NeurReg: Neural Registration and Its Application to Image Segmentation

Wentao Zhu, Andriy Myronenko, Ziyue Xu, Wenqi Li
Holger Roth, Yufang Huang[1], Fausto Milletari, Daguang Xu
NVIDIA [1]Cornell University
wentaoz,amyronenko,ziyuex,wenqil,hroth,fmilletari,daguangx@nvidia.com
yfhuang1992new@gmail.com

Abstract

Registration is a fundamental task in medical image analysis which can be applied to several tasks including image segmentation, intra-operative tracking, multi-modal image alignment, and motion analysis. Popular registration tools such as ANTs and NiftyReg optimize an objective function for each pair of images from scratch which is time-consuming for large images with complicated deformation. Facilitated by the rapid progress of deep learning, learning-based approaches such as VoxelMorph have been emerging for image registration. These approaches can achieve competitive performance in a fraction of a second on advanced GPUs. In this work, we construct a neural registration framework, called NeurReg, with a hybrid loss of displacement fields and data similarity, which substantially improves the current state-of-the-art of registrations. Within the framework, we simulate various transformations by a registration simulator which generates fixed image and displacement field ground truth for training. Furthermore, we design three segmentation frameworks based on the proposed registration framework: 1) atlas-based segmentation, 2) joint learning of both segmentation and registration tasks, and 3) multi-task learning with atlas-based segmentation as an intermediate feature. Extensive experimental results validate the effectiveness of the proposed NeurReg framework based on various metrics: the endpoint error (EPE) of the predicted displacement field, mean square error (MSE), normalized local cross-correlation (NLCC), mutual information (MI), Dice coefficient, uncertainty estimation, and the interpretability of the segmentation. The proposed NeurReg improves registration accuracy with fast inference speed, which can greatly accelerate related medical image analysis tasks.

1. Introduction

Image registration tries to establish the correspondence between objects, edges, surfaces or landmarks in different images and it is critical to many clinical tasks such as image fusion, organ atlas creation, and tumor growth monitoring [17]. Manual image registration is laborious and lacks reproducibility which causes potentially clinical disadvantage. Therefore, automated registration is desired in many clinical settings. Generally, registration can be necessary to analysis sequential data [46] or a pair of images from different modalities, acquired at different times, from different viewpoints or even from different patients. Thus designing a robust image registration can be challenging due to the high variability.

Traditional registration methods are based on estimation of the displacement field by optimizing certain objective functions. Such displacement field can be modeled in several ways, e.g. elastic-type models [5, 36], free-form deformation (FFD) [34], Demons [40], and statistical parametric mapping [2]. Beyond the deformation model, diffeomorphic transformations [22] preserve topology with exact inverse transforms and many methods adopt them such as LDDMM [8], SyN [3] and DARTEL [1]. One limitation of these methods is that the optimization can be computationally expensive.

Deep learning-based registration methods have recently been emerging as a viable alternative to the above conventional methods [31, 39, 43]. These methods employ sparse/weak label of registration field, or conduct supervised learning purely based on registration field, inducing high sensitivity on registration field during training. Recent unsupervised deep learning-based registrations, such as VoxelMorph [6], are facilitated by a spatial transformer network [19, 14, 13]. VoxelMorph is also further extended to diffeomorphic transformation and Bayesian framework [13]. NMSR employs self-supervised optimization and multi-scale registration to handle domain shift and large deformation [44]. However from the performance perspective, most of these registration methods have comparable accuracy as traditional iterative optimization methods, although with potential speed advantages.

In this work, we design a deep learning-based registra-

tion framework with a hybrid loss based on data similarity and registration field supervision. The framework, as illustrated in Fig. 1, is motivated by the fact that supervised learning can predict accurate displacement field, while unsupervised learning can extract visual representations that generalize well to unseen images. More specifically, we build a registration simulator which models random translation, rotation, scale, and elastic deformation. For training, we employ the registration simulator to generate a fixed image with its corresponding displacement field ground truth from a given moving image. Then, we use a U-Net to parameterize the displacement field and a spatial transform network to warp the moving image towards the generated fixed image [32, 19]. In addition to the hybrid loss and registration simulator for registration itself, we further investigated its potential in segmentation, a common application of deformable registration. We design two different multi-task learning networks between segmentation and registration. Also, to fully exploit the capacity with one moving/atlas image, we further design a dual registration to enforce registration loss and segmentation loss from both a random training image and a random image from registration simulator.

Our main contributions are as follows: 1) we design a registration simulator to model various transformations, and employ a hybrid loss with registration field supervision loss and data similarity loss to benefit from both accurate supervision and powerful generalization of deep appearance representation. 2) We design a dual registration scheme in the multi-task learning between registration and segmentation to fully exploit the capacity of one moving/atlas image. We expect the network to align the moving image with a pair of random images from the registration field and training set. A residual segmentation block is further developed to boost the performance as illustrated in Fig. 2. 3) We validate our framework on two widely used public datasets: the Hippocampus and Prostate datasets from the Medical Segmentation Decathlon [38]. Our framework outperforms three popular public toolboxes, ANTs [4], NiftyReg [29] and VoxelMorph [6], on both registration and segmentation tasks based on several evaluation metrics.

2. Related Work

Deep learning-based medical image registration can be primarily categorized into three classes, deep iterative registration, supervised, and unsupervised transformation estimation [17]. Early deep learning-based registrations directly embed deep learning as a visual feature extractor into traditional iterative registration based on handcrafted metrics such as sum of squared differences (SSD), cross-correlation (CC), mutual information (MI), normalized cross correlation (NCC) and normalized mutual information (NMI). Wu et al. [42] employed a stacked convolutional auto-encoder to extract features for mono-modal deformable registration based on NCC. Blendowski et al. [9] combine CNN feature with MRF-based feature in mono-modal registration. Reinforcement learning is further used to mimic the iterative transformation estimation. Liao et al. [24] use reinforcement learning and a greedy supervision to conduct rigid registration. Kai et al. [26] further use Q-learning and contextual feature to perform rigid registration. Miao et al. [28] then employ multi-agent-based reinforcement learning in the rigid registration. Krebs et al. [23] conduct deformable registration by reinforcement learning on low resolution deformation with fuzzy action control. The iterative approaches can be relatively slow compared with the direct transformation estimation.

Supervised transformation uses a neural network to estimate transformation parameters directly which can significantly speed up the registration. AIRNet uses a CNN to directly estimate rigid transformation [11]. Rothe et al. [31] use a U-Net to estimate the deformation field. Cao et al. [10] perform displacement field estimation based on image patches and an equalized activate-points guided sampling is proposed to facilitate the training. Sokooti et al. [39] employ random deformation field to augment the dataset and design a multi-scale CNN to predict a deformation field. Uzunova et al. [41] use a CNN to fit registration field from statistical appearance models. The supervised transformation estimation heavily relies on the quality/diversity of registration field ground truth generated synthetically or manually from expert. Unsupervised registration is desirable to learn representation from data to increase generalization.

Unsupervised transformation estimation mainly uses spatial transformer networks (STN) to warp moving image with estimated registration field, and training of the estimators relies on the design of data similarity function and smoothness of estimated registration field [19]. Neylon et al. [30] model the relationship between similarity function and TRE. VoxelMorph is designed as a general method for unsupervised registration and further extended to variational inference for deformation field [13, 6, 7]. Adversarial similarity network further employs a discriminator to automatically learn the similarity function [15]. Jiang et al. [20] instead learn parameterization of multi-grid B-Spline by a CNN. NMSR employs self-supervised optimization and multi-scale registration to handle domain shift and large deformation [44]. These unsupervised transformation estimations produce comparable accuracy and are significantly faster than the traditional registration tools [6, 13].

Different from the aforementioned methods, we design a neural registration, NeurReg, by taking advantage of both accurate supervision on registration field and good generalization of unsupervised similarity objective function. For registration-based segmentation, a dual registration framework is proposed by enforcing the moving/atlas image to align with random images from both training set and regis-

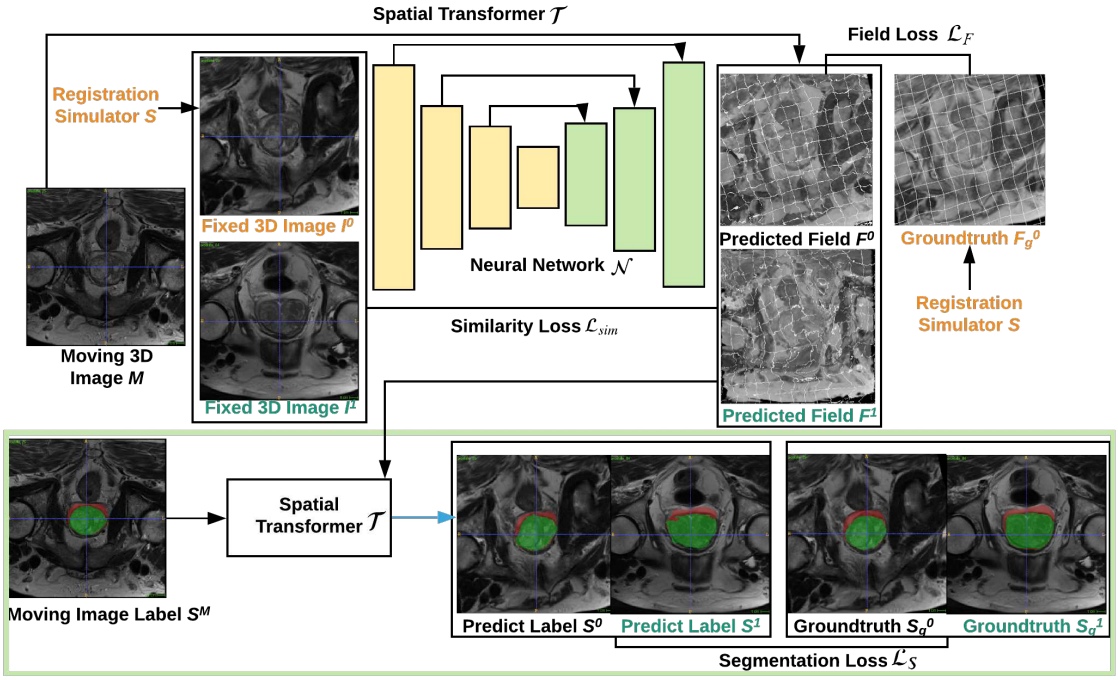

Figure 1. Framework of the proposed NeurReg. In the registration module, we generate a random simulated displacement field and fixed image given a moving image. During training, a hybrid loss consisting of displacement field loss and data similarity loss is employed to take advantage of accurate field supervision and powerful generalization of appearance similarity. For the registration-based segmentation, a dual registration scheme is designed with multi-task loss for registration and segmentation. The dual registration enforces the moving/atlas image to be aligned with random images from both the registration simulator and the training set.

tration simulator which is capable to model various transformations. The NeurReg obtains better performance than ANTs [4], NiftyReg [29] and VoxelMorph [6] and can be a new baseline for medical image registration.

3. Neural Registration Framework

In this section, we introduce three main components in the neural registration framework, registration simulator, neural registration and registration-based segmentation, as illustrated in Fig. 1.

3.1. Registration Simulator

Given a moving/atlas image $\boldsymbol{M}$, we generate a random deformation $\boldsymbol{F}_g^0$ and a fixed image $\boldsymbol{I}^0$ from a registration simulator $\mathcal{S}$. The generated random deformation field $\boldsymbol{F}_g^0$ can be used as an accurate supervised loss $\mathcal{L}_F$ for predicted registration field $\boldsymbol{F}^0$ parameterized by a neural network $\mathcal{N}$. To fully learn representation from data, a similarity loss $\mathcal{L}_{sim}$ is employed to measure the discrepancy between the fixed image $\boldsymbol{I}_0$ and warped image $\mathcal{T}(\boldsymbol{M}, \boldsymbol{F}^0)$ through spatial transformer $\mathcal{T}$ [19].

To fully model various transformations, we simulate random rotation, scale, translation and elastic deformation in the registration simulator $\mathcal{S}$ [37]. We uniformly generate rotation angles $\boldsymbol{a} \sim U(0, \boldsymbol{A})$ for all the dimensions. After that, we uniformly sample scale factors $\boldsymbol{c} \sim U(\boldsymbol{C}_{min}, \boldsymbol{C}_{max})$. If the element in $\boldsymbol{c}$ is smaller than one, it shrinks the moving image. Otherwise, it enlarges the moving image. We then uniformly sample a translation factor $\boldsymbol{l} \sim U(-\boldsymbol{L}, \boldsymbol{L})$. To model the elastic distortion, we firstly randomly generate coordinate offset from Gaussian distribution with standard deviation $\gamma \sim U(0, \Gamma)$. We further apply a multidimensional Gaussian filter with standard deviation $\sigma \sim U(\boldsymbol{\Sigma}_{min}, \boldsymbol{\Sigma}_{max})$ to make the generated coordinate offset smooth and realistic.

During training, we generate the registration field ground truth $\boldsymbol{F}_g^0$ and fixed image $\boldsymbol{I}^0$ on the fly for each batch

$$\begin{aligned} \boldsymbol{F}_g^0 &= \mathcal{S}(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{l}, \gamma, \sigma | \boldsymbol{M}), \\ \boldsymbol{I}^0 &= \mathcal{T}(\boldsymbol{M}, \boldsymbol{F}_g^0). \end{aligned} \tag{1}$$

For the registration-based segmentation, the segmentation ground truth $\boldsymbol{S}_g^0$ of fixed image $\boldsymbol{I}^0$ can be generated by

$$\boldsymbol{S}_g^0 = \mathcal{T}(\boldsymbol{S}^M, \boldsymbol{F}_g^0), \tag{2}$$

where $\boldsymbol{S}^M$ is the segmentation ground truth of the moving/atlas image $\boldsymbol{M}$ in the training set.

3.2. Neural Registration

After we obtain the fixed image $\boldsymbol{I}^0$ with moving image $\boldsymbol{M}$, we design the neural registration to estimate the registration field $\boldsymbol{F}^0$. We employ a neural network $\mathcal{N}$ to parameterize the registration field

$$\boldsymbol{F}^0 = \mathcal{N}(\boldsymbol{M}, \boldsymbol{I}^0; \boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{\theta}$ is the parameters in the neural network.

Different from unsupervised transformation estimation, we design a registration field supervised loss $\mathcal{L}_F$ based on generated registration field ground truth $\boldsymbol{F}_g^0$ from registration simulator

$$\mathcal{L}_F(\boldsymbol{F}^0, \boldsymbol{F}_g^0; \boldsymbol{\theta}) = \frac{1}{|\Omega|} \sum_{\boldsymbol{p} \in \Omega} ||\boldsymbol{F}^0(\boldsymbol{p}) - \boldsymbol{F}_g^0(\boldsymbol{p})||_{L_2}, \tag{4}$$

where $\boldsymbol{p}$ is the pixel position in the image coordinate space Ω. The field supervised loss in Eq. 4 is the endpoint error (EPE) which is an accurate loss for image matching measuring the alignment of two displacement fields [47]. The registration field ground truth is smooth and the model can learn the smoothness of estimated registration field from registration simulator $\mathcal{S}$. Unlike the bending energy loss used in other unsupervised transformation estimation [6], the registration field supervised loss is accurate and the loss can steadily decrease with the decrease of data similarity loss experimentally in Fig. 5.

However, the model might heavily rely on the quality and diversity of the simulated registration field. To improve the model's generalization ability, we further employ data similarity loss as an auxiliary loss inspired by the unsupervised transformation estimation. To train the network in an end-to-end manner, a spatial transformer network $\mathcal{T}$ is used to obtain the reconstructed image $\boldsymbol{I}_R^0$ by warping the moving image $\boldsymbol{M}$ with the estimated registration field $\boldsymbol{F}^0$ [19]

$$\boldsymbol{I}_R^0 = \mathcal{T}(\boldsymbol{M}, \boldsymbol{F}^0). \tag{5}$$

We use the negative normalized local cross-correlation which is robust to evaluate similarity between MRI images

$$\mathcal{L}_{sim}(\boldsymbol{I}^0, \boldsymbol{I}_{\boldsymbol{R}}^0; \boldsymbol{\theta}) = \\ -\frac{1}{|\Omega|} \sum_{\boldsymbol{p} \in \Omega} \frac{\left(\sum_{\boldsymbol{p}_i} (\boldsymbol{I}^0(\boldsymbol{p}_i) - \overline{\boldsymbol{I}^0(\boldsymbol{p})})(\boldsymbol{I}_R^0(\boldsymbol{p}_i) - \overline{\boldsymbol{I}_R^0(\boldsymbol{p})})\right)^2}{\sum_{\boldsymbol{p}_i} (\boldsymbol{I}^0(\boldsymbol{p}_i) - \overline{\boldsymbol{I}^0(\boldsymbol{p})})^2 \sum_{\boldsymbol{p}_i} (\boldsymbol{I}_R^0(\boldsymbol{p}_i) - \overline{\boldsymbol{I}_R^0(\boldsymbol{p})})^2}, \tag{6}$$

where $\boldsymbol{p}_i$ is the pixel position within a window around $\boldsymbol{p}$, and $\overline{\boldsymbol{I}^0(\boldsymbol{p})}$ and $\overline{\boldsymbol{I}_R^0(\boldsymbol{p})}$ are local means within the window around pixel position $\boldsymbol{p}_i$ in $\boldsymbol{I}^0$ and $\boldsymbol{I}_R^0$ respectively.

For NeurReg, we employ the hybrid loss between data similarity loss and registration field supervised loss to train the neural network $\mathcal{N}$

$$\mathcal{L}_{reg}(\boldsymbol{F}^0, \boldsymbol{F}_g^0, \boldsymbol{I}^0, \boldsymbol{I}_R^0; \boldsymbol{\theta}) = \mathcal{L}_F(\boldsymbol{F}^0, \boldsymbol{F}_g^0; \boldsymbol{\theta}) \\ + \lambda \mathcal{L}_{sim}(\boldsymbol{I}^0, \boldsymbol{I}_R^0; \boldsymbol{\theta}), \tag{7}$$

where λ is the hyper-parameter to balance the two losses. In the inference, given a moving/atlas image $\boldsymbol{M}$ and a fixed image $\boldsymbol{I}$, the registration field $\boldsymbol{F}$ can be estimated instantly by Eq. 3 and reconstructed image $\boldsymbol{I}_R$ can be further calculated from Eq. 5.

3.3. Registration-Based Segmentation

The estimated registration field can be applied to transforming segmentation mask for image segmentation purpose. From section 3.2, we obtain the estimated registration field $\boldsymbol{F}$ from Eq. 3 given a moving/atlas image $\boldsymbol{M}$ and a test image as the fixed image $\boldsymbol{I}$. With the segmentation ground truth $\boldsymbol{S}^M$ of moving image, we can further obtain the predicted segmentation mask $\boldsymbol{S}$ through warping $\boldsymbol{S}^M$ with nearest neighbor re-sampling.

More importantly, if ground truths of the segmentation are available during training, we can utilize them to further tune the registration network. This joint learning of segmentation and registration tasks can be beneficial for both because the tasks are highly correlated [33].

In registration-based segmentation by multi-task learning (MTL), we introduce a dual registration scheme to fully exploit the power of registration simulator in the proposed NeurReg as illustrated in Fig. 1. We expect the moving/atlas image $\boldsymbol{M}$ to be aligned with images $\boldsymbol{I}^1$ and $\boldsymbol{I}^0$ from both the dataset and registration simulator $\mathcal{S}$. In the segmentation scenario, the registration simulator $\mathcal{S}$ acts as a data augmentation which is crucial to medical image segmentation because medical image dataset is typically small and dense annotation of segmentation is expensive.

The other registration in dual registration for moving/atlas image $\boldsymbol{M}$ and a random image $\boldsymbol{I}^1$ from dataset can be obtained

$$\begin{aligned} \boldsymbol{F}^1 &= \mathcal{N}(\boldsymbol{M}, \boldsymbol{I}^1; \boldsymbol{\theta}), \\ \boldsymbol{I}_R^1 &= \mathcal{T}(\boldsymbol{M}, \boldsymbol{F}^1). \end{aligned} \tag{8}$$

Given the segmentation ground truth $\boldsymbol{S}^M$ of a moving/atlas image $\boldsymbol{M}$, we can obtain the segmentation ground truth $\boldsymbol{S}_g^0$ from Eq. 2. If the segmentation ground truth $\boldsymbol{S}_g^1$ is available, we can further design a segmentation loss based on dual registration into the framework with Tversky loss [35, 45]

$$\begin{aligned} &\mathcal{L}_{seg}(\boldsymbol{S}^0, \boldsymbol{S}_g^0, \boldsymbol{S}^1, \boldsymbol{S}_g^1) = \mathcal{D}(\boldsymbol{S}^0, \boldsymbol{S}_g^0) + \mathcal{D}(\boldsymbol{S}^1, \boldsymbol{S}_g^1), \\ &\mathcal{D}(\boldsymbol{S}^0, \boldsymbol{S}_g^0) = -\frac{1}{C} \sum_{c=0}^{C-1} \frac{\sum_{\boldsymbol{p}} 2\boldsymbol{S}^0(\boldsymbol{p}) \boldsymbol{S}_g^0(\boldsymbol{p})}{\sum_{\boldsymbol{p}} \boldsymbol{S}^0(\boldsymbol{p}) + \boldsymbol{S}_g^0(\boldsymbol{p})}, \end{aligned} \tag{9}$$

where $\boldsymbol{S}_g^0$ and $\boldsymbol{S}_g^1$ are one hot form of segmentation ground truth and $\boldsymbol{S}^0$ and $\boldsymbol{S}^1$ are continuous values with interpolation order one in the spatial transformer $\mathcal{T}$, C is the number

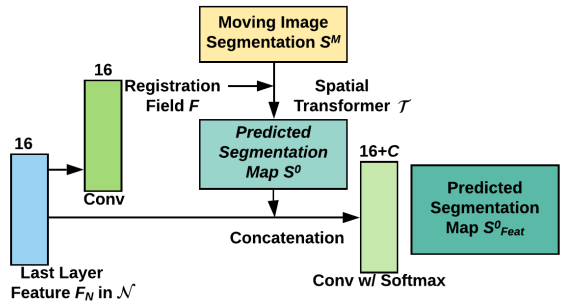

Figure 2. A residual block designed to boost the segmentation.

of classes. The loss function $\mathcal{L}_{reg}^{MTL}$ in the MTL turns to

$$\begin{aligned}\mathcal{L}_{reg}^{MTL}(\boldsymbol{F}^0, \boldsymbol{F}_g^0, \boldsymbol{I}^0, \boldsymbol{I}_R^0, \boldsymbol{I}^1, \boldsymbol{I}_R^1; \boldsymbol{\theta}) = \mathcal{L}_F(\boldsymbol{F}^0, \boldsymbol{F}_g^0; \boldsymbol{\theta}) \\ + \lambda\big(\mathcal{L}_{sim}(\boldsymbol{I}^0, \boldsymbol{I}_R^0; \boldsymbol{\theta}) + \mathcal{L}_{sim}(\boldsymbol{I}^1, \boldsymbol{I}_R^1; \boldsymbol{\theta})\big) \\ + \beta\mathcal{L}_{seg}(\boldsymbol{S}^0, \boldsymbol{S}_g^0, \boldsymbol{S}^1, \boldsymbol{S}_g^1),\end{aligned} \tag{10}$$

where β controls the weight of segmentation loss. At inference time, we obtain the final segmentation prediction by taking argmax along the class dimension.

Inspired by residual network and boosting concept [18, 16], we further extend the framework by introducing an extra convolutional layer with predicted segmentation $\boldsymbol{S}^0$ and the last layer feature $\boldsymbol{F}_N$ from neural network $\mathcal{N}$ as input and softmax activation function to further improve the segmentation as illustrated in Fig. 2

$$\boldsymbol{S}_{Feat}^0 = \operatorname{Softmax}(\operatorname{Conv}([\boldsymbol{F}_N, \boldsymbol{S}^0])), \tag{11}$$

where $\boldsymbol{S}_{Feat}^0$ is the segmentation prediction using aligned segmentation prediction $\boldsymbol{S}^0$ as feature and $[\cdot, \cdot]$ is the concatenation along channel dimension. The segmentation loss based on Tversky loss in Eq. 9 can be easily adapted in the feature based segmentation.

4. Experiments

We conduct experiments on the Hippocampus and Prostate MRI datasets from medical segmentation decathlon to fully validate the proposed NeurReg [38].

4.1. Datasets and Experimental Settings

On the Hippocampus dataset, we randomly split the dataset into 208 training images and 52 test images. There are two foreground categories in the segmentation, hippocampus head and hippocampus body. Because the image size is within $48 \times 64 \times 48$ voxels, we use a $5 \times 5 \times 5$-voxel window in the similarity loss $\mathcal{L}_{sim}$ in Eq. 6.

On the Prostate dataset, we randomly split the dataset into 25 training images and seven test images. The Prostate dataset consists of T2 weighted and apparent diffusion coefficient (ADC) MRI scans. Because of the low signal to noise ratio in ADC scans, we only use the T2 weighted channel. There are two foreground categories, prostate peripheral zone and prostate central gland. Because the image size is around $96 \times 240 \times 240$ which is larger than the Hippocampus dataset, we use a $9 \times 9 \times 9$-voxel window in Eq. 6.

We re-sample the MR images to $1 \times 1 \times 1\ \mathrm{mm}^3$ spacing. To reduce the discrepancy of the intensity distributions of the MR images, we calculate the mean and standard deviation of each volume and clip each volume up to six standard deviations. Finally, we linearly transform each 3D image into range $[0, 1]$.

The hyper-parameters in the registration simulator $\mathcal{S}$ are empirically set as $\boldsymbol{A} = (\frac{1}{6}\pi, \frac{1}{6}\pi, \frac{1}{6}\pi)$, $\boldsymbol{C}_{min} = (0.75, 0.75, 0.75)$, $\boldsymbol{C}_{max} = (1.25, 1.25, 1.25)$, $\boldsymbol{L} = (0.02, 0.02, 0.02)$, $\Gamma = 1000$, $\Sigma_{min} = 10$, $\Sigma_{min} = 13$ based on data distribution. We use 3D U-Net type network as $\mathcal{N}$ [32, 12]. There are four layers in the encoder with numbers of channels $(16, 32, 32, 32)$ with stride as two in each layer. After that we use two convolutional layers with number of channels 32 and 32. For the decoder, we use four blocks of up-sampling, concatenation and convolution with numbers of channels $(32, 32, 32, 16)$. Finally, we use a convolutional layer with number of channels of 16. LeakyReLU is used with slop of 0.2 in each convolutional layer in the encoder and decoder [27]. We use Adam optimizer with learning rate 10^{-4} [21], and numbers of epochs of 1,500 and 2,000 for the Hippocampus and Prostate datasets respectively. Because of small prostate dataset, large image size and slow training, we initialize the model by pretrained model on Hippocampus dataset. Because the registration field supervised loss in Eq. 4 is relative large, we set λ and β as 10 to balance the three losses. Most of the hyper-parameter settings are the same as VoxelMorph for fair comparisons [6]. And we further use the recommended hyper-parameters for ANTs and NiftyReg in [6], because the field of view in the hippopotamus and prostate images is already roughly aligned during image acquisition. We use three scales with 200 iterations each, B-Spline SyN step size of 0.25, updated field mesh size of five for ANTs. For NiftyReg, we use three scales, the same local negative cross correlation objective function with control point grid spacing of five voxels and 500 iterations.

4.2. Registration Performance Comparisons

To construct a registration test dataset which can be used to quantitatively compare different registrations, we randomly generate two registration fields for each test image in the test dataset. We use five evaluation metrics in the registration, average time in seconds for inference on one pair of images, average registration field endpoint error (EPE) which is the same as Eq. 4, average mean square error over

reconstructed image and fixed image (MSE), average normalized local cross-correlation with neighbor volume size $5 \times 5 \times 5$ the same as negative value in Eq. 6 (NLCC), and average mutual information with 100 bins (MI). The comprehensive comparisons over ANTs [4], NiftyReg [29] and VoxelMorph [6] are listed in the table 1 and 2 for the Hippocampus and Prostate datasets respectively.

From the table 1 and 2, our registration achieves substantial improvement on EPE which is one of the most accurate ways to evaluate the performance of registration in equation 4. Our registration obtains the best performance on the MSE which is a robust metric for data with the same distribution in the current scenario. The NeurReg yields the best MI on the two dataset and the best NLCC on the Hippocampus dataset. The hybrid loss obtains comparable EPE metric and does much better on the other metrics compared with model without similarity loss. The multi-task learning-based registration obtains the comparable registration performance as the base methods. The improvement over the three mostly used registration toolboxes on the two datasets with the five metrics confirms the robustness and good performance of the proposed NeurReg. The improvement probably is because the registration field guided learning in NeurReg leads to an optimal convergent point in the learning.

We further qualitatively compare the four registration methods by visualizing the registration field, reconstructed image and difference image between reconstructed image and fixed image in Fig. 3 and 4. We randomly choose two test cases from the two datasets respectively. For the difference image visualization, we multiply six on the pixel value to increase the intensity for visual purpose.

From the Fig. 3 and 4, NeurReg performs much better than ANTs, NiftyReg and VoxelMorph on the registration field estimation. The difference images of neural registration are smooth without sharp large error points and the overall error is smaller. The ANTs, NiftyReg and VoxelMorph directly optimize the similarity loss which can be considered as the difference image. We introduce the registration field guided supervision and the model converges a better solution for both registration field estimation and appearance reconstruction.

We visualize the field training loss of VoxelMorph and NeurReg to further analysis the learning/optimization on the Hippocampus dataset in Fig. 5. We can find the registration field supervision loss is easier to optimize than bending energy used in VoxelMorph and other registrations.

4.3. Segmentation Performance Comparisons

We use Dice coefficient (Dice = $\frac{2TP}{2TP+FN+FP}$) as the evaluation metric, where TP, FN, and FP are true positives, false negatives, false positives, respectively. For segmentation based on VoxelMorph and NeurReg, we use 10% training images based on NLCC of reconstructed image as atlases and conduct majority voting in the multi-atlas based segmentation. Because of heavy computational cost of ANTs and NiftyReg, we only use the top one training image based on NLCC in the atlas set from NeurReg as atlas. The multi-task learning from Eq. 10 is denoted by VoxelMorph (MTL) and Ours (MTL). The residual segmentation block in Fig. 2 is denoted by VoxelMorph (Feat.) and Ours (Feat.). We also compare with an advanced segmentation

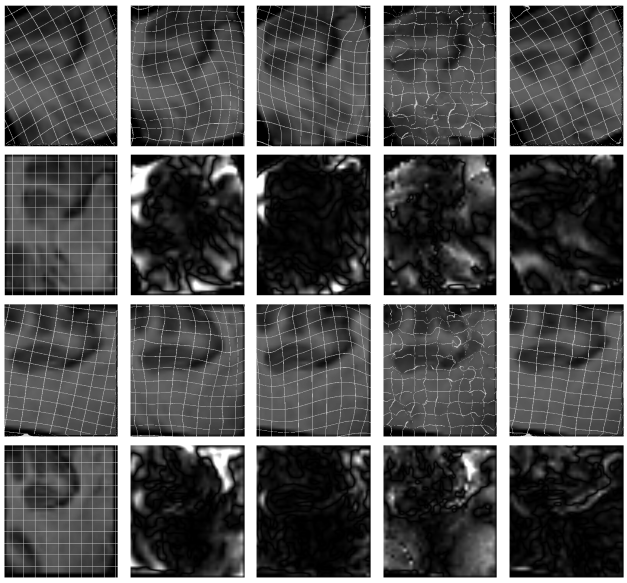

Figure 3. Visualization of registration results from ANTs, NiftyReg, VoxelMorph and ours on the Hippocampus test dataset. The images in the first column are original fixed/moving images with mesh denoting the registration field ground truth. The images from the second column to the last column are reconstructed images with estimated registration field and difference images from ANTs, NiftyReg, VoxelMorph and ours respectively. The first two rows are from axial direction and the last two rows are from sagittal direction. NeurReg obtains the best registration field estimation and smooth difference images with the least error.

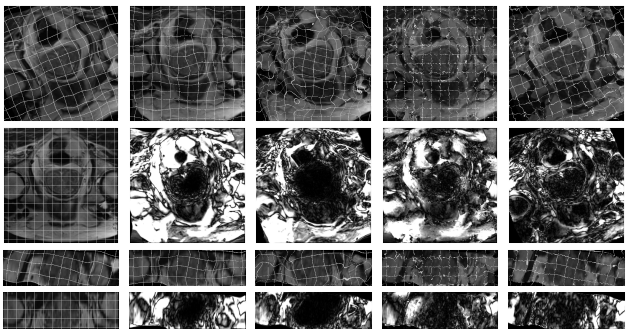

Figure 4. Visualization of registration results from ANTs, NiftyReg, VoxelMorph and ours on the Prostate test datasets. NeurReg obtains the best registration field and difference images.

Method	Time (s)	EPE (mm)	MSE	NLCC	MI
ANTs	289.05±135.71	4.267±1.809	0.008±0.006	0.624±0.115	0.934± 0.276
NiftyReg	991.48±420.90	4.113±1.450	0.002±0.002	0.795±0.058	1.484± 0.195
VoxelMorph	0.03± 0.18	6.222± 1.573	0.004± 0.002	0.723± 0.047	1.006± 0.134
VoxelMorph (MTL)	0.03±0.14	6.236±1.579	0.007±0.004	0.598±0.050	0.775±0.120
VoxelMorph (Feat.)	0.04±0.16	6.242±1.573	0.005±0.003	0.687±0.050	0.911±0.133
Ours (NeurReg)	0.03±0.15	0.957 ± 0.312	$\mathbf{0.001 \pm 0.153}$	$\mathbf{0.808 \pm 0.069}$	$\mathbf{1.520 \pm 0.191}$
Ours w/o $\mathcal{L}_{sim}$	0.03±0.18	$\mathbf{0.950 \pm 0.330}$	0.002 ± 0.002	0.762±0.087	1.386 ± 0.219
Ours (MTL)	0.06±0.19	1.277±0.382	0.002±0.001	0.749±0.070	1.337±0.164
Ours (Feat.)	0.05±0.20	1.146±0.339	0.002±0.001	0.782±0.068	1.410±0.176

Table 1. Registration comparisons on the Hippocampus dataset. NeurReg is the best based on all the metrics. Best scores are in bold face.

Method	Time (s)	EPE (mm)	MSE	NLCC	MI
ANTs	5851.84±2450.31	24.371±8.535	0.021±0.005	0.304±0.074	0.334±0.173
NiftyReg	2307.32±662.08	25.556±7.595	0.008±0.003	0.407±0.096	0.821±0.242
VoxelMorph	0.92±1.71	26.483±7.450	0.010±0.001	$\mathbf{0.504 \pm 0.033}$	0.472±0.104
VoxelMorph (MTL)	1.07±1.60	26.480±7.432	0.012±0.002	0.423±0.029	0.376±0.089
VoxelMorph (Feat.)	1.20±1.72	26.489±7.441	0.013±0.002	0.433±0.027	0.356±0.084
Ours (NeurReg)	0.78±1.34	5.228 ± 1.169	$\mathbf{0.006 \pm 0.002}$	0.363±0.057	$\mathbf{0.860 \pm 0.231}$
Ours w/o $\mathcal{L}_{sim}$	0.76±1.33	$\mathbf{5.082 \pm 1.173}$	0.008 ± 0.002	0.280±0.052	0.735±0.204
Ours (MTL)	1.90±3.12	6.991±1.452	0.008±0.002	0.292±0.039	0.700±0.198
Ours (Feat.)	1.70±2.15	6.532±1.687	0.008±0.003	0.321±0.045	0.741±0.215

Table 2. Registration comparisons on the Prostate dataset.

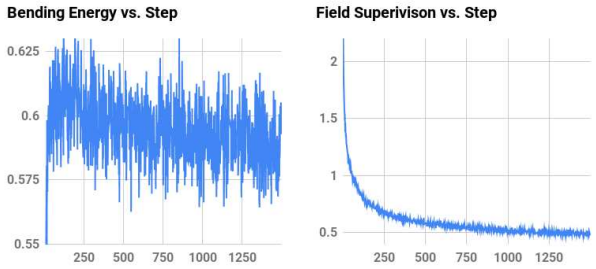

Figure 5. Visualization of bending energy used in VoxelMorph (left) and registration field supervision loss (right) in our registration on the Hippocampus dataset.

model which uses 3D U-Net[1] based on 18 residual blocks in the encoder and ImageNet pretrained Res-18 weight [25]. The comparison results are listed in table 3 and 4.

Table 3 demonstrates better accuracy of our registration-based segmentations compared to VoxelMorph on all the three frameworks. The NeurReg with residual segmentation block obtains comparable performance as the advanced 3D U-Net with more parameters and pretrained model [25]. The atlas-based segmentation with fast inference speed is comparable with NiftyReg and ANTs and it confirms the robustness of proposed registration. From table 4, NeurReg

[1]https://developer.nvidia.com/clara

Method	Dice (%)
ANTs	80.86±5.13/78.34±5.24
NiftyReg	80.53±4.86/77.92±5.47
VoxelMorph	78.92±6.79/75.85±8.13
VoxelMorph (MTL)	86.69±5.04/85.72±3.97
VoxelMorph (Feat.)	86.68±4.24/85.68±4.46
3D U-Net	88.66±3.09/86.76±3.26
Ours	78.99±6.24/78.72±7.47
Ours (MTL)	88.95±3.66/87.10±3.72
Ours (Feat.)	$\mathbf{89.18 \pm 3.50/87.39 \pm 3.42}$

Table 3. Segmentation comparisons on the Hippocampus dataset. Ours is the best and comparable with an advanced 3D UNet.

Method	Dice (%)
ANTs	28.52±18.94/57.58±24.09
NiftyReg	27.88±17.25/56.67±23.71
VoxelMorph	22.03±12.33/53.40±21.50
VoxelMorph (MTL)	25.50±14.63/63.62±18.73
VoxelMorph (Feat.)	38.97±16.61/76.72±5.62
Ours	21.70±11.95/55.80±17.53
Ours (MTL)	31.57±22.49/73.84±12.68
Ours (Feat.)	$\mathbf{44.30 \pm 17.60/82.38 \pm 3.46}$

Table 4. Segmentation comparisons on the Prostate dataset.

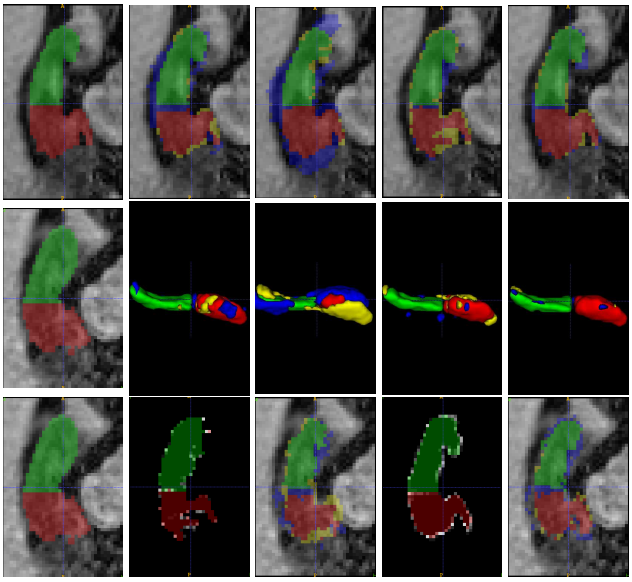

Figure 6. Visualization of segmentation results from ANTs, NiftyReg, VoxelMorph and Ours (Feat.) on the Hippocampus dataset. The images in the first column are test image and moving images from VoxelMorph and NeurReg. The columns from second row to the last row are segmentation results from ANTs, NiftyReg, VoxelMorph and ours. The images in the last row are uncertainty map and MTL prediction backpropagated map to the moving image from VoxelMorph and Ours. The green and red region represents ground truth or true positive. Blue is the false positive region and yellow is the false negative region.

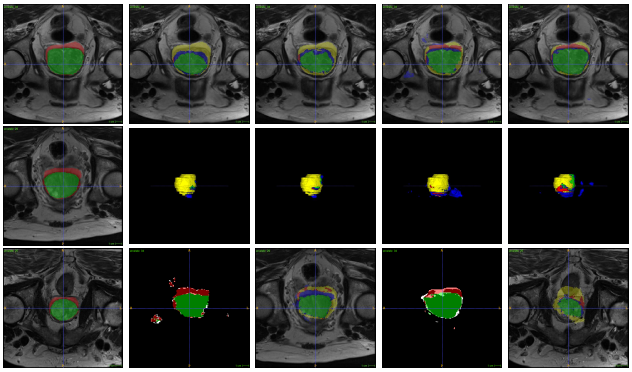

Figure 7. Visualization of segmentation results from ANTs, NiftyReg, VoxelMorph and Ours (Feat.) on the Prostate dataset.

with residual segmentation block achieves the best performance which might be because of the dual registration and residual segmentation block.

We visualize the segmentation from the four registration methods in Fig. 6 and 7. The figures demonstrate NeurReg achieves the best segmentation on the randomly chosen two test images from the two datasets.

4.3.1 Uncertainty Estimation and Interpretability for Segmentation

Uncertainty estimation is one of the most important features in medical image analysis to assist radiologists. For multi-atlas based segmentation, each atlas can be considered as a prior. We can simply derive the segmentation uncertainty by

$$\boldsymbol{U} = 1 - \frac{\sum_{i=1}^{N_{atlas}} \mathbb{I}(\boldsymbol{S}_i = \boldsymbol{S})}{N_{atlas}}, \tag{12}$$

based on empirical estimation, where $\boldsymbol{S}$ is the prediction after majority voting and $\boldsymbol{S}_i$ is the prediction from i_{th} atlas, $\mathbb{I}$ is an indicator image. If the pixel value in the uncertainty map is large, it means the prediction in the current pixel has a high uncertainty. From the uncertainty maps in last row of Fig. 6 and 7, the high uncertainty regions are along the edges of the predictions which are error prone areas.

The other advantage of registration-based segmentation is the registration field provides the interpretability of segmentation prediction. From the registration field, we can build the connection between the MTL prediction and ground truth of the moving image. We use ITK to approximately calculate the inverse registration field in the last row of Fig. 6 and 7. Through the prediction backpropagated map based on top one image of NLCC, we can find the reason of the prediction even based on appearance in the image space.

5. Conclusion

In this work, we developed a registration simulator to synthesize images under various plausible transformations. Then we design a hybrid loss between registration field supervision loss and data similarity loss in NeurReg. The registration field supervision provides an accurate field loss and is easy to optimize. The data similarity loss improves the model generalization ability. We further extend the registration framework to multi-task learning with segmentation and propose a dual registration to fully exploit the generalization of representational similarity loss on random fixed images. A residual segmentation block is designed to further boost the segmentation performance. Extensive experimental results demonstrate our NeurReg yields the best registration on several metrics and best segmentation with uncertainty and interpretability on the two public datasets. In future works, it is promising to generalize NeurReg and explore the applicability to multi-modal image registration.

References

[1] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007.

[2] J. Ashburner and K. J. Friston. Voxel-based morphometrythe methods. *Neuroimage*, 11(6):805–821, 2000.

[3] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.

[4] B. B. Avants, N. Tustison, and G. Song. Advanced normalization tools (ants). *Insight j*, 2:1–35, 2009.

[5] R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Computer vision, graphics, and image processing*, 46(1):1–21, 1989.

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9252–9260, 2018.

[7] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 2019.

[8] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.

[9] M. Blendowski and M. P. Heinrich. Combining mrf-based deformable registration and deep binary 3d-cnn descriptors for large lung motion estimation in copd patients. *International journal of computer assisted radiology and surgery*, 14(1):43–52, 2019.

[10] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen. Deformable image registration based on similarity-steered cnn regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 300–308. Springer, 2017.

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

[12] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

[13] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer, 2018.

[14] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum. 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*, pages 204–212. Springer, 2017.

[15] J. Fan, X. Cao, Z. Xue, P.-T. Yap, and D. Shen. Adversarial similarity network for evaluating image alignment in deep learning based registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 739–746. Springer, 2018.

[16] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[17] G. Haskins, U. Kruger, and P. Yan. Deep learning in medical image registration: A survey. *arXiv preprint arXiv:1903.02026*, 2019.

[18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[19] M. Jaderberg, K. Simonyan, A. Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.

[20] P. Jiang and J. A. Shackleford. Cnn driven sparse multi-level b-spline image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9281–9289, 2018.

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

[22] J. Krebs, H. e Delingette, B. Mailhé, N. Ayache, and T. Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging*, 2019.

[23] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen. Robust non-rigid registration through agent-based action learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 344–352. Springer, 2017.

[24] R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D. Comaniciu. An artificial agent for robust image registration. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[25] S. Liu, D. Xu, S. K. Zhou, O. Pauly, S. Grbic, T. Mertelmeier, J. Wicklein, A. Jerebko, W. Cai, and D. Comaniciu. 3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 851–858. Springer, 2018.

[26] K. Ma, J. Wang, V. Singh, B. Tamersoy, Y.-J. Chang, A. Wimmer, and T. Chen. Multimodal image registration with deep context reinforcement learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 240–248. Springer, 2017.

[27] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3, 2013.

[28] S. Miao, S. Piat, P. Fischer, A. Tuysuzoglu, P. Mewes, T. Mansi, and R. Liao. Dilated fcn for multi-agent 2d/3d medical image registration. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[29] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–284, 2010.

[30] J. Neylon, Y. Min, D. A. Low, and A. Santhanam. A neural network approach for fast, automated quantification of dir performance. *Medical physics*, 44(8):4126–4138, 2017.

[31] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec. Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 266–274. Springer, 2017.

[32] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[33] S. Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

[34] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

[35] S. S. M. Salehi, D. Erdogmus, and A. Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International Workshop on Machine Learning in Medical Imaging*, pages 379–387. Springer, 2017.

[36] D. Shen and C. Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE transactions on medical imaging*, 21(11):1421, 2002.

[37] P. Y. Simard, D. Steinkraus, J. C. Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *Icdar*, volume 3, 2003.

[38] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.

[39] H. Sokooti, B. de Vos, F. Berendsen, B. P. Lelieveldt, I. Išgum, and M. Staring. Nonrigid image registration using multi-scale 3d convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–239. Springer, 2017.

[40] J.-P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical image analysis*, 2(3):243–260, 1998.

[41] H. Uzunova, M. Wilms, H. Handels, and J. Ehrhardt. Training cnns for image registration from few samples with model-based data augmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 223–231. Springer, 2017.

[42] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen. Unsupervised deep feature learning for deformable registration of mr brain images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 649–656. Springer, 2013.

[43] X. Yang, R. Kwitt, M. Styner, and M. Niethammer. Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396, 2017.

[44] W. Zhu, Y. Huang, M. A. Vannan, S. Liu, D. Xu, W. Fan, Z. Qian, and X. Xie. Neural multi-scale self-supervised registration for echocardiogram dense tracking. *arXiv preprint arXiv:1906.07357*, 2019.

[45] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie. Anatomynet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics*, 46(2):576–589, 2019.

[46] W. Zhu, C. Lan, J. Xing, W. Zeng, Y. Li, L. Shen, and X. Xie. Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[47] Y. Zhu, Z. Lan, S. Newsam, and A. G. Hauptmann. Guided optical flow learning. *arXiv preprint arXiv:1702.02295*, 2017.