



U-RSNet: An unsupervised probabilistic model for joint registration and segmentation

Liang Qiu, Hongliang Ren *

Department of Biomedical Engineering, National University of Singapore, Singapore

ARTICLE INFO

Article history:

Received 22 October 2020

Revised 20 February 2021

Accepted 12 April 2021

Available online 17 April 2021

Communicated by Zidong Wang

Keywords:

Deformable registration

Segmentation

Deep learning

Joint learning framework

ABSTRACT

Medical image segmentation and registration have vital roles in computer-assisted diagnosis procedures, challenging tasks suffering from various limitations and artifacts inside images. Recently, deep learning techniques accomplish these two tasks and achieve outstanding performances. However, most deep learning-based methods overlook the potential correlation between each other. In this paper, an unsupervised probabilistic model named U-RSNet is proposed to realize concurrent medical image registration and segmentation in one framework. Specifically, the unsupervised segmentation branch is derived from Bayesian inference. The prior warped atlas for segmentation can be obtained by deforming a known probabilistic atlas by the corresponding invertible deformation field with a well-behaved diffeomorphic guarantee, which can perfectly integrate these two tasks to form a complete intelligent prediction system. In this case, the segmentation performance could be largely improved based on the warped probabilistic atlas obtained from the registration branch. Experiments on human brain 3D magnetic resonance images have demonstrated the effectiveness of our approach. We trained and validated U-RSNet with 1000 images and tested its performances on four public datasets. We showed our method successfully realized concurrent segmentation and registration and yielded better segmentation results than a separately trained network.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Artificial intelligence (AI) has been successfully applied in many healthcare areas, such as robotic surgery, virtual nursing assistance, clinical judgment or diagnosis, and medical image analysis. Medical image registration and segmentation as two significant tasks in medical research and clinical application have led to quite a lot of AI-powered tools and solutions. Medical image segmentation is the process to extract the target region automatically or semi-automatically within different imaging modalities, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound Imaging (US), Cone-beam CT (CBCT), and X-ray. Segmentation can assist in border detection, visual augmentation, tumor detection, and some other computer-aided diagnosis based on the extracted boundaries. However, it still suffers from the limitations and artifacts of medical imaging, such as noise and intensity inhomogeneity across various modalities. Traditional

segmentation methods are usually computationally intensive and time-consuming, not suitable for current automatic clinical diagnosis and intelligent healthcare. Fortunately, deep learning techniques have achieved outstanding performances in segmentation tasks, which are quite fast and accurate.

Similarly, deformable registration is also a vital component of medical image analysis procedures, aiming to transform different medical images into one coordinate system and establish anatomical correspondence regardless of imaging modalities, viewpoints, and time. It has been successfully used in tumor growth process monitoring, atlas-based segmentation, and organ template creation. Deformable registration approaches usually include two stages: an initial low-dimensional rigid or affine transformation, then followed by a much more complex deformable transformation that can capture the subtle and localized deformation. Traditional deformable registration methods have obvious drawbacks, such as intensive computation and local minimum problems. Simultaneously, deep learning techniques that excel at learning complex contextual features can achieve fast performances and provide comparable accuracy and robustness.

However, most existing methods only conduct registration and segmentation separately and neglect the potential connection between these two tasks. Besides, extensive and high-quality

* Corresponding author at: Department of Electronic Engineering and Shun Hing Institute of Advanced Engineering, The Chinese University of Hong Kong (CUHK), Hong Kong, Department of Biomedical Engineering, National University of Singapore, Singapore.

E-mail address: hlen@ieee.org (H. Ren).

annotated datasets are usually required to train data-hungry deep learning models challenging and intractable when facing a large amount of raw data. In this paper, to cope with the problems mentioned above, a joint learning framework based on probabilistic inference, called U-RSNet, is proposed for concurrent medical image registration and segmentation, unsupervised and adaptive to various unlabeled medical image data. Our contributions are summarized as follows:

- 1) A joint learning framework is proposed by incorporating a diffeomorphic registration network and a Bayesian segmentation network. By the invertible deformation field, the two tasks can be seamlessly integrated into one system.
- 2) Our proposed method is unsupervised for both registration and segmentation without the need for any known transformations and manually segmented images for training, which is efficient and adaptive to various brain MRI datasets.
- 3) Our proposed end-to-end framework can realize concurrent registration and segmentation within 14 s, suitable for clinics' time-critical tasks.
- 4) To demonstrate the effectiveness and generalizability of our U-RSNet, we test it on four widely-used public brain MRI datasets, which presents largely improved segmentation performances compared with a separately-trained network in terms of Dice scores.

2. Related work

2.1. Learning-based medical image segmentation

Medical image segmentation constitutes a primary step in automatic analysis and diagnosis, such as cell counting, tumorous tissue identification, and growth prediction [1,2]. However, the high variability of human anatomy and various medical imaging modalities have posed significant challenges to guarantee the segmentation performance. Besides, artifacts and noises that lead to medical imaging uncertainty also make the segmentation process much more complicated. Recently, a surge of deep learning-based methods that exploit convolutional neural network (CNN) for medical image segmentation has won tremendous success in both accuracy and real-time performance. In terms of the image dimensionality, the deep learning approaches could fall into 2D CNN-based methods and 3D CNN-based methods.

To process volumetric images, 2D CNN-based methods perform the convolution operations slice-by-slice. Then all the segmentation slices are combined into a segmentation volume. One of the most representative networks is called U-Net [3], whose distinctive characteristic is the skip connection, which can provide the deconvolution layers with significant corresponding high-resolution features to enhance supervision in different feature levels compared with traditional coder-decoder structures. Several modified variations have coped with varying tasks of segmentation for outstanding performance. For instance, an attention U-Net with a new attention gate for medical images could learn how to focus on target anatomical structures [4].

Moreover, to further enhance the connection and reduce the semantic gap between encoder and decoder sub-networks, A nested U-Net called U-Net++ has been proposed. The skip connection at each level has been re-designed by using a series of nested dense convolutional blocks. Experiments with different datasets such as liver and chest CT scans demonstrated U-Net++ with deep supervision achieved much better performance against U-Net [5].

Because of the 2D convolution operated on each slice of volumetric images, the spatial information embedded among slices has been ignored, which will affect the segmentation performance. The most obvious solution is to leverage 3D CNN-based

approaches. One representative is called 3D U-Net [6], which extends the original U-Net into a 3D version by converting all the 2D operations into 3D counterparts and has achieved great success. However, all the methods mentioned above rely on manually annotated images and are susceptible to intensity variance. To overcome this problem, a Bayesian brain MRI segmentation method was presented by exploiting a probabilistic atlas with deformable registration and obtaining outstanding segmentation accuracy of brain Magnetic Resonance (MR) images [7].

2.2. Learning-based medical image registration

Classical registration algorithms [8–10] can provide impressive performances with rigorous mathematical derivation to solve an optimization problem, but intensive computation and slow run-time have extremely hindered their development and application. In contrast, deep learning techniques are well-suited for medical image registration in computational efficiency and accuracy. Recently, different learning-based methods for medical image registration have been proposed for various medical applications. For example, to estimate the brain MR images' deformation field, a well-designed encoder-decoder network named Quicksilver was proposed [11]. Furthermore, a multi-scale 3D CNN named RegNet was designed to predict 3D chest CT data's deformation with artificially generated displacement vector field (DVF) [12].

Meanwhile, contextual information is considered to boost the feature representation by exploiting multi-scaled MRI contents. Though the aforementioned deep learning-based registration methods have presented competitive registration performances, they are all supervised methods. To be more specific, ground truth warp fields should be provided to guarantee practical training, which is expensive and impractical to acquire, particularly for medical images.

Given rarely available large quantities of ground truth warp fields and their potential problems, unsupervised learning methods have been adopted for medical image registration, which is efficient and adaptive to various raw datasets and has attracted the medical imaging community's attention. For example, a self-supervised, fully convolutional network (FCN) was built to learn the deformable registration field of inter-subject 3D brain MRIs with spatial transformation function and multi-resolution strategy [13]. Recently, anatomical segmentations have been incorporated into the objective function to learn the 3D voxel correspondence, demonstrated to have an accurate registration performance [14]. In general, although all the unsupervised methods mentioned above formulate the image registration as optimization problems based on different similarity metrics and regularization terms, they cannot guarantee the diffeomorphism. To alleviate this problem, an unsupervised inference algorithm was derived based on a probabilistic generative model using diffeomorphic deformation representation [15].

2.3. Joint optimization of segmentation and registration

Medical image segmentation and registration should not be treated as totally isolated tasks, and they can be highly correlated by exploiting the auxiliary anatomical information and the invertible transformation. Specifically, segmentation can be accomplished by registering an image to be segmented to an atlas image to find the corresponding labels, which can be called atlas-based segmentation [16]. In addition, reference information from the adjacent slices in a volume instead of atlases can also be exploited via registration techniques for medical image segmentation. For example, a registration-based geodesic active contour model was proposed to segment the outer wall of abdominal aortic aneurysm from 3D black-blood MRI, which is challenging due to the low intensity contrast with respect to neighboring tissues [17]. An auxiliary registration optimization term measuring the correlation between the current and

reference contours is introduced to improve the segmentation performance, where the manual initial delineations should be carefully determined to avoid accumulating error in the following iterative optimization process. Although it could provide effective segmentation performance, the relatively time-consuming computing process could be a problem in clinical emergencies or time-critical medical applications. Likewise, registration can be improved by incorporating the auxiliary segmentation information into the training procedure, which can be called weakly-supervised registration and demonstrated to be effective [14,18].

Contrary to the above single-task methods, alternative or joint optimization of both registration and segmentation has been investigated, which can simultaneously complete those two tasks. Specifically, alternative registration and segmentation methods conduct one-step registration followed by one-step segmentation, which assists the counterpart in an iterative interleaving way. For example, a generative probabilistic framework was proposed to combine registration, classification, and bias correction in a circular procedure by establishing a maximum a posteriori model [19]. Alternative estimation of model parameters offers better results than the simple serial implementation of each component. Differently, joint registration and segmentation approaches update the parameters simultaneously in each iteration. For instance, a joint registration and segmentation model derived from optical flow and active contour theory were proposed [20]. It successfully combines the non-rigid registration method's strength with an active contour framework embedded with local segmentation constraints, which can theoretically register any anatomical contours.

Nevertheless, obvious drawbacks for those traditional alternatives or joint optimization approaches are intensive computation and slow runtime for each image pair, intractable for some time-critical medical applications such as emergency diagnosis and treatment. To overcome the limitations, deep learning-based joint optimization methods have been developed. For example, by coupling registration and segmentation, a multi-task deep neural network named U-ReSNet was proposed, which led to promising results on published OASIS dataset [21]. Similarly, both tasks' hybrid deep learning network was well-designed, which was evaluated on longitudinal white matter tract changes and achieved consistent improvements [22]. Moreover, a joint semi-supervised learning framework named DeepAtlas was proposed by optimizing the segmentation and registration branch network alternatively, which provides a general solution to cope with joint training when many images are with a few manual segmentations [23]. In addition, a learning-based adversarial optimization approach for joint registration and segmentation has been proposed, which was demonstrated effective in prostate CT radiotherapy. It leverages a generator network to estimate the deformation field among images. These two sub-networks form an end-to-end closed loop system to improve the performances of both tasks [24]. However, the aforementioned joint-task methods cannot guarantee the diffeomorphism of the registration field, which may exert a negative effect on both tasks. Furthermore, all these methods are supervised; that is to say, large and high-quality segmentation annotations are acquired to train the networks, impractical, and troublesome when dealing with a large amount of raw data. Therefore, in this paper, we propose a novel unsupervised joint learning framework for concurrent registration and segmentation.

3. Materials and methods

3.1. System overview

In this section, our unsupervised joint learning framework for both registration and segmentation will be described generally.

Let m and f represent the moving 3D image and the fixed 3D image, respectively. In addition, a probabilistic atlas \mathbf{A} corresponding to the fixed image f is provided to realize unsupervised learning with the aid of an invertible deformation field. We assume all the images are preprocessed to be affinely aligned, with only non-linear misalignment left.

The overview of our approach is illustrated in Fig. 1. Our proposed neural network's whole architecture comprises two branches, which correspond to the registration sub-network and the segmentation sub-network, respectively. For image registration, we mainly adopt the probabilistic model to guarantee the diffeomorphic characteristic [15]. RegNet $\mathcal{R}_\theta(m, f)$ from the registration sub-network takes a moving 3D image m and a fixed 3D image f as inputs and outputs a probabilistic velocity field \mathbf{v} sampled from corresponding mean $\mu_{v|m, f}$ and variance $\sum_{v|m, f}$. By using squaring and scaling integration layers, the velocity field \mathbf{v} is transformed into a diffeomorphic deformation field ϕ_v , which can warp m to obtain $m \circ \phi_v$ to optimally fit f by the spatial transformation network STN [25]. In addition, our segmentation branch is designed based on Bayesian inference. Because of the invertible deformation field with a well-behaved diffeomorphic guarantee, the negative velocity field $-\mathbf{v}$ can be integrated to compute the inverse ϕ_v^{-1} efficiently, which can be applied to the probabilistic atlas \mathbf{A} to obtain warped probabilistic atlas $\mathbf{A}(l, \phi_v^{-1})$, which is treated as prior information for segmentation of the image m . Both moving 3D image m and $\mathbf{A}(l, \phi_v^{-1})$ are input into SegNet $\mathcal{S}_\theta(m)$ followed by a global max-pooling layer to output the Gaussian parameters which are used to model likelihood $p(\mathbf{D}|\mathbf{S}, \xi_D)$. Based on the prior $\mathbf{A}(l, \phi_v^{-1})$ and the likelihood $p(\mathbf{D}|\mathbf{S}, \xi_D)$, the maximum a posteriori (MAP) segmentation can be predicted effectively.

3.2. Diffeomorphic registration

This section describes the particular architecture for the diffeomorphic registration network we used in our framework. We adopted the same architecture as described in the probabilistic diffeomorphic model VoxelMorph-diff [15] to implement the registration of input image pair (m, f) , which can provide a differentiable, invertible deformation field for the image deformable transformation. The CNN $\mathcal{R}_\theta(m, f)$ parameterized by θ to estimate $\mu_{v|m, f}$ and $\sum_{v|m, f}$ can be trained to learn parameters θ by the approximate posterior inference:

$$L(\theta|f, m) = -E_q[\log p(f|\mathbf{v}, m)] + \text{KL}[q_\theta(\mathbf{v}|m, f) \| p(\mathbf{v})] \quad (1)$$

whose rigorous theoretical derivation can be easily found in [15]. After obtaining $\mu_{v|m, f}$ and $\sum_{v|m, f}$, the stationary velocity field \mathbf{v} could be generated by resampling from a normal distribution $\mathbf{v} \sim \mathcal{N}(\mu_{v|m, f}, \sum_{v|m, f})$, which can be further processed into a deformation field ϕ_v .

The particular architecture of RegNet $\mathcal{R}_\theta(m, f)$ is similar to 3D U-Net [6], including both encoder, decoder and direct skip connection. Besides, related attributes have been adjusted to fit our application, such as the number of the convolution kernels and the architecture's depth, as shown in Fig. 2. It takes the 3D images m and f as input and outputs voxel-wise mean $\mu_{v|m, f}$ and variance $\sum_{v|m, f}$ of the stationary velocity field. It should be noted that $\mathcal{R}_\theta(m, f)$ is general CNN block built with different architectures to realize similar feature extraction and provide ideal output, but the specific design is not our focus in the paper.

In the encoder stage, 3D convolutions are applied to obtain increasingly abstract features of the original input images as the process goes deeper. The kernel size and the stride in each convolution are set to 3 and 2, respectively, followed by a LeakyReLU. The

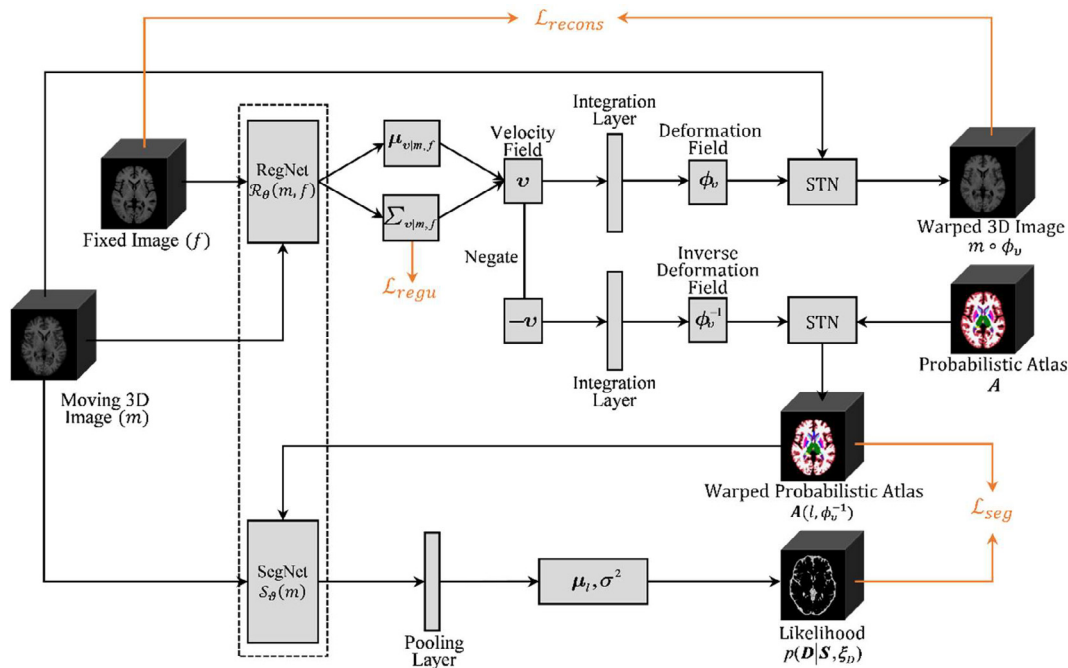


Fig. 1. Overview of U-RSNet, which is a joint learning framework for both image registration and segmentation. The architecture is composed of two branches, the registration sub-network, and segmentation sub-network. $\mathcal{R}_v(m, f)$ takes a moving 3D image m and a fixed 3D image f as inputs and outputs a probabilistic velocity field v sampled from corresponding mean and variance. By using squaring and scaling integration layers, the velocity field v is transformed into a diffeomorphic deformation field ϕ_v , which can warp m to obtain $m \circ \phi_v$ by spatial transformation network STN. Meanwhile, the negative velocity field $-v$ can be integrated to compute the inverse ϕ_v^{-1} easily, which can be applied to the probabilistic atlas A to obtain a warped probabilistic atlas $A(l, \phi_v^{-1})$ for MAP segmentation inference.

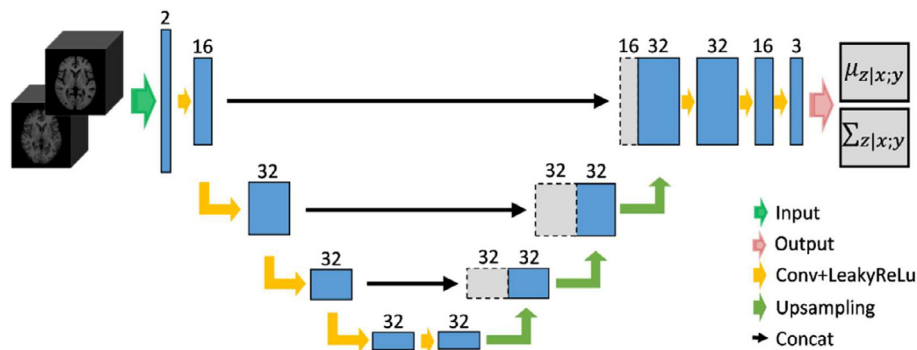


Fig. 2. The architecture of diffeomorphic registration 3D U-Net.

3.3. 3D image segmentation

The segmentation sub-network is modeled based on Bayesian inference, which is specially designed by using the warped probabilistic atlas as the segmentation prior obtained from the registration sub-network. In this section, we take the brain MR image as an example to show the detailed derivation process. Let \mathbf{D} denotes the intensity of the 3D human brain MR images, and let \mathbf{S} represents the corresponding segmentation with L anatomical labels. Then the MAP estimation of the segmentation can be formulated as follow:

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmax}} p(\mathbf{S}|\mathbf{D}) = \underset{\mathbf{S}}{\operatorname{argmax}} p(\mathbf{D}|\mathbf{S})p(\mathbf{S}) \quad (2)$$

where the prior $p(\mathbf{S})$ presents the probabilistic spatial distribution of segmentation labels parameterized by ξ_s , while the likelihood $p(\mathbf{D}|\mathbf{S})$ describes the potential relationship between the MR image intensity and segmentation labels parameterized by ξ_D .

To solve the intractability of integrals over ξ_S and ξ_D , the point estimates are exploited as follow:

$$\begin{aligned} \{\hat{\xi}_S, \hat{\xi}_D\} &= \underset{\{\xi_S, \xi_D\}}{\operatorname{argmax}} p(\xi_S, \xi_D | \mathbf{D}) \\ &= \underset{\{\xi_S, \xi_D\}}{\operatorname{argmax}} p(\xi_S) p(\xi_D) \sum_{\mathbf{S}} p(\mathbf{D} | \mathbf{S}, \xi_D) p(\mathbf{S} | \xi_S) \end{aligned} \quad (3)$$

Considering point estimates shown in (3), the MAP segmentation can be obtained efficiently:

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmax}} p(\mathbf{S} | \hat{\xi}_S, \hat{\xi}_D, \mathbf{D}) = \underset{\mathbf{S}}{\operatorname{argmax}} p(\mathbf{D} | \mathbf{S}, \hat{\xi}_D) p(\mathbf{S} | \hat{\xi}_S) \quad (4)$$

In our framework, the initial prior segmentation information is defined based on a given probabilistic atlas \mathbf{G} where each voxel is associated with a prior probability vector corresponding to observed segmentation labels at the specific location. It could be acquired by averaging a group of manually annotated brain MR Images with various modalities. Then the initial atlas \mathbf{G} should be transformed into probabilistic atlas \mathbf{A} corresponding to the fixed image f by deformable registration field ϕ_{AG} , to guarantee the consistency of registration and segmentation tasks, shown as follows:

$$\hat{G}(x) = \underset{l}{\operatorname{argmax}} G(l, x) \quad (5)$$

$$\phi_{AG}^* = \underset{\phi_{AG}}{\operatorname{argmin}} [\mathcal{L}_{\text{sim}}(f_{\text{seg}}, \hat{G} \circ \phi_{AG}) + \lambda_{\text{reg}} \mathcal{L}_{\text{smooth}}(\phi_{AG})] \quad (6)$$

$$A(l, x) = G(l, x) \circ \phi_{AG}^*(x) \quad (7)$$

where $G(l, x)$, $A(l, x)$ means the probability of each anatomical label $l = 1, \dots, L$ at each location $x \in \Omega$ (Ω represents the location space in \mathbf{G} and \mathbf{A}), f_{seg} is the segmentation of fixed image f , $\mathcal{L}_{\text{sim}}(\cdot, \cdot)$ denotes the similarity loss, $\mathcal{L}_{\text{smooth}}(\cdot)$ is the regularization loss and λ_{reg} is the trade-off hyper-parameter. Specifically, Eq. (6) represents a typical deformable registration optimization based on energy function considering both image similarity and smoothness regularization.

Then the prior for a specific segmentation task can be obtained by assuming independence over voxels:

$$p(\mathbf{S}|\xi_S, \mathbf{A}) = \prod_{j \in \Omega} A(S_j, \phi_v^{-1}(x_j)) \quad (8)$$

where the distribution of ξ_S can be defined as:

$$p(\xi_S) \propto \exp(-\kappa \|\nabla \mathbf{u}\|) \quad (9)$$

which is used to guarantee the smoothness of the deformation field by penalizing the spatial gradient $\nabla \mathbf{u}$. However, our registration branch has already guaranteed the diffeomorphic characteristic. Hence, Eq. (9) could be neglected in our final loss function, as shown in the next section.

Moreover, the likelihood for observed intensities at each voxel location are assumed to obey independent Gaussian distributions:

$$p(\mathbf{D}|\mathbf{S}, \xi_D) = \prod_{j \in \Omega} \mathcal{N}(D_j; \eta_{S_j}, \sigma_{S_j}^2) \quad (10)$$

where $\mathcal{N}(\cdot; \eta, \sigma)$ denotes the standard Gaussian distribution, D_j is the intensity at location j , and a flat prior is applied $p(\xi_D) \propto 1$.

Then the final predicted segmentation in each independent voxel can be obtained as follows:

$$\hat{S}_j = \underset{l}{\operatorname{argmax}} \mathcal{N}(D_j; \eta_l, \sigma_l^2) A(l, \phi_v^{-1}(x_j)) \quad (11)$$

SegNet $S_\theta(m)$ in the framework is a convolutional network for segmentation prediction. In our application, we adopt a 3D U-Net style architecture like $\mathcal{R}_\theta(m, f)$. Instead of recovering the dimension to half of the input size, we further increase the spatial dimension to the full size with the number of the feature channels equaling to that of segmentation labels. Then an additional convolutional layer followed by a pooling layer will take the combination of the previous output and the warped probabilistic atlas $\mathbf{A}(l, \phi_v^{-1})$ as input, and outputs the corresponding Gaussian parameters which are used to model likelihood $p(\mathbf{D}|\mathbf{S}, \xi_D)$.

3.4. Loss function and implementation details

This section proposes a loss function for the joint learning framework U-RSNet, which consists of two components: \mathcal{L}_{reg} that indicates the loss of registration sub-network, \mathcal{L}_{seg} which denotes the loss of segmentation sub-network.

The registration loss \mathcal{L}_{reg} includes two parts according to the probabilistic model VoxelMorph-diff [15] shown in formula (1): reconstruction loss $\mathcal{L}_{\text{recon}}$ that encourages image $m \circ \phi_v$ and image f as similar as possible, and regulation loss $\mathcal{L}_{\text{regu}}$ that encourages the posterior $q_\theta(v|m, f)$ to be close to the prior $p(v)$, shown as follows:

$$\mathcal{L}_{\text{recon}} = \frac{1}{2\sigma^2 K} \sum_k \|f - m \circ \phi_{v_k}\|^2 \quad (12)$$

$$\mathcal{L}_{\text{regu}} = [\mu_{v|m,f}^T \Gamma_v \mu_{v|m,f} + \text{tr}(\lambda \mathbf{D} \sum_{v|m,f} -\log \sum_{v|m,f})]/2 \quad (13)$$

where K is the sampling number, Γ and \mathbf{D} are the precision matrix and the graph degree matrix respectively, and λ is the parameter to control the scale of v .

In addition, the segmentation loss \mathcal{L}_{seg} can be acquired based on the mathematical derivation shown in the above section, as follow:

$$\mathcal{L}_{\text{seg}} = - \sum_{j \in \Omega} \log \sum_{l=1}^L \mathcal{N}(D_j; \eta_l, \sigma_l^2) A(l, \phi_v^{-1}(x_j)) \quad (14)$$

Then we can combine all the loss items to obtain the objective loss function:

$$\mathcal{L}_{\text{RSNet}} = \alpha \mathcal{L}_{\text{reg}} + \mathcal{L}_{\text{seg}} = \alpha (\mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{regu}}) + \mathcal{L}_{\text{seg}} \quad (15)$$

where α represents the hyper-parameter.

4. Results

4.1. Medical image data acquisition and preprocessing

In this paper, experiments on human brain 3D MR images have been performed to demonstrate the effectiveness of our approach. To train our unsupervised U-RSNet, a large number of brain images are needed. In order to represent the generalization of the overall population, a large-scale multi-site dataset constructed from eight public datasets has been applied to demonstrate the effectiveness of our method. To be more specific, we randomly collect 1000 brain MRI images with various age ranges and health conditions from 4 public datasets: OASIS [26], ADNI [27], ABIDE [28] and ADHD200 [29], where 800 images for training and 200 images for validation, then we test our approach on another four different datasets: CUMC12 [30], IBSR18 [30], MGH10 [30], and LPBA40 [31]. In order to alleviate the possible biases stemming from image acquisition procedures, heterogeneous demographics and other factors, all the raw data mentioned above need to go through a preprocessing pipeline to generate suitable images to feed into our U-RSNet, which is a necessary preliminary step to improve data harmonization and the reliability of our learned model. Specifically, skull stripping will firstly be applied to all the datasets to get corresponding linearly resampled brain-masks ($256 \times 256 \times 256$ resolution) with 1 mm isotropic voxels using FreeSurfer [32], followed by image normalization, image cropping, and manually refinement, which result in well-defined images with $160 \times 192 \times 224$ resolution. Moreover, all the brain images are segmented into 28 specific anatomical structures with integer labels by using nearest neighbor resampling, as shown in Table 1. Furthermore, all the pre-processed images are aligned with a brain atlas image with affine transformation by using Advanced Normalization Tools (ANTs) [33], which is open-source software for efficient medical image registration. Last but not least, the probabilistic atlas \mathbf{A} used in our experiment is the deformed version of a public available atlas \mathbf{G} [34], which is registered to the fixed image using ANTs as well. In the way, registration and segmentation tasks can be effectively integrated into one framework.

Table 1
Regions of interest (Rols) in our dataset.

ID	ROI name	ID	ROI name
1	Left Cerebral White Matter	15	Right Cerebral White Matter
2	Left Cerebral Cortex	16	Right Cerebral Cortex
3	Left Lateral Ventricle	17	Right Lateral Ventricle
4	Left Cerebellum white Matter	18	Right Cerebellum white Matter
5	Left Cerebellum Cortex	19	Right Cerebellum Cortex
6	Left Thalamus Proper	20	Right Thalamus Proper
7	Left Caudate	21	Right Caudate
8	Left Putamen	22	Right Putamen
9	Left Pallidum	23	Right Pallidum
10	Left Hippocampus	24	Right Hippocampus
11	Left Amygdala	25	Right Amygdala
12	Left VentralDC	26	Right VentralDC
13	Left Choroid Plexus	27	Right Choroid Plexus
14	4rd Ventricle	28	Brain Stem

4.2. Baseline methods

In order to test and analyze the segmentation performance of our joint learning framework U-RSNet, the reimplementation of an affine version of the atlas-based segmentation [35] using deep learning method is exploited as the baseline to analyze the impact deriving from the registration branch. This baseline, denoted as seg-aff-atlas, also leverages formula (14) to conduct segmentation but without deformable transformation, namely $\phi_v^{-1} = Id$. We evaluate the performances of these segmentation methods by computing the Dice metric, which measures the volume overlap of the corresponding anatomical structures shown in Table 1. Moreover, we also employ the Euclidean surface distance (ESD) computed using the strategy described in [15] to further evaluate the segmentation performances, which is more sensitive to the small segmentation error and less biased by the region sizes of the anatomical structures. Specifically, we implement the ESD metric by making use of the surface distance transform of a specific image, which is a derived representation of the image where each voxel has a value equal to its distance to an ROI. In addition, 10,000 points along each anatomical surface structure are sampled, which we find to be sufficient to estimate accurate measures.

Similarly, to assess the registration performance, a comparison to other well-performed methods is provided as well. The baseline methods we select here are two top-performing traditional methods: Symmetric Normalization (SyN) and NiftyReg (CC), and one deep learning-based method: VoxelMorph-diff [15] corresponding to the registration sub-network in U-RSNet. SyN algorithm integrated in ANTs package is implemented with the step size of 0.25, the Gaussian smoothing of (9, 0.2), and 201 iterations for three scales. NiftyReg is also an efficient open-source registration tool with the parameter setting as follows: grid spacing of 5, maximum iteration number of 1000, and a standard deviation of the Gaussian kernel of 5. Additionally, the parameter selection for VoxelMorph-diff is as follows: learning rate of 1×10^{-4} , regularization parameter of 1, batch size of 1, step per epoch of 100 and the number of epochs of 1000. All the registration methods are evalu-

ated using the Dice metric and the Jacobian determinant $\det(J_\phi) = \det(\nabla \phi)$ where non-positive values mean folding voxels.

4.3. Implementation and comparison

Our proposed U-RSNet is implemented based on Keras with a Tensorflow backend, and it is implemented on Nvidia Titan RTX GPU with 24 GB memory. The batch size and step of the epoch are set to 1 and 100, respectively. Besides, The ADAM optimizer is used in our application with a learning rate of 1×10^{-4} . Considering the limited memory of the GPU, each training batch is composed of one pair of medical image volumes. Before we start the joint training of U-RSNet, we can pre-train the registration sub-network separately for a sufficient number of epochs. Here we set the number of pre-training epochs to 600 and the number of joint training epochs to 400. Specifically, the probabilistic diffeomorphic registration framework VoxelMorph-diff is applied as the registration sub-network in our model, and we also adopt the same parameter setting in our experiments, where λ shown in formula (13) is set to 25 with best performance [15]. Moreover, the hyper-parameter α in the proposed loss function formula (15) needs to be carefully tuned during the training and the validation steps. Table 8 presents the average Dice scores for the validation dataset with different values of α . The results with comparable good performances vary smoothly over a large range $\alpha \in [1, 100]$ values, indicating that the proposed model is robust to the selection of α . For very small values of α (e.g. 0.01), both registration and segmentation performances are decreased, probably due to less attention is paid to the registration part, which leads to poor segmentation prior. For very large α values (e.g. 10000), the training process becomes unstable and converges much slower. In our experiments, the loss weight is set to $\alpha = 10$ based on the hyper-parameter analysis and fine-tuning. Furthermore, to obtain the transformation ϕ_{AG}^* between the initial atlas G and the warped atlas A according to formula (6) which is a general energy optimization expression, we exploited an specific iteratively algorithm, namely SyN algorithm integrated in ANTs package, whose implementation is with the step size of 0.25, the Gaussian smoothing of (9, 0.2), and 201 iterations for three scales.

To evaluate the registration and the segmentation performances of U-RSNet, we test it on four public datasets and compare the results with the corresponding baseline methods mentioned above, respectively.

As shown in Table 2, four different deformable registration algorithms are compared regarding the Dice score, computational time, and the number of voxels with non-positive Jacobian determinant. In addition, the affine registration is also listed as a reference. We find that the deep learning-based method VoxelMorph-diff achieves better registration accuracy than traditional methods, and the diffeomorphic constraint leads to a smooth and reasonable deformation field. Furthermore, U-RSNet achieves comparable performance against VoxelMorph-diff with a slight improvement (0.5% = 0.768–0.763). Besides, both VoxelMorph-diff and U-RSNet require less than 2 s, running on GPU, which is over 100 times faster than traditional ANTs(SyN) and NiftyReg running on CPU.

Table 2
Summary of registration results on total dataset (mean/std. dev.).

Method	Averaged dice	GPU sec	CPU sec	$\det(J_\phi) < 0$
Affine	0.667 (0.173)	~	~	~
ANTs(SyN)	0.750 (0.141)	~	2143 (371)	9765 (2821)
NiftyReg	0.758 (0.141)	~	214 (54)	66,895 (913)
VoxelMorph-diff	0.763 (0.125)	0.74 (0.007)	~	3.3 (4.5)
U-RSNet	0.768 (0.124)	1.60 (0.005)	~	4.7 (5.7)

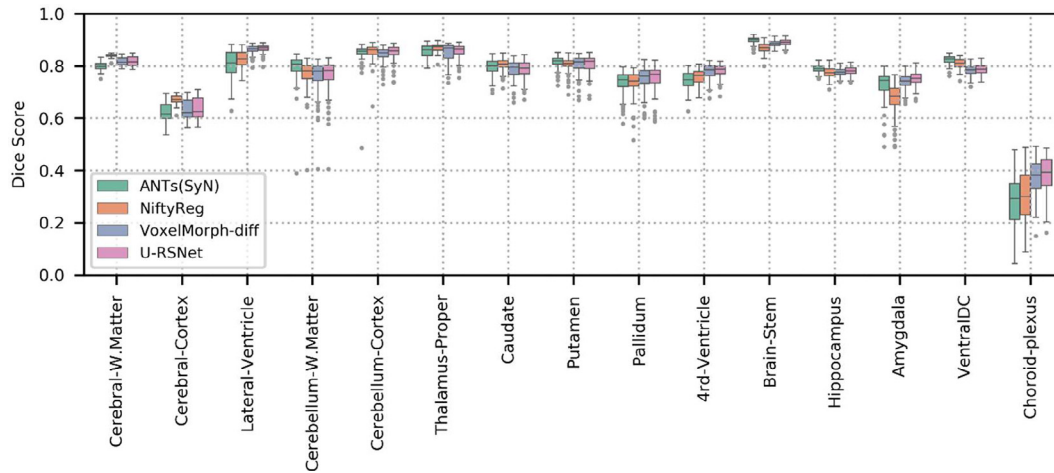


Fig. 3. Boxplot of Dice scores of different anatomical structures for 4 registration methods: ANTs(SyN), NiftyReg, VoxelMorph-diff and U-RSNet. Dice scores of the corresponding anatomical structures from the left and the right brain hemispheres are averaged into one score.

Table 3

Summary of segmentation dice results on total dataset (mean/std. dev.).

Method	Cerebral-w.matter	Cerebral-cortex	Lateral-ventricle	Cerebellum-w.matter	Cerebellum-cortex
seg-aff-atlas	0.839 (0.023)	0.727 (0.036)	0.657 (0.136)	0.794 (0.059)	0.845 (0.034)
U-RSNet	0.857 (0.029)	0.690 (0.089)	0.802 (0.050)	0.781 (0.075)	0.860 (0.050)
Method	Thalamus-Proper	Caudate	Putamen	Pallidum	4th-Ventricle
seg-aff-atlas	0.831 (0.032)	0.732 (0.045)	0.790 (0.043)	0.734 (0.054)	0.715 (0.074)
U-RSNet	0.884 (0.023)	0.762 (0.085)	0.802 (0.053)	0.766 (0.058)	0.777 (0.055)
Method	Brain-Stem	Hippocampus	Amygdala	VentralDC	Choroid-plexus
seg-aff-atlas	0.873 (0.029)	0.749 (0.048)	0.741 (0.049)	0.775 (0.052)	0.155 (0.122)
U-RSNet	0.897 (0.016)	0.816 (0.025)	0.774 (0.033)	0.816 (0.018)	0.448 (0.069)

To present the performances of different registration methods, the averaged Dice scores of the corresponding anatomical structures from both the left and the right brain hemispheres are illustrated in Fig. 3.

Furthermore, the segmentation performances on the total dataset in terms of the Dice score are presented in Table 3. We can find that our proposed U-RSNet almost achieves consistent improvements with the higher mean Dice scores compared with the seg-aff-atlas method except for cerebral-cortex and cerebellum-w.matter, which demonstrates our approach is much more accurate with the help from the diffeomorphic registration branch. Moreover, outstanding improvements on some specific ROIs in terms of Dice score can be observed, such as lateral ventricle (+14.5%), thalamus proper (+5.3%), 4th-ventricle (+6.2%), hippocampus (+6.7%) and choroid-plexus (+29.3%). Moreover, we could find that those particular ROIs with high Dice scores obtained by U-RSNet are deep brain structures with relatively regular shapes, while the performances on cerebral-cortex (−3.7%) and cerebellum-w.matter (−1.3%) are decreased to some extent. We further checked the segmentation performances on cerebral-cortex and cerebellum-w.matter for each subject image, and found that there are no consistent evidences to show U-RSNet is totally inferior to seg-aff-atlas. There could be two potential reasons resulting in such experimental result. Firstly, the prior label distribution information described by our probabilistic atlas \mathbf{G} computed from a dataset of 20 T1-weighted MRI scans with manual parcellations [34], may not sufficiently reflect the real probability distribution of cerebral-cortex and cerebellum-w.matter due to their structural complexity and variability, compared with other organs with regular shapes and locations. Secondly, in order to realize joint learning in our framework, we try to obtain the probabilistic atlas \mathbf{A}

from \mathbf{G} to match the fixed image f as much as possible. However, the probabilistic atlas \mathbf{A} and the fixed image f could be built from different datasets, and they have completely different properties as well, namely, the former representing label probabilities while the latter representing a template with an average intensity. Therefore, there is always a gap between these two components, which could lead to additional uncertainty or inaccuracy when warping the probabilistic atlas \mathbf{A} with the inverse deformation field ϕ_o^{-1} to provide the segmentation prior, especially on those anatomical sites with irregular and complex shapes. Likewise, for better illustration, a boxplot of the segmentation results is displayed in Fig. 4, which can provide a straightforward comparison between these two algorithms. On top of those results, we further evaluated the segmentation performance with the ESD metric, which may be less biased by the anatomical region size and structure. The related surface results for each ROI on the total dataset are presented in Table 4, which can demonstrate that our U-RSNet almost achieves significantly smaller surface distances (lower is better), particular on lateral ventricle and choroid-plexus, except for cerebral-cortex and cerebellum-w.matter. This experimental findings in terms of ESD is consistent to those in terms of Dice metric, which further demonstrates the superiority of U-RSNet against seg-aff-atlas. In addition, U-RSNet can generate the segmentation output less than 14 s, which is satisfactory for time-critical applications.

To further demonstrate the generalizability of U-RSNet, we test its performance on each dataset, respectively, as shown in Tables 5–7. We observe that our U-RSNet performs slightly better than VoxelMorph-diff (VM-diff) on all the sub-datasets with an overall increase of 0.5% (=0.768–0.763) given all the ROIs on total dataset. More importantly, U-RSNet provides a consistent increase of segmentation Dice scores on all the sub-datasets, and have 5.1%

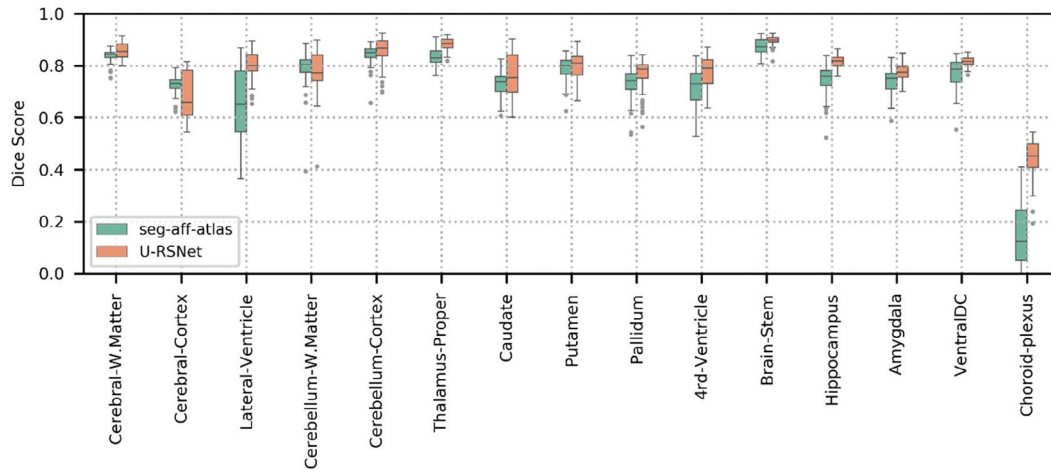


Fig. 4. Boxplot of Dice scores of various anatomical structures for 2 segmentation methods: seg-aff-atlas and U-RSNet. Dice scores of the corresponding anatomical structures from the left and the right brain hemispheres are averaged into one score.

Table 4

Summary of segmentation ESD results on total dataset (mean/std. dev.).

Method	Cerebral-w.matter	Cerebral-cortex	Lateral-ventricle	Cerebellum-w.matter	Cerebellum-Cortex
seg-aff-atlas	1.587 (0.121)	1.625 (0.113)	1.948 (0.484)	1.858 (0.465)	1.833 (0.266)
U-RSNet	1.536 (0.212)	1.869 (0.292)	1.298 (0.148)	1.966 (0.539)	1.699 (0.368)
Method	Thalamus-Proper	Caudate	Putamen	Pallidum	4th-Ventricle
seg-aff-atlas	1.675 (0.265)	1.532 (0.180)	1.569 (0.222)	1.691 (0.271)	1.674 (0.496)
U-RSNet	1.346 (0.155)	1.440 (0.282)	1.531 (0.217)	1.548 (0.244)	1.451 (0.511)
Method	Brain-Stem	Hippocampus	Amygdala	VentralDC	Choroid-plexus
seg-aff-atlas	1.558 (0.239)	1.581 (0.207)	1.527 (0.225)	1.512 (0.227)	3.151 (0.169)
U-RSNet	1.433 (0.129)	1.321 (0.098)	1.436 (0.195)	1.353 (0.104)	1.609 (0.573)

Table 5

Registration results on 4 public datasets (mean/std. dev.).

Method	CUMC12	IBSR18	LPBA40	MGH10
	Dice $\det(J_\phi) < 0$	Dice $\det(J_\phi) < 0$	Dice $\det(J_\phi) < 0$	Dice $\det(J_\phi) < 0$
VM-diff	0.785 (0.119) 3.0 (2.8)	0.763 (0.116) 1.4 (2.7)	0.762 (0.132) 4.6 (5.5)	0.744 (0.116) 1.6 (2.0)
U-RSNet	0.791 (0.119) 3.0 (3.6)	0.770 (0.114) 2.8 (4.8)	0.765 (0.131) 6.9 (6.2)	0.749 (0.114) 1.5 (2.9)

(=0.782–0.731) significant improvement on total dataset. In addition, we can also find that the average ESD values of U-RSNet on each individual dataset are consistent smaller than those of seg-aff-atlas, which further validates the effectiveness and robustness of U-RSNet. Furthermore, some examples of different registration and segmentation methods are illustrated in Fig. 5 for a more intuitive comparison.

Given all the experimental result analysis, it can be demonstrated that U-RSNet can realize concurrent registration and segmentation in an unsupervised way, and achieve comparable (registration) or largely improved (segmentation) performance over the separately-learned sub-task networks with the state-of-the-art accuracy and generalizability, which fully reveals the feasibility and the effectiveness of our proposed framework. In addition, we can further interpret the experimental results through more specific analyses to understand the relationship between these two sub-tasks. We can find that U-RSNet obtains comparable registration performances with a slight improvement in contrast to VM-diff, which may involve two possible reasons. On one hand, VM-diff is an outstanding diffeomorphic registration model which could provide relatively accurate registration results as well as differentiable deformation fields, and it is well integrated into our model as a registration sub-network. However, the diffeomorphic

characteristic may potentially hinder further registration improvement in terms of Dice scores compared with other registration models, due to its strong constrain on allowable transformations to ensure desirable topology-preserving properties. On the other hand, the probabilistic atlas could be easily warped by the inverse deformation field deriving from the registration sub-network to provide the prior information for the segmentation task, which is the key component of our joint learning model to connect these two sub-tasks. Particularly, the obtained warped probabilistic atlas, as the segmentation prior probability, can be directly utilized in the derivation and optimization procedure, leading to significant improvement of segmentation performance effectively. However, this segmentation optimization procedure could only indirectly and limitedly affect the registration sub-network through its sub-branch for inverse transformation, resulting in very slight improvements as shown in our results.

5. Discussion and conclusions

We proposed an unsupervised probabilistic model for joint registration and segmentation, named U-RSNet, which establishes a principled and practical connection between these two tasks and

Table 6

Segmentation dice results on 4 public datasets (mean/std. dev.).

Method	CUMC12	IBSR18	LPBA40	MGH10	Total Sets
seg-aff-atlas	0.744 (0.172)	0.736 (0.169)	0.719 (0.184)	0.753 (0.153)	0.731 (0.176)
U-RSNet	0.813 (0.106)	0.799 (0.111)	0.761 (0.119)	0.798 (0.110)	0.782 (0.116)

Table 7

Segmentation ESD results on 4 public datasets (mean/std. dev.).

Method	CUMC12	IBSR18	LPBA40	MGH10	Total Sets
seg-aff-atlas	1.680 (0.527)	1.711 (0.487)	1.834 (0.774)	1.626 (0.382)	1.755 (0.642)
U-RSNet	1.360 (0.206)	1.464 (0.379)	1.603 (0.382)	1.507 (0.311)	1.522 (0.363)

Table 8Dice score of validation data with different hyper-parameter α .

Setting	0.01	1	10	20	100	10,000
Reg. Dice	0.664 (0.166)	0.758 (0.125)	0.760 (0.124)	0.761 (0.126)	0.759 (0.122)	0.752 (0.125)
Seg. Dice	0.689 (0.196)	0.780 (0.118)	0.784 (0.114)	0.783 (0.116)	0.780 (0.114)	0.775 (0.121)

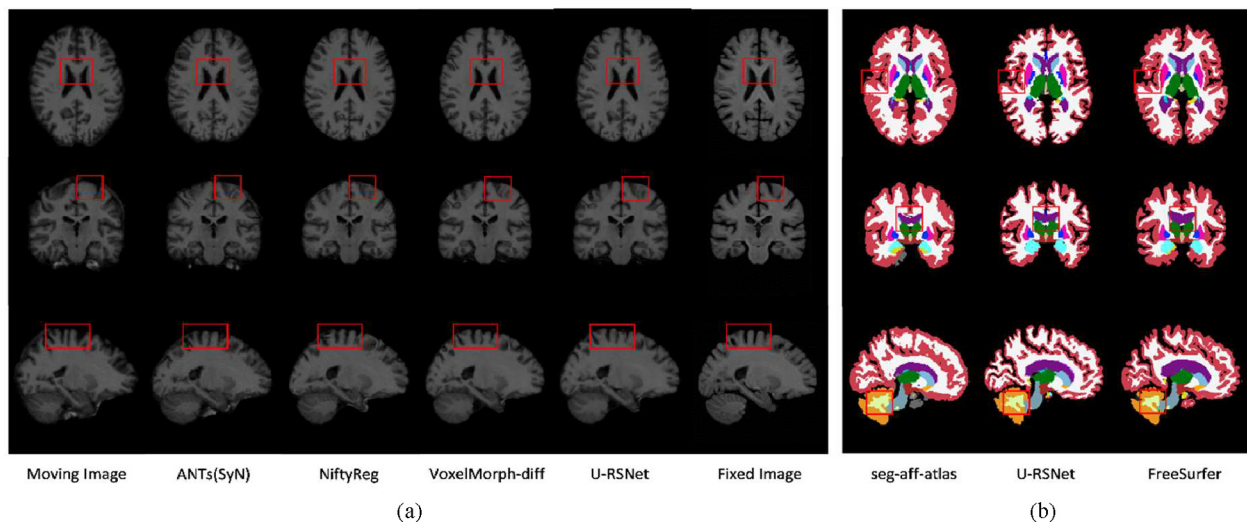


Fig. 5. (a) Registration examples on the IBSR18 dataset with 4 different methods: ANTs(SyN), NiftyReg, VoxelMorph-diff and U-RSNet, which register the moving image (Column 1) to the fixed image (Column 6). (b) Segmentation examples on IBSR18 dataset with 3 different methods: seg-aff-atlas, U-RSNet, and FreeSurfer.

shows significant improvement on segmentation. Besides, our proposed method is unsupervised for both registration and segmentation without the need for any known transformations and manually segmented images, which has been demonstrated to be efficient and adaptive to various raw brain MRI datasets.

Besides, our framework can generate the results for both registration and segmentation simultaneously within several seconds in one framework, which is particularly suitable for time-critical clinical applications. To demonstrate our proposed method's feasibility and effectiveness, experiments on four public datasets yield concurrent improvements for the registration and the segmentation compared with independently trained networks.

Notably, multi-site datasets with sufficient medical images are collected in this paper for the training, validation and testing of our model, which is essential for the learning-based large population analysis, similar to the implementations mentioned in [15,18]. Although pooling data from different sites could enhance the data generalization, naive combination of datasets is probably ill-advised due to introducing various biases in training data stemming from different demographics, image acquisition and other factors [36]. Specifically, demographics bias refers to that the sub-

ject selection cannot represent the overall or target population, and image acquisition bias mainly depends on the scanning effect and the acquisition protocol such as magnetic field strength, manufacturer, pulse sequences, etc. All these biases may lead to testing performance drop indicating poor generalization, which is also called potential data shift problem between training and testing datasets. In order to relieve data shift and improve data harmonization, necessary preliminary steps such as atlas-based normalization, quality control and outlier removal should be implemented to improve the model reliability based on multi-site studies [37], similar to our preprocessing pipeline introduced in Section 4.1. However, it also emerges that substantial biases in multi-site datasets remain despite preprocessing techniques are exploited to enhance data consistency across scanners. Moreover, preprocessing such as registration and segmentation with specific tools (e.g. FreeSurfer used in our experiments) may also introduce extra biases, related to the accuracy of the tools and certain potential image attributes. Therefore, particular attention should be paid to potential data shift when dealing with multi-site medical data, which remains an open challenge in both academic research and clinic applications [38]. Recently, domain adaptation and transfer learning methods, as

powerful tools, are proposed to cope with data shift problems, which could effectively improve the prediction accuracy on the target domain [39,40]. In the future, we will also try to incorporate these advanced techniques into our framework and study their impacts on the performances of both registration and segmentation tasks. Furthermore, although FreeSurfer is widely exploited to produce automated brain segmentation as the silver standard ground truth, making broad range of experiments doable, potential biases could also be introduced to affect the model accuracy. Thus we will train and test our model with manual segmentation treated as the golden standard in the next step to further analyze the performance of our model and assess the corresponding differences.

To sum up, our joint learning neural network U-RSNet, is a general unsupervised framework with state-of-the-art performances on brain MR images, which could also be widely applied in other image modalities or anatomies for both registration and segmentation. Therefore, we expect our multi-task model could be an automatic clinical tool to speed up the medical analysis procedure and improve diagnostic accuracy.

CRedit authorship contribution statement

Liang Qiu: Conceptualization, Methodology, Software, Data curation, Writing - original draft. **Hongliang Ren:** Conceptualization, Methodology, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Shun Hing Institute of Advanced Engineering (SHIAE project #BME-p1-21) at the Chinese University of Hong Kong (CUHK), and Singapore Academic Research Fund under Grant R397000353114 and R397000350118.

References

- [1] B. Ibragimov, D.A. Toesca, D.T. Chang, Y. Yuan, A.C. Koong, L. Xing, Automated hepatobiliary toxicity prediction after liver stereotactic body radiation therapy with deep learning-based portal vein segmentation, *Neurocomputing* 392 (2020) 181–188.
- [2] M. Baldeon-Calisto, S.K. Lai-Yuen, AdaResU-Net: multiobjective adaptive convolutional neural network for medical image segmentation, *Neurocomputing* 392 (2020) 325–340.
- [3] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, 2015, pp. 234–241.
- [4] O. Oktay, et al., Attention u-net: learning where to look for the pancreas, *arXiv preprint arXiv:1804.03999*, 2018.
- [5] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: a nested u-net architecture for medical image segmentation, in: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 3–11.
- [6] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, 2016, pp. 424–432.
- [7] A.V. Dalca, E. Yu, P. Golland, B. Fischl, M.R. Sabuncu, J.E. Iglesias, Unsupervised deep learning for Bayesian brain MRI segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 356–365.
- [8] J. Ashburner, A fast diffeomorphic image registration algorithm, *Neuroimage* 38 (1) (2007) 95–113.
- [9] B.B. Avants, C.L. Epstein, M. Grossman, J.C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, *Med. Image Anal.* 12 (1) (2008) 26–41.
- [10] V. Vishnevskiy, T. Gass, G. Szekely, C. Tanner, O. Goksel, Isotropic total variation regularization of displacements in parametric image registration, *IEEE Trans. Med. Imag.* 36 (2) (2017) 385–395.
- [11] X. Yang, R. Kwitt, M. Styner, M. Niethammer, Quicksilver: Fast predictive image registration—a deep learning approach, *NeuroImage* 158 (2017) 378–396.
- [12] H. Sokooti, B. de Vos, F. Berendsen, B.P. Lelieveldt, I. Išgum, M. Staring, Nonrigid image registration using multi-scale 3D convolutional neural networks, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 232–239.
- [13] H. Li, Y. Fan, Non-rigid image registration using self-supervised fully convolutional networks without training data, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, IEEE, pp. 1075–1078.
- [14] Y. Hu et al., Weakly-supervised convolutional neural networks for multimodal image registration, *Med. Image Anal.* 49 (2018) 1–13.
- [15] A.V. Dalca, G. Balakrishnan, J. Guttag, M.R. Sabuncu, Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces, *Med. Image Anal.* 57 (2019) 226–236.
- [16] J.M. Lötjönen et al., Fast and robust multi-atlas segmentation of brain magnetic resonance images, *Neuroimage* 49 (3) (2010) 2352–2365.
- [17] Y. Wang et al., Segmentation of lumen and outer wall of abdominal aortic aneurysms from 3D black-blood MRI with a registration based geodesic active contour model, *Med. Image Anal.* 40 (2017) 1–10.
- [18] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, VoxelMorph: a learning framework for deformable medical image registration, *IEEE Trans. Med. Imag.* (2019).
- [19] J. Ashburner, K.J. Friston, Unified segmentation, *Neuroimage* 26 (3) (2005) 839–851.
- [20] V. Duay, X. Bresson, J.S. Castro, C. Pollo, M.B. Cuadra, J.-P. Thiran, An active contour-based atlas registration model applied to automatic subthalamic nucleus targeting on MRI: method and validation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2008, pp. 980–988.
- [21] T. Estienne et al., U-ReSNet: ultimate coupling of registration and segmentation with deep nets, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 310–319.
- [22] B. Li et al., A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 645–653.
- [23] Z. Xu, M. Niethammer, DeepAtlas: joint semi-supervised learning of image registration and segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 420–429.
- [24] M.S. Elmahdy, J.M. Wolterink, H. Sokooti, I. Išgum, M. Staring, Adversarial optimization for joint registration and segmentation in prostate CT radiotherapy, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 366–374.
- [25] M. Jaderberg, K. Simonyan, A. Zisserman, Spatial transformer networks, in: *Advances in Neural Information Processing Systems*, 2015, pp. 2017–2025.
- [26] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, R.L. Buckner, Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults, *J. Cogn. Neurosci.* 19 (9) (2007) 1498–1507.
- [27] S.G. Mueller et al., Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI), *Alzheimer's Dementia* 1 (1) (2005) 55–66.
- [28] A. Di Martino et al., The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism, *Mol. Psychiatry* 19 (6) (2014) 659.
- [29] M.P. Milham, D. Fair, M. Mennes, S.H. Mostofsky, The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience, *Front. Syst. Neurosci.* 6 (2012) 62.
- [30] A. Klein et al., Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration, *Neuroimage* 46 (3) (2009) 786–802.
- [31] D.W. Shattuck et al., Construction of a 3D probabilistic atlas of human cortical structures, *Neuroimage* 39 (3) (2008) 1064–1080.
- [32] B. Fischl, FreeSurfer, *Neuroimage* 62 (2) (2012) 774–781.
- [33] B.B. Avants, N.J. Tustison, G. Song, P.A. Cook, A. Klein, J.C. Gee, A reproducible evaluation of ANTs similarity metric performance in brain image registration, *Neuroimage* 54 (3) (2011) 2033–2044.
- [34] O. Puonti, J.E. Iglesias, K. Van Leemput, Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling, *NeuroImage* 143 (2016) 235–249.
- [35] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based tissue classification of MR images of the brain, *IEEE Trans. Med. Imag.* 18 (10) (1999) 897–908.
- [36] C. Wachinger, B.G. Becker, A. Rieckmann, S. Pölsterl, Quantifying confounding bias in neuroimaging datasets with causal inference, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 484–492.
- [37] E. Ferrari, P. Bosco, G. Spera, M.E. Fantacci, A. Retico, Common pitfalls in machine learning applications to multi-center data: tests on the ABIDE i and ABIDE ii collections, in: *Joint Annual Meeting ISMRM-ESMRMB*, 2018.

- [38] B. Glocker, R. Robinson, D.C. Castro, Q. Dou, E. Konukoglu, Machine learning with multi-site imaging data: an empirical study on the impact of scanner effects, arXiv preprint arXiv:1910.04597, 2019.
- [39] Y.-W. Luo, C.-X. Ren, D. Dao-Qing, H. Yan, *Unsupervised domain adaptation via discriminative manifold propagation*, *IEEE Trans. Pattern Anal. Mach. Intell.* (2020).
- [40] C.-X. Ren, J. Feng, D.-Q. Dai, S. Yan, *Heterogeneous domain adaptation via covariance structured feature translators*, *IEEE Trans. Cybern.* (2019).



Liang Qiu received the M.S. degree in control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2016. He is currently pursuing the Ph.D degree with the Department of Biomedical Engineering, National University of Singapore (NUS), Singapore. His research interests include medical robotics, vision-based surgical navigation, deformable registration and deep learning.



Hongliang Ren received his Ph.D. in Electronic Engineering (Specialized in Biomedical Engineering) from The Chinese University of Hong Kong (CUHK) in 2008. He serves as an Associate Editor for IEEE Transactions on Automation Science & Engineering (T-ASE) and Medical & Biological Engineering & Computing (MBEC). He has navigated his academic journey through Chinese University of Hong Kong, Johns Hopkins University, Children's Hospital Boston, Harvard Medical School, Children's National Medical Center, United States, and National University of Singapore. He is currently Associate Professor, Department of Electronic Engineering at Chinese University of Hong Kong, and Adjunct Associate Professor, Department of Biomedical Engineering at National University of Singapore. His areas of interest include biorobotics, intelligent control, medical mechatronics, soft continuum robots, soft sensors, and multisensory learning in medical robotics. He is the recipient of NUS Young Investigator Award and Engineering Young Researcher Award, IAMBE Early Career Award 2018, Interstellar Early Career Investigator Award 2018, and ICBHI Young Investigator Award 2019.