subject on a GeForce GTX 1080 Ti, significantly faster than the MR-dependent baseline, which runs for 30 minutes per subject on an Intel i7-8700 CPU. The run times for the compared methods in [4] are not reported.

Table 1. Absolute percentage error over (SUVR) tau load in the temporal meta-ROI ('SUVR-temp') and in the whole brain ('SUVR-all'), as well as the mean residual error over the deformation fields ("MRE"), for three different setups of our approach (a)-(c) and the MR-dependent baseline ("MR-baseline"), evaluated on the BioFINDER test set. The table reports the mean ± standard deviation for the BioFINDER test set.

	Ours, setup (a)	Ours, setup (b)	Ours, setup (c)	MR-baseline
SUVR-temp	$2.30\% \pm 2.41\%$	$2.14\% \pm 2.57\%$	$\mathbf{1.86}\% \pm \mathbf{1.95}\%$	$2.04\% \pm 2.58\%$
SUVR-all	$1.70\% \pm 1.55\%$	$1.25\% \pm 2.01\%$	$\mathbf{1.21}\% \pm \mathbf{1.31}\%$	$2.17\% \pm 2.52\%$
MRE	$4.2\text{mm} \pm 5.3\text{mm}$	$3.2\text{mm} \pm 3.7\text{mm}$	$\mathbf{2.9\text{mm} \pm 2.2\text{mm}}$	$6.7\text{mm} \pm 4.6\text{mm}$

Table 2. Absolute percentage error over (SUVR) tau load in the temporal meta-ROI ('SUVR-temp') and in the whole brain ('SUVR-all') for the top-performing setup from Table 1, and for the methods in [4], evaluated on the ADNI test set. The table reports the mean ± standard deviation.

	[4], mean templ.	[4], linear templ.	[4], PCA templ.	Ours, setup (c)
SUVR-temp	$1.72\% \pm 2.48\%$	$1.77\% \pm 2.91\%$	$1.57\% \pm 1.47\%$	$\mathbf{1.53}\% \pm \mathbf{1.38}\%$
SUVR-all	$1.47\% \pm 2.21\%$	$1.53\% \pm 2.67\%$	$1.41\% \pm 1.34\%$	$\mathbf{1.16}\% \pm \mathbf{0.89}\%$

4   Concluding Discussion

This work proposes a template-free approach to MR-less spatial normalization of [18F]flortaucipir PET scans, enabled by a deep neural network consisting of CNN regressors and spatial transformer layers. The network is trained and validated on 199 PET scans from the BioFINDER study.

We evaluate our method on 37 additional BioFINDER subjects, and outperform the standard, significantly slower, MR-dependent baseline. This shows that our method is robust and able to handle the high diversity among the subjects in the dataset without needing manual supervision. In addition, we compare the proposed method to the three methods in [4] evaluated on 111 test subjects from the ADNI database, which is a less challenging dataset. We perform better than
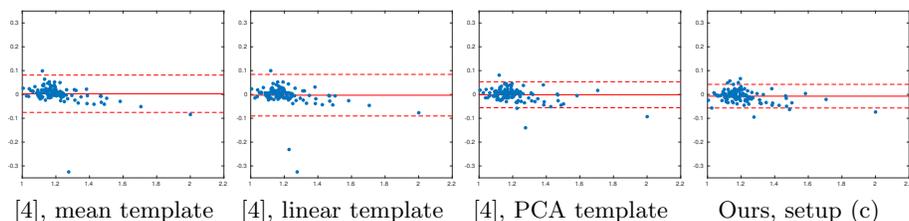

[4], mean template   [4], linear template   [4], PCA template   Ours, setup (c)

Fig. 2. Bland-Altman plot of the temporal meta-ROI SUVR for the compared methods. The figures show the difference in tau deposition between the MR-less and the MR-dependent method against the average of the two methods.

the template-based approaches, despite that (i) our network is trained exclusively on BioFINDER data and run without fine-tuning on the ADNI test set and (ii) the compared approaches include non-rigid deformation while our does not.

Two possible extensions are (i) a more thorough evaluation of the optimal number of rigid and affine transformations used in the iterative scheme, and (ii) to implement and evaluate non-rigid deformation on top of these. A third extension could be to consider other loss functions, such as the difference between the estimated and the ground truth deformation fields.

It would be informative to evaluate our approach on other PET ligands than the tau tracer, such as a β-amyloid tracer. Further, the proposed approach could, with negligible modifications, be applied to tasks involving other modalities, e.g. as an alternative to multi-atlas based brain region parcellation on MR images. This would yield a larger selection of literature to compare with, as well as establish the applicability of our method. To allow for such future comparisons, we will make the code publicly available.

To enable a stand-alone tool for MR-less analysis of [18F]flortaucipir PET scans, we do not only need MR-less spatial normalization. Region-wise tau quantification demands region segmentations, which are most often derived from the MR image. It is questionable whether the anatomical information in a PET scan alone is sufficient for retrieving these regions, but a tool able to do so would surely challenge the MR-dependent gold standard in tau PET analysis.

References

1. Avants, B.B., et al.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3), 2033–2044 (2011)
2. Balakrishnan, G., et al.: An unsupervised learning model for deformable medical image registration. In: Proc. CVPR. pp. 9252–9260 (2018)
3. Bourgeat, P., et al.: Comparison of MR-less PiB SUVR quantification methods. Neurobiology of Aging **36**, S159–S166 (2015)
4. Bourgeat, P., et al.: PET-only 18F-AV1451 tau quantification. In: Proc. ISBI. pp. 1173–1176 (2017)
5. Dozat, T.: Incorporating Nesterov momentum into Adam (2016)
6. Edison, P., et al.: Comparison of MRI based and PET template based approaches in the quantitative analysis of amyloid imaging with PIB-PET. NeuroImage **70**, 423–433 (2013)
7. Fripp, J., et al.: Appearance modeling of 11C PiB PET images: characterizing amyloid deposition in Alzheimer's disease, mild cognitive impairment and healthy aging. NeuroImage **43**(3), 430–439 (2008)
8. Fripp, J., et al.: MR-less high dimensional spatial normalization of 11 C PiB PET images on a population of elderly, mild cognitive impaired and Alzheimer disease patients. In: Proc. MICCAI. pp. 442–449 (2008)
9. Hansson, O., et al.: Tau pathology distribution in Alzheimer's disease corresponds differentially to cognition-relevant functional brain networks. Front. Neurosci. **11**, 167 (2017)
10. Hu, Y., et al.: Label-driven weakly-supervised learning for multimodal deformable image registration. In: Proc. ISBI. pp. 1070–1074 (2018)

11. Hutton, C., et al.: Quantification of 18 F-florbetapir PET: comparison of two analysis methods. Eur. J. Nucl. Med. Mol. Imaging **42**(5), 725–732 (2015)
12. Jaderberg, M., et al.: Spatial transformer networks. In: Proc. NIPS. pp. 2017–2025 (2015)
13. Kang, S.K., et al.: Adaptive template generation for amyloid PET using a deep learning approach. Human Brain Mapping (2018)
14. Krebs, J., et al.: Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 101–109. Springer (2018)
15. Kuhn, F.P., et al.: Comparison of PET template-based and MRI-based image processing in the quantitative analysis of C 11-raclopride PET. Eur. J. Nucl. Med. Mol. Imaging **4**(1), 7 (2014)
16. Lilja, J., et al.: Spatial normalization of [18F] flutemetamol PET images utilizing an adaptive principal components template. J. Nucl. Med. (2018)
17. Lundqvist, R., et al.: Implementation and validation of an adaptive template registration method for 18F-flutemetamol imaging data. J. Nucl. Med. **54**(8), 1472–1478 (2013)
18. Ossenkoppele, R., et al.: Discriminative accuracy of [18F]flortaucipir positron emission tomography for Alzheimer disease vs other neurodegenerative disorders. JAMA **320**(11), 1151–1162 (2018)
19. Penny, W.D., et al.: Statistical parametric mapping: the analysis of functional brain images. Elsevier (2011)
20. de Vos, B.D., et al.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 204–212. Springer (2017)