Biomedical Signal Processing and Control 80 (2023) 104294

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

# Joint few-shot registration and segmentation self-training of 3D medical images

Huabang Shi [a], Liyun Lu [a], Mengxiao Yin [a,b], Cheng Zhong [a,b], Feng Yang [a,b,*]

[a] School of Computer, Electronics and Information, Guangxi University, Nanning, Guangxi 530004, China
[b] Guangxi Key Laboratory of Multimedia Communications Network Technology, Guangxi University, Nanning, Guangxi 530004, China

ARTICLE INFO

*Keywords:*
Image registration
Image segmentation
Few-shot
Self-training
Pseudo-labels

ABSTRACT

Medical image segmentation and registration are very important related steps in clinical medical diagnosis. In the past few years, deep learning techniques for joint segmentation and registration have achieved good results in both segmentation and registration tasks through one-way assisted learning or mutual utilization. However, they often rely on large labeled datasets for supervised training or directly use pseudo-labels without quality estimation. We propose a joint registration and segmentation self-training framework (JRSS), which aims to use segmentation pseudo-labels to promote shared learning between segmentation and registration in scenarios with few manually labeled samples while improving the performance of dual tasks. JRSS combines weakly supervised registration and semi-supervised segmentation learning in a self-training framework. Segmentation self-training generates high-quality pseudo-labels for unlabeled data by injecting noise, pseudo-labels screening, and uncertainty correction. Registration utilizes pseudo-labels to facilitate weakly supervised learning, and as input noise as well as data augmentation to facilitate segmentation self-training. Experiments on two public 3D medical image datasets, abdominal CT and brain MRI, demonstrate that our proposed method achieves simultaneous improvements in segmentation and registration accuracy under few-shot scenarios. Outperforms the single-task fully-supervised training state-of-the-art model in the metrics of Dice similarity coefficient and standard deviation of the Jacobian determinant.

## 1. Introduction

Image registration and segmentation are two highly related techniques that are critical for many medical image analysis tasks. They play important roles in medical information fusion, tumor growth detection, image-guided surgical treatment and radiation, and so on. Image registration seeks to map the corresponding anatomical structures of two input images (moving image and fixed image) to a transformation of the same coordinate system; while image segmentation seeks to identify voxel regions belonging to the same tissue structure based on image intensity distribution characteristics. In many AI-powered methods, registration and segmentation complement each other. For example, image warping through image registration is used for data augmentation of segmented samples; segmentation can provide weak supervision for image registration beyond image intensity and be used to evaluate registration quality. Segmentation methods based on deep learning (DL) techniques as well as convolutional neural networks (CNNs) are faster and better than classical methods when there are enough labeled samples for training [1]. Likewise, DL-based image registration (DLIR) achieves similar performance to iterative optimization-based methods but is much faster [2,3]. To take full advantage of the unique advantages of these two methods, the DL-based joint registration and segmentation (JRS) method are proposed. For example, dual-task learning methods based on multiple decoders [4,5]; weakly supervised registration learning methods assisted by automatic segmentation [6–8]; and segmentation self-training methods via registration as data augmentation [9,10]. These methods combine registration and segmentation or in the same framework, the two are joined by loss functions or as a unidirectional auxiliary task.

However, in some JRS methods, when segmentation is used for weakly supervised registration, the quality of segmentation pseudo-labels is often ignored [7]; when registration is used for segmentation, the noise effect of registration deformation on the input is not considered [10]. And unlabeled data is usually excluded from loss calculations. Other methods often require a large amount of labeled data for supervised training. However, acquiring the manually segmented labels of 3D

* Corresponding author.
*E-mail address:* fyang@foxmail.com (F. Yang).

https://doi.org/10.1016/j.bspc.2022.104294
Received 15 June 2022; Received in revised form 15 August 2022; Accepted 8 October 2022
Available online 20 October 2022

medical images, especially the deformation field ground truth, requires considerable expertise and time, which limits the improvement of the accuracy and robustness of deep learning models [11].

To overcome the lack of expert annotations for 3D medical images, and the limitation of other JRS methods, we propose a Joint Registration and Segmentation Self-training framework (JRSS) that uses pseudo-labels to provide additional supervised learning for segmentation and registration. JRSS allows the dual tasks to learn and promote each other in cyclic iterations, which together improve the dual-task performance of multi-organ segmentation and deformation registration in 3D medical images in few-shot scenarios. JRSS integrates multiple correction methods of injected noise, threshold screening and uncertainty estimation [12–14] to ensure forward optimization of pseudo-labels from coarse to fine. The quality-assessed and screened pseudo-labels facilitate the learning of weakly supervised registrations, and data deformations from the registrations act as input noise and data augmentation. JRSS progressively increases the weakly supervised training of segmentation and registration through joint self-training, realizing joint learning and knowledge complementation of segmentation and registration. We conduct experiments on two medical image datasets that differ widely in shape, size, and intensity distribution. The brain MRI dataset LPBA40 [16] and the Pancreas-CT dataset [15,17,18] from TCIA [19] are used to conduct extensive experiments, respectively, to fully verify the performance of JRSS in few-shot scenarios. The main contributions of this research are summarized as follows:

- A novel dual-task self-training framework for joint weakly supervised registration and segmentation pseudo-label learning.
- Using threshold screening and uncertainty estimation to correct the learning of noisy pseudo-labels from coarse to fine improves dual-task performance in a mutually reinforcing manner.
- Proposed method effectively improves the performance of single-task training registration and segmentation models in few-shot scenarios.

## 2. Related work

In recent years, DL-based pseudo-labeling strategies have provided an important method to enhance model performance with unlabeled data for label-scarce medical image tasks [20–24]. In the dual-task of medical image registration and segmentation, segmentation (pseudo) labels provide additional weakly supervised constraints for registration learning independent of image intensity. And registration can inherently provide a more reasonable data augmentation scheme for segmentation than random flipping, scaling, and affine deformation [25]. Furthermore, we treat the deformed image as a kind of input noise.

### 2.1. Weakly supervised registration

Compared with unsupervised registration, weakly supervised learning is closer to traditional feature-based registration methods to a certain extent. Label-Reg [26] is a representative work of weakly supervised registration learning. Label-Reg incorporates the segmentation label Dice loss into the loss function of weakly supervised training, which avoids the difficulty of measuring the similarity of multimodal image intensity so that the model can register images independently of the modality. Subsequently, segmentation label pixel-level similarity measures based on Dice coefficients are adopted by registration models such as VoxelMorph [2], PDD-Net [27], and AC-RegNet [6], and weighted with standard loss functions based on image intensity similarity. These works show that weakly supervised training of registration networks with Dice loss of segmentation labels can effectively improve registration accuracy. It also reflects the fact that weakly supervised methods rely on human-annotated anatomical knowledge, not just the statistical properties of image intensity matching, compared to other unsupervised learning, and the results are more in line with the doctor-annotated regions of interest (ROIs). However, the above weakly supervised methods rely too much on labeled data to improve the registration accuracy, and it is often difficult to ensure the smoothness of the deformation field in large deformation registration.

### 2.2. Joint registration and segmentation learning (JRS)

Recent joint registration and segmentation learning (JRS) methods have achieved remarkable results [9,28–32]. On the one hand, utilizing unsupervised registration or domain adaptation [33,34], a portion of labeled data can be aligned to unlabeled data to generate a new labeled training set that fits the structural features of the target set, which can be fully supervised for segmentation or registration learning. These methods utilize registration for label propagation, showing the advantages of using registration for data augmentation in JRS learning. On the other hand, the use of segmentation labels can make the registration network pay additional attention to the ROIs to constrain the network optimization and achieve more refined registration [4,5,35].

Our proposed method uses a similar model architecture as DeepAtlas [7], DeepRS [34] and RSegNet [8]. They study the mutual assistance of semi-supervised segmentation and weakly supervised registration through joint loss functions and segmentation pseudo-labels, demonstrating the advantages of joint learning of segmentation and registration. However, the segmentation network of DeepAtlas sets the Dice loss to 0 when the unlabeled data is input, which means that the unlabeled data will not participate in the optimization of the segmentation network. DeepRS utilizes GAN-based alignment confidence maps to measure registration, providing a weighted loss for weakly supervised segmentation. However, the deformation field is prone to interference from labels in the background, resulting in distorted deformations [49]. RSegNet requires fully supervised training on labeled data to ensure that pseudo-labels are beneficial for registration, the unlabeled data will not participate in training. Another JRS method, Cross-Stitch [36] uses units named cross-stitch to fuse the dual tasks of segmentation and registration at the architecture level, but also needs to rely on supervised training with labeled data. Our proposed JRSS uses a joint self-training strategy of registration and segmentation, builds a knowledge bridge for dual tasks via pseudo-labels, and multiple correction methods ensure a virtuous cycle of pseudo-labels, overcoming the limitations faced by the above JRS methods.

### 2.3. Noisy Pseudo-label learning

Semi-supervised segmentation self-training with pseudo-labels has achieved good performance [11,20–23], but the pseudo-labels generated by the model will still contain noisy predictions, so screening pseudo-labels is an essential process to avoid poor quality pseudo-labels affect the iterative training of the model. Our proposed method combines the recent SOTA semi-supervised self-training framework to screen pseudo-labels to ensure that the segmentation network is well-learned. The learning of noisy pseudo-labels is corrected by pseudo-label threshold screening [37] and uncertainty estimation [21,38]. Specifically, using the Kullback-Leibler divergence (KL-divergence) between the soft probability predictions of the teacher network and the student network as the uncertainty of the pseudo-label, so that the student model can get more reliable guidance from the teacher model.

## 3. Methods

The goal of this paper is to improve the performance of segmentation and registration networks by making full use of unlabeled data through joint registration and segmentation self-training, while only relying on a small number of human segmentation labels (few-shot). In the following sections, the proposed method of JRSS is mainly described, including the joint self-training architecture of dual models, noisy pseudo-label learning, weakly supervised registration learning, and their respective

loss functions.

### 3.1. Overview

Fig. 1 shows the model architecture of the proposed JRSS, which consists of two parts: a semi-supervised segmentation module (left) and a weakly supervised registration module (right). The segmentation network predicts pseudo-labels for unlabeled images, allowing the training of registration to be weakly supervised through anatomical similarity loss, and the training process of the registration network also guides the segmentation learning of unlabeled images.

Let $X = (X_i)_{i=1}^{K}$ represent $K$ samples in the training set, $E_l = \{X_i^l, S_i\}_{i=1}^{N}$ represent $N$ labeled samples, $E_u = \{(X_i^u)\}_{i=1}^{J}$ represents $J$ unlabeled samples, $N+J=K$. Where $X_i^l$ represents the $i$-th labeled sample, $S_i$ is its one-hot segmentation label, and $X_i^u$ represents the $i$-th unlabeled sample. And segmentation network $\psi_s$ with learnable parameters $\theta_s$ and registration network $\psi_r$ with parameters $\theta_r$.

Consider a pair of randomly picked images from $X$, one as a moving image $M$, and a fixed image $F$, along with their segmentation (pseudo) labels $S_M$ and $S_F$. During training, using $M$ and $F$ as input, the registration network $\psi_r$ receives stacked inputs of $M$ and $F$ and predicts a deformation field $\mathcal{D} = \psi_r(M, F;\ \theta_r))$, where $\mathcal{D}: \mathbb{R}^d \to \mathbb{R}^d$ ($d$ is the image dimension). Another registration component is the differentiable deformable module used in the Spatial Transformer Networks (STN) [39], which deforms the moving image $M$ using $\mathcal{D}$, resulting in warped images $M^w = M \circ \mathcal{D}$ and warped segmentations $S_M^w = S_M \circ \mathcal{D}$. The segmentation network $\psi_s$ predicts pseudo-labels $\widehat{S}_M = \text{argmax}(\psi_s(M;\ \theta_s))$ and $\widehat{S}_F = \text{argmax}(\psi_s(F;\ \theta_s))$, respectively, for weakly supervised learning of registration. During training, the parameter $\theta_s$ is optimized to minimize the weighted sum of Dice loss $\mathcal{L}_d$ and pixel-wise weighted cross-entropy loss $\mathcal{L}_c$ of the segmentation labels, and the uncertainty loss $\mathcal{L}_u$ of softened probabilistic prediction. The parameter $\theta_r$ is optimized to minimize the weighted sum of deformation field regularization loss $\mathcal{L}_r$, image intensity similarity loss $\mathcal{L}_i$ and weakly supervised Dice loss $\mathcal{L}_d$. The following sections describe the JRSS training strategy and loss function in detail.

### 3.2. Backbone and mono networks

JRSS uses the SOTA model obelisk-CNN (Seg-Net) [40] for abdominal segmentation as the backbone architecture for segmentation and registration networks. As shown in Fig. 2, it is 3D-UNet [41] that replaces the two-layer obelisk plugin. *obelisk1* has 512 spatial filter offsets and one-eighth resolution differentiable sampling grid, and the other has 128 spatial filter offsets and quarter resolution differentiable sampling grids. Each differentiable sampling grid is followed by a $1 \times 1 \times 1$ convolution to control the number of output channels. Since the top two layers of the original 3D-UNet are replaced with sparse convolutions, only about 230k trainable parameters remain. Compared with traditional convolution kernels, the sparse deformable convolution kernel of obelisk requires less of trainable parameters and less memory while obtaining high-quality results [40]. Except for the input and output layers, the registration network has the same architecture as the segmentation network. The registration adopts a similar way to VoxelMorph [2], connecting hybrid-obelisk-CNN and STN to form a differentiable deformable registration module. We train the single-task segmentation and registration network (called Mono-Net) as baselines for the proposed JRSS performance.

The single-task segmentation network is trained using the Dice loss $\mathcal{L}_d$ on labeled data; the registration network is trained using the deformation field regularization loss $\mathcal{L}_r$, the image intensity similarity loss $\mathcal{L}_i$, and the weakly supervised Dice loss $\mathcal{L}_d$. The loss function for registration is unsupervised or weakly supervised according to the experimental setting.

### 3.3. Joint registration and segmentation self-training

To fully utilize unlabeled data to participate in network training and provide accurate and sufficient anatomical segmentation labels through

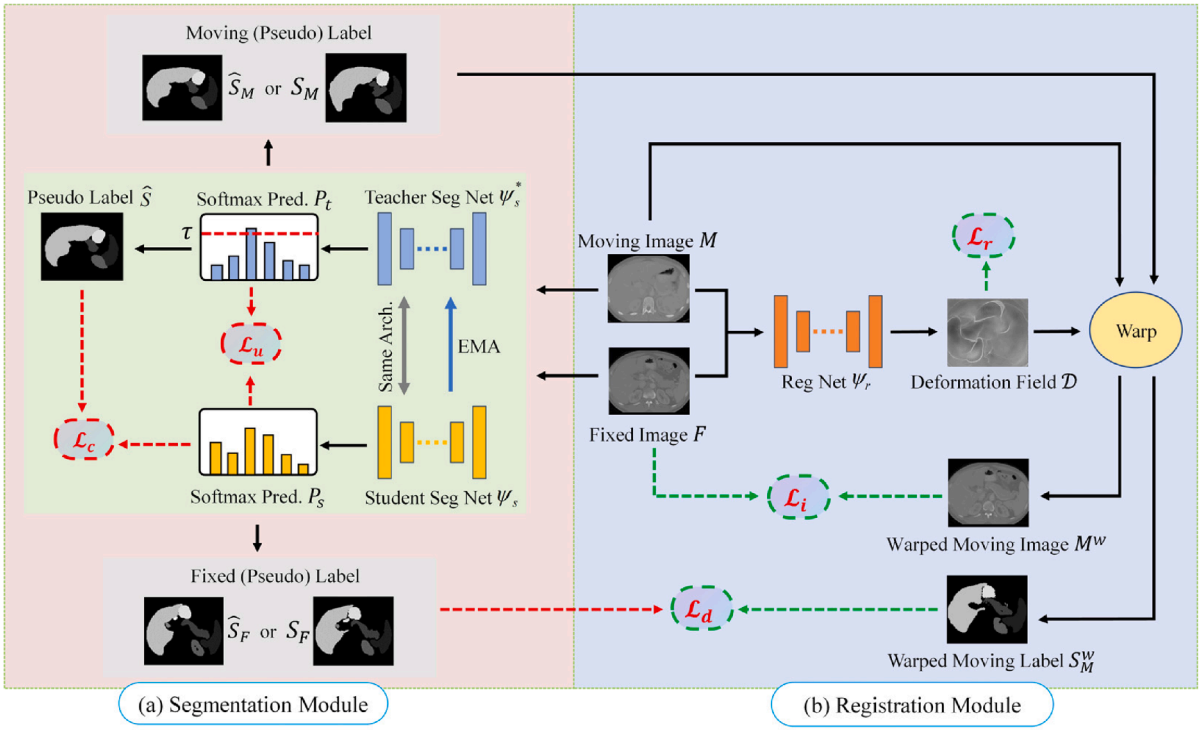



**Fig. 1.** Model architecture for JRSS. (a) Segmentation module predicts pseudo-labels on unlabeled data to provide weakly supervised learning for (b) Registration module. We compute the KL-divergence as uncertainty estimation loss $\mathcal{L}_u$ on the softmax probability map of the student seg net and teacher seg net to correct pseudo-label learning. The deformation field $\mathcal{D}$ of the registration module will be used as input noise as well as data augmentation to improve the segmentation performance.

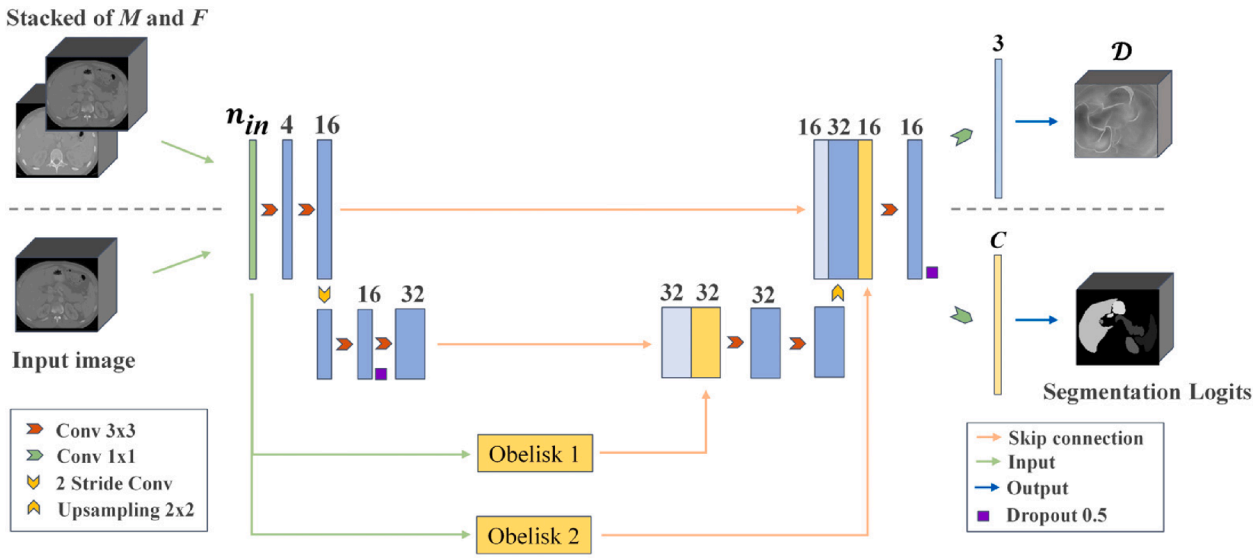



**Fig. 2.** Each convolution of hybrid-obelisk-CNN is a $3 \times 3 \times 3$ kernel, followed by a BatchNorm layer and a LeakyReLU activation layer with parameter 0.2. The encoder uses a convolution kernel with stride = 2 to reduce the spatial resolution by half, while the decoder uses an up sampling layer. Dropout is applied to the last layer of down sampling and up sampling with a dropout rate of 0.5. Replace the original third layer and bottle neck with two obelisk plugins.

automatic segmentation, JRSS adopts a joint self-training strategy as shown in Fig. 3. The self-training strategy of JRSS follows the following four steps and performs iterative training.

1) Pre-trained Teacher Segmentation Network: For $N$ labeled data, optimize the segmentation network parameter $\theta_s$ to minimize the soft multi-class Dice loss which can addresses imbalances inherently:

$$\mathcal{L}_d(S_i, X_i^l) = 1 - \frac{1}{C}\sum_{c=1}^{C} Dice\left(S_i^c, \psi_s^c(X_i^l;\theta_s)\right), \forall i = 1, \ldots N,$$

$$where \quad \mathrm{Dice}(S^c, \widehat{S}^c) = 2\cdot\frac{|S^c \cap \widehat{S}^c|}{|S^c| + |\widehat{S}^c|} \tag{1}$$

where $C$ denotes the number of classes and is also the number of channels of model prediction; $S^c$ indicates the voxels of structure $c$. The training of the segmentation network at this stage is formulated as:

$$\theta_s^* = \underset{\theta_s}{\mathrm{argmin}}\left\{\mathcal{L}_d(S_i, X_i^l)\right\} \tag{2}$$

2) Pseudo-label Screening: Use teacher segmentation network with pre-trained parameters $\theta_s^\star$ to generate pseudo-labels for unlabeled data:

$$\widehat{S}_i = \arg\max\left(\psi_s(X_i^u;\theta_s^*)\right), \forall i = 1, \cdots, J \tag{3}$$

The segmentation network predicts a pseudo-label for each unlabeled sample. To obtain a high-quality pseudo-label, for the softmax probability distribution predicted by the model:$P_i = \mathrm{softmax}(\psi_s(X_i^u;\theta_s^*))$, we set a class-level threshold $\tau$ in a similar way to FixMatch [37] to keep only predictions with $\max(P_i^c) > \tau$, where $P_i^c$ is the prediction probabilities channel of class $c$ (out of $C$). Then, we use $\widehat{S}_i = \mathrm{argmax}\ (P_i)$ as a one-hot pseudo-label. Using hard pseudo-labels can make pseudo-labels closer to entropy minimization, because the predicted value of the model on unlabeled data should be low-entropy (high confidence) [42], which is more conducive to the next step for student segmentation network to learn.

3) Automatic Correction of Pseudo-label Learning: Then, data with (pseudo) labels are combined into a new training set $X^{'} = \left\{\left(X_1^l, S_1\right), \left(X_2^u, \widehat{S}_2\right), \cdots, \left(X_i^*, *\right)_{i=1}^{K}\right\}$ to train a new student segmentation network, and add noise to the model using dropout [11]. Compute the pixel-wise weighted cross-entropy loss for labeled data and pseudo-labeled data:

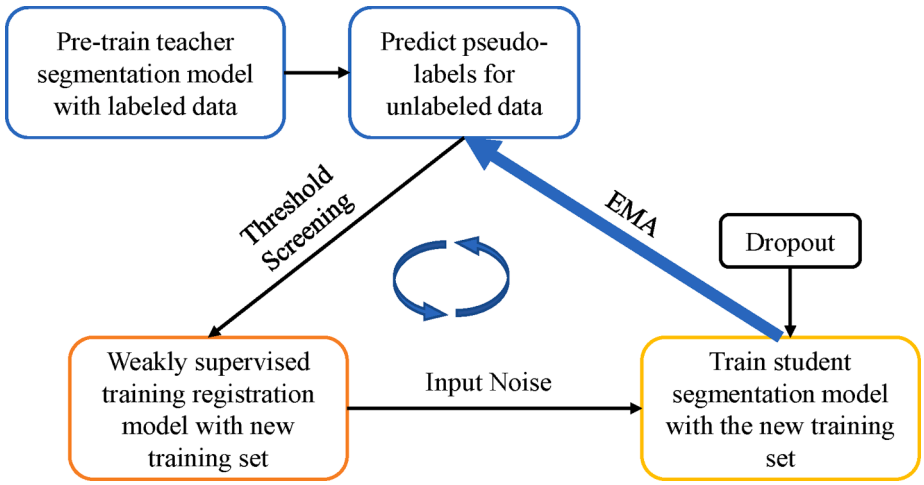



**Fig. 3.** Joint self-training strategy for JRSS.

$$\mathcal{L}_c\left(S_i, X_i^l, \widehat{S}_i, X_i^u\right) = \frac{1}{N}\sum_{i=1}^{N} CE\left(S_i, \psi_s^{noised}\left(X_i^l;\theta_s\right)\right) + \frac{1}{J}\sum_{i=1}^{J} CE\left(\widehat{S}_i, \psi_s^{noised}\left(X_i^u;\theta_s\right)\right) \tag{4}$$

When dropout is used as model noise, the student is forced to mimic a more powerful ensemble model, which ensures that the segmentation network in self-training will continue to make progress.

In addition, for unlabeled data, we apply the pre-trained registration network to predict the deformation field $\mathcal{D}$ on the current moving image for data augmentation, and as a kind of input noise to enhance the learning of student model. To avoid the segmentation network in training being negatively affected by noisy labels, the KL-divergence $\mathrm{KL}(\mathbf{a}\|\mathbf{b}) = \sum \mathbf{a}\log\frac{\mathbf{a}}{\mathbf{b}}$ between the two softmax predictions of the teacher model and the student model is used as the uncertainty estimation [21,38] to correct the pseudo-labels Interference with student learning:

$$\mathcal{L}_u(P_t, P_s) = \frac{1}{C}\sum_{c=1}^{C}\left(KL\left(P_t^c \| P_s^c\right) + KL\left(P_s^c \| P_t^c\right)\right) \tag{5}$$

where $P_t^c$ and $P_s^c$ are the softmax probability channel of class $c$ in the predictions of the teacher and student models, respectively. By minimizing their KL-divergence, the student model can obtain knowledge distillation from the teacher model [46].

We also enable the segmentation network to share the knowledge of the registration network by minimizing the Dice loss $\mathcal{L}_d$ in weakly supervised registration learning.

Hence, the total loss $\mathcal{L}_s = \mathcal{L}_c(S_i, X_i^l, \widehat{S}_i, X_i^u) + \lambda_u\mathcal{L}_u(P_t, P_s) + \lambda_d\mathcal{L}_d(S_F, S_M, \mathcal{D})$ used for student segmentation network training. Where $\lambda_u$ and $\lambda_d$ are scalar hyperparameters, representing the weight of uncertainty loss and weakly supervised Dice loss in the total segmentation loss, respectively. The training of the segmentation network is formulated as:

$$\theta_s^* = \underset{\theta_s}{\operatorname{argmin}}\{\mathcal{L}_s\} \tag{6}$$

4) Iterative Training: Treat the exponential moving average (EMA) of the student's weights $\theta_s$ as a new teacher segmentation model and go back to step **2)**.

### 3.4. Weakly supervised registration using combined data

In the weakly supervised registration learning process, in addition to the image intensity similarity loss $\mathcal{L}_i$ and the deformation field regularization loss $\mathcal{L}_r$, by adding the Dice loss $\mathcal{L}_d$, the overlap between the warped moving labels and the fixed labels is encouraged to maximize, so as to optimize the registration network. Therefore, in JRSS, for images without segmentation labels, a new training set $X^*$ with (pseudo) labeled data is used to provide weakly supervised training for the registration network.

1) Soft Multi-class Dice Loss $\mathcal{L}_d$: Soft multi-class Dice loss between warped labels and fixed labels:

$$\mathcal{L}_d(S_F, S_M, \mathcal{D}) = 1 - \frac{1}{C}\sum_{c=1}^{C} \mathcal{D}ice\left(S_F^c, S_M^c D\right) \tag{7}$$

If the unlabeled image is not segmented to obtain satisfactory pseudo-labels, then $\mathcal{L}_d = 0$, and the registration network performs unsupervised training through the following two loss functions.

2) Intensity Similarity LossL: Like VoxelMorph [2], we use a standard mean square error (MSE) loss that encourages aligned image intensities to achieve maximum similarity:

$$\mathcal{L}_i(M^w, F) = MSE(M^w, F) \tag{8}$$

3) **Regularization Loss** $\mathcal{L}_r$: We use the diffusion regularizer on the spatial gradients of the deformation fields$\nabla D(o)$, to encourage registration to output a smooth D, same as VoxelMorph [2]:

$$\mathcal{L}_r(D) = \sum_{o\in\Omega}||\nabla D(o)||^2 \tag{9}$$

where $\Omega$ denotes the set of voxels of D, **o** denotes the displacement of each voxel. Therefore, the training of weakly supervised registration network can be formulated as:

$$\theta_r^* = \underset{\theta_r}{\operatorname{argmin}}\{\mathcal{L}_i(M^w, F) + \lambda_r\mathcal{L}_r(D) + \lambda_d\mathcal{L}_d(S_F, S_M, D)\} \tag{10}$$

where $\lambda_r$, $\lambda_d$ are scalar hyperparameters, representing the weight of regularization loss and weakly supervised Dice loss, respectively.

## 4. Experimental results

### 4.1. Datasets and preprocess

We evaluate our technique using two segmentation and registration benchmark datasets: The LONI Probabilistic Brain Atlas (LPBA40) dataset [16], and the abdominal CT scans from collections of The Cancer Imaging Archive (TCIA) project [17–19]. We have conducted extensive experiments on different organs and modalities.

1) LPBA40: The LPBA40 dataset contains 40 T1-weighted MRI 3D brain scans, each scan containing subcortical segmentation of 56 structures manually annotated by experts. To facilitate numerical statistics, we divide the 56 structures into 5 large regions and 5 independent structures according to [16]. We resampled all MRI scans to an isotropic resolution of $1^3$ mm, and then cropped them to a size of $160 \times 192 \times 160$ by center. To obtain convincing experimental results, the LPBA40 dataset is randomly divided into 28:2:10 for training, validation, and testing, respectively. In the registration experiment, we randomly select 2 scans from the test set as fixed images.

2) TCIA: The TCIA [19] dataset was originally a pancreas CT dataset introduced by Roth et al. [15,17]. More abdominal organs were subsequently annotated by Gibson et al. [18], extending the manual segmentation labels to 8 ROIs and retaining 43 abdominal CT scans. Some of these tissues and organs are extremely challenging for registration networks due to their small size, variable shape, and difficulty in image intensity similarity measurement. We preprocessed using the following steps: all scans were resampled to an isotropic resolution of $1^3$ mm, center-cropped to a suitable field of view of $144 \times 144 \times 144$. In the experimental phase, the TCIA dataset is randomly divided into 30:3:10 for training, validation, and testing, respectively. The registration experiment also requires randomly selecting 2 scans from the test set as fixed images. For the TCIA dataset, the traditional iterative registration tool Advanced Normalization Tools (ANTs) [43] was used for affine pre-registration before training, however, the initial average Dice overlap was only 33 % (see Table 2).

During training, we simulate few-shot scenarios by selecting a specified number of scans and their labels from the training set, using all labels from the validation and test sets. All scans are normalized to $\mathcal{N}(0, 1)$ by an end-to-end automated pipeline before being fed into models for training and testing.

**Table 1**
Dice similarity coefficient (DSC: Avg. (Std.)) of segmentation results in few-shot scenarios (only 5 labeled data are used for training).

| Methods ($N = 5$) | DSC ↑ | |
|---|---|---|
| | LPBA40 | TCIA |
| VM-UNet | 0.671 (0.125) | 0.642 (0.144) |
| DeepAtlas | 0.683 (0.149) | 0.669 (0.172) |
| RSegNet | 0.726 (0.133) | 0.701 (0.167) |
| Mono-Seg | 0.679 (0.138) | 0.644 (0.140) |
| (Our) JRSS-Seg | **0.795** (0.126) | **0.753** (0.151) |

**Table 2**
Dice similarity coefficient, standard deviation of Jacobian determinant, and runtime (DSC, Std(Jac) and Runtime: Avg. (Std.)) of registration results in few-shot scenes (only 5 labeled data were used for training).

| Methods ($N=5$) | LPBA40 | | TCIA | | Run Time / s |
|---|---|---|---|---|---|
| | DSC ↑ | Std(Jac) ↓ | DSC ↑ | Std(Jac) ↓ | |
| Initial | 0.576 (0.162) | – | 0.332 (0.258) | – | – |
| ANTs (SyN) | 0.765 (0.126) | 0.304 (0.109) | 0.561 (0.169) | 0.422 (0.114) | 29.44 (4.23) |
| VoxelMorph | 0.727 (0.134) | 0.665 (0.122) | 0.453 (0.278) | 0.793 (0.120) | 0.29 (0.15) |
| PDD-Net | 0.711 (0.129) | 0.445 (0.146) | 0.468 (0.167) | 0.638 (0.172) | 0.46 (0.40) |
| DeepAtlas | 0.702 (0.123) | 0.772 (0.139) | 0.460 (0.221) | 0.862 (0.138) | 0.30 (0.11) |
| RSegNet | 0.724 (0.114) | 0.589 (0.104) | 0.472 (0.144) | 0.718 (0.133) | 0.41 (0.39) |
| Mono-Reg | 0.721 (0.133) | 0.548 (0.131) | 0.459 (0.179) | 0.691 (0.141) | 0.36 (0.47) |
| (Our) JRSS-Reg | **0.759** (0.130) | 0.544 (0.105) | **0.539** (0.174) | 0.684 (0.157) | 0.36 (0.47) |

### 4.2. Baseline methods

To investigate the performance of the proposed method, we use SyN [47] implemented in ANTs as the traditional registration baseline. In addition, JRSS is compared with existing semi-supervised segmentation and weakly supervised registration methods, including:

1) **VoxelMorph** [2]: VoxelMorph (VM) is the most important baseline method in the field of DL-based registration. Its proposed combination of UNet architecture CNN and STN is the dominant model architecture of recent DLIR methods [44]. Since the registration module of JRSS is consistent with VM, we add $\mathcal{L}_d$ for weakly supervised registration training. Furthermore, we use the backbone of VM (which we call VM-UNet) as the segmentation baseline.

2) **PDD-Net** [27]: PDD-Net Uses obelisk-CNN as the discrete registration model for feature extraction, the registration network is trained with weak supervision. PDD-Net can efficiently handle large deformation registration in the abdomen.

3) **DeepAtlas** [7]: DeepAtlas (DA) proposes a method of joint semi-supervised segmentation and weakly supervised registration learning to reduce annotation dependence. However, DA does not perform pseudo-label learning and correction.

4) **RSegNet** [8]: RSegNet is another JRS method. Segmentation consistency and better deformation regularity are achieved by introducing consistency supervision into the JRS framework. However, RSegNet requires fully supervised training with labeled data to ensure segmentation pseudo-labels are beneficial for registration.

5) **Mono-Net**: We use separately trained segmentation and registration networks as baselines. We perform separate fully-supervised pre-training ($N=K$) of segmentation and registration networks as upper bound models for segmentation and registration, respectively. In addition, the registration network performs separate unsupervised pre-training as another baseline.

6) **JRSS** (ours): Since joint training from scratch is time-consuming, we use the two separately pre-trained semi-supervised Mono-Nets to initialize the models for joint self-training. This is an important reason why we choose to use hybrid-obelisk-CNN as the backbone, which is memory efficient and easier to train. DeepAtlas and RSegNet initialize the joint model with the pre-trained single-task network in the same way.

### 4.3. Implementation details

The proposed method is implemented by PyTorch [45] framework on an Nvidia V100 GPU with 32 GB of video memory. During training, all models use Adam as the optimizer, the learning rate decays exponentially with $\gamma = 0.99$ throughout, and the batch size is set to 2. For the segmentation model, we make default settings according to the backbone [40]. The weight decay is $10^{-4}$, the initial learning rate is 0.001, $\lambda_u = 0.5$, $\lambda_d = 1$ if labeled, else 0. For the registration model, we make default settings according to VoxelMorph [2]. The initial learning rate is $4 \times 10^{-4}$, $\lambda_d = 0.1$ if labeled else 0, $\lambda_r = 0.025$. We use a warm-up mechanism for the initial 300 steps with a higher $\lambda_r$ to stabilize the deformation field. For baselines, we reproduced using their official open-source code and default parameters and evaluated the model with the best performance on the validation set.

For the training set $X$ with $K$ samples, take $N$ samples as labeled samples. We set $\mathcal{N} \in \{1, 3, 5, 7\}$ to verify the performance of JRSS based on the fact that medical image samples and their labels are rare in this experiment and the actual situation. All baselines and JRSS do the same semi-supervised training.

1) **Baselines Training Details**: For baselines, In addition to the above default settings, when the input data has no segmentation labels, $\lambda_d = 0$, and only unsupervised registration is performed at this time. The training process of the baseline models always uses random affine deformations as data augmentation. Models trained with full supervision provide an upper bound on the performance of JRSS models.

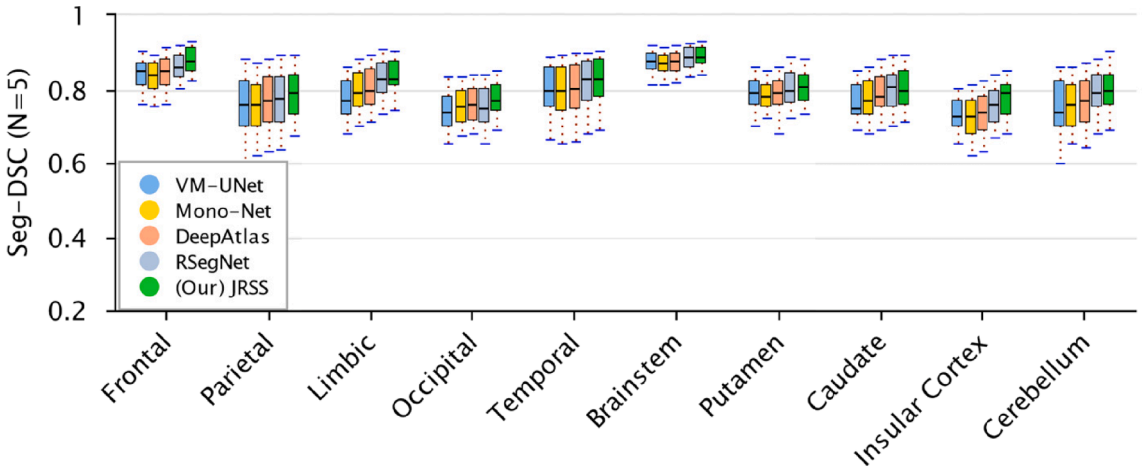
2) **JRSS Training Details**: For JRSS, according to the proposed joint self-training framework, semi-supervised training of the segmentation model is performed first. Since the segmentation accuracy of brain images is easier to improve and each organ remains stable, we empirically set a higher segmentation threshold $\tau = 0.7$ for the LPBA40 dataset. We set the threshold for TCIA as $\tau = 0.6$. Our ablation experiments further confirm such a setting. Joint training starts only when the segmentation network can predict pseudo-labels that exceed the initial threshold.

In joint self-training, first, registration weakly supervised training is performed using the new training set $X^{'}$. Second, refer to [21], dropout is applied to the last layer of down-sampling and up-sampling, and will randomly drop some channels of the feature map with a rate of 0.5. Use the same new training set $X^{'}$ to train the student segmentation model. Finally, the EMA of the student segmentation model parameters is used as the new teacher segmentation network to predict pseudo-labels with a decay of 0.99 according to [14]. When the registration and segmentation networks are alternately trained, one is fixed and the other is trained. In one round of joint self-training iterations, we determine by validation set accuracy that the ratio of the number of epochs required to train the registration network to the segmentation network is 5:1. In the experiment, JRSS achieved the indicators shown in this paper after three rounds of joint self-training iterations. The proposed method as well as all baselines are trained three times and the average results are reported.

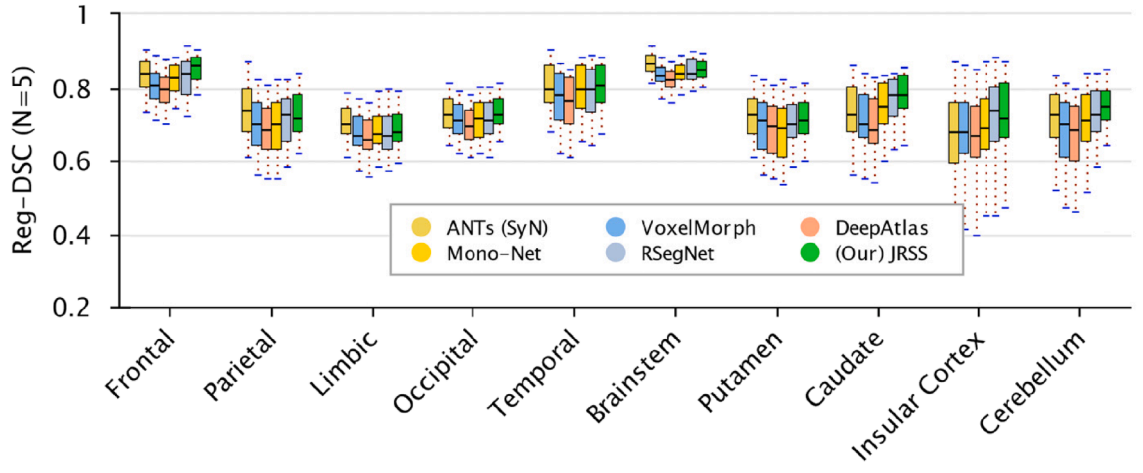
### 4.4. Results and comparison

We use a series of metrics to evaluate the performance of the proposed method for segmentation and registration, including the Dice similarity coefficient (DSC) of the segmentation labels, and the standard deviation of the Jacobian determinant (std(Jac)) to measure the predicted deformation field smoothness. The standard deviation (std) of these metrics are also provided to evaluate the stability of these models. In addition, the runtime is used to measure the application performance of the methods.

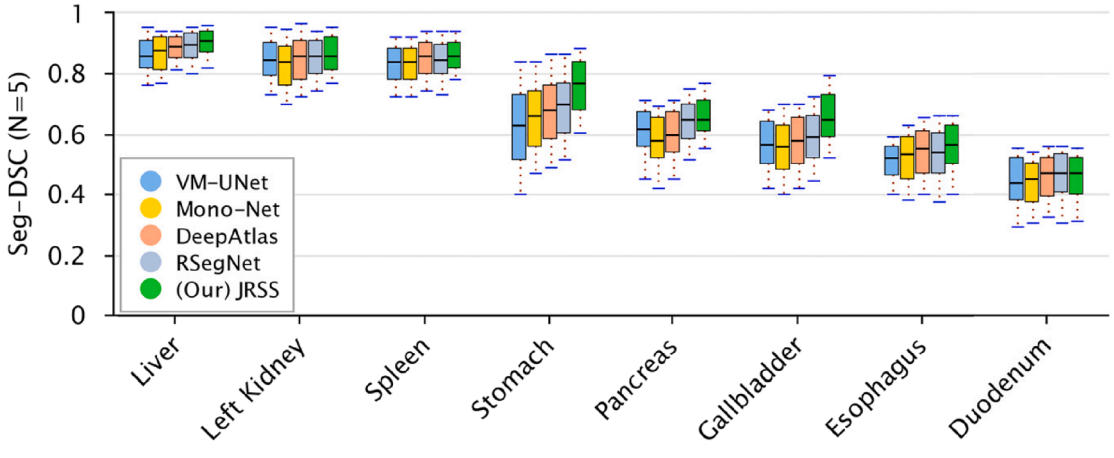
1) **Segmentation Results**: Fig. 4 (a) and (c) show the DSC scores for each structure on the LPBA40 and TCIA test set under few-shot scenario ($N=5$). We find that with the joint self-training segmentation network in JRSS, with the help of pseudo-label generation, denoising and screening, our JRSS performs much better than the single networks. Comparable DSC results were achieved for all structures, both on abdomen and brain segmentation. We found that the high-quality pseudo-labels provided by joint self-training significantly improved some specific ROIs, such as stomach, pancreas, gallbladder, and insular




(a) LPBA40 segmentation DSC.




(b) LPBA40 registration DSC.

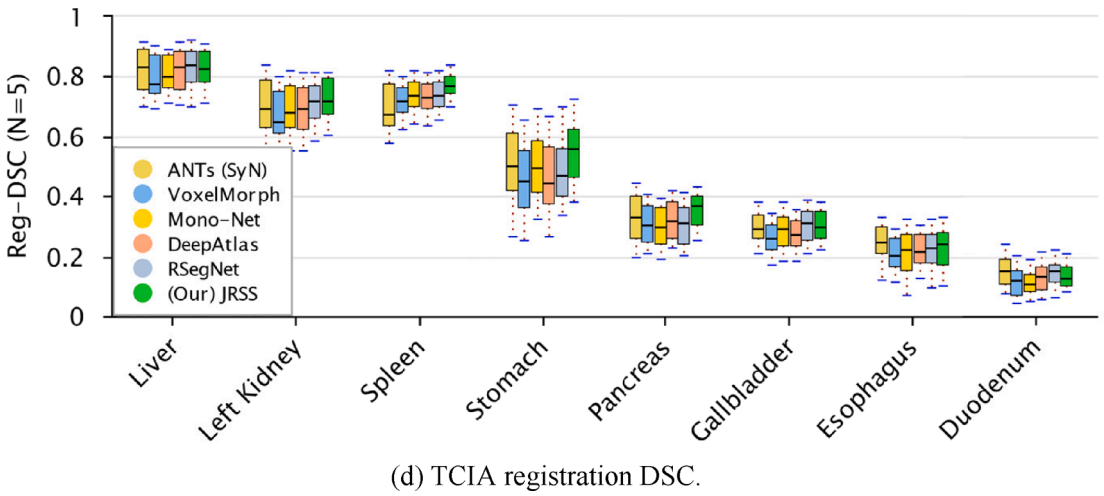



(c) TCIA segmentation DSC.

**Fig. 4.** Boxplots depict the average DSC scores of segmentation and registration of each anatomical structure on LPBA40 (top) and TCIA (bottom) test sets by different methods when only five labeled data ($N = 5$) were used for training.

cortex in the brain. These ROIs tend to make segmentation based on image intensity more difficult due to their variable structures and blurred boundaries. The intensity distribution-independent structural features provided by pseudo-labels can significantly improve the segmentation performance of these regions.

As shown in Table 1, the DSC scores of all structures are 11.7 % (LPBA40) and 10.1 % (TCIA) higher than the average of the other three DL-based methods (VM-UNet, DeepAtlas and Mono-Seg), respectively.




(d) TCIA registration DSC.

**Fig. 4.** (*continued*).

RSegNet strengthens joint learning compared to DeepAtlas through the consistency constraints of the registered bidirectional deformation. However, both of them lack the ability to provide qualified pseudo-label supervision for unlabeled data, and perform inferior to JRSS in few-shot scenarios.

Furthermore, to show these improvements more intuitively, Fig. 5 shows the examples of segmentation results. The segmentation results of JRSS are visually significantly improved, reflected in fine boundaries and few false segmentations.

2) **Registration Results**: As shown in (b) and (d) in Fig. 4, our JRSS shows advantages over the other three semi-supervised methods in the registration of all structures and achieves accuracy close to traditional methods. Benefiting from the improved segmentation pseudo-labels, the registration achieves significant improvements in the stomach in the abdomen, and insular cortex in the brain under the constraints of more precise organ boundaries.

As shown in Table 2, we compare the mean DSC, computation time and standard deviation of the Jacobian determinant for six different deformable registration methods. In addition, the initial affine pre-registration is also listed as a reference. We find that DL-based methods can achieve comparable registration accuracy with a small amount of label supervision compared to traditional methods. Constraints based on regularization loss can produce relatively smooth and reasonable deformation fields. In addition, JRSS has the highest average DSC, and Mono-Net has similar accuracy to other single-task networks, but the large receptive field of the obelisk plugin provides advantages for smooth deformation fields (lower Jacobian standard deviation). Notably, a comparison with Mono-Net reveals that JRSS benefits from self-training segmentation, providing additional pseudo-label guidance on top of a small number of ground-truth labels. Other JRS methods, DeepAtlas and RSegNet seem to be affected by noisy pseudo-labels and perform less than expected in weakly supervised registration. Furthermore, our JRSS and other learning-based methods only need to run in less than one second on GPU, which is more than 100 times faster than ANTs (SyN) running on CPU.

Fig. 6 shows examples of registration results using six different methods. We can observe that in terms of our selection of highlighted ROIs for the brain and abdomen, JRSS can produce registration results that are closer to fixed images than networks trained separately. Specifically, JRSS can effectively handle the blurred boundaries and easily deformed stomach and pancreas registrations, which are difficult to capture by other registrations based only on image intensity.

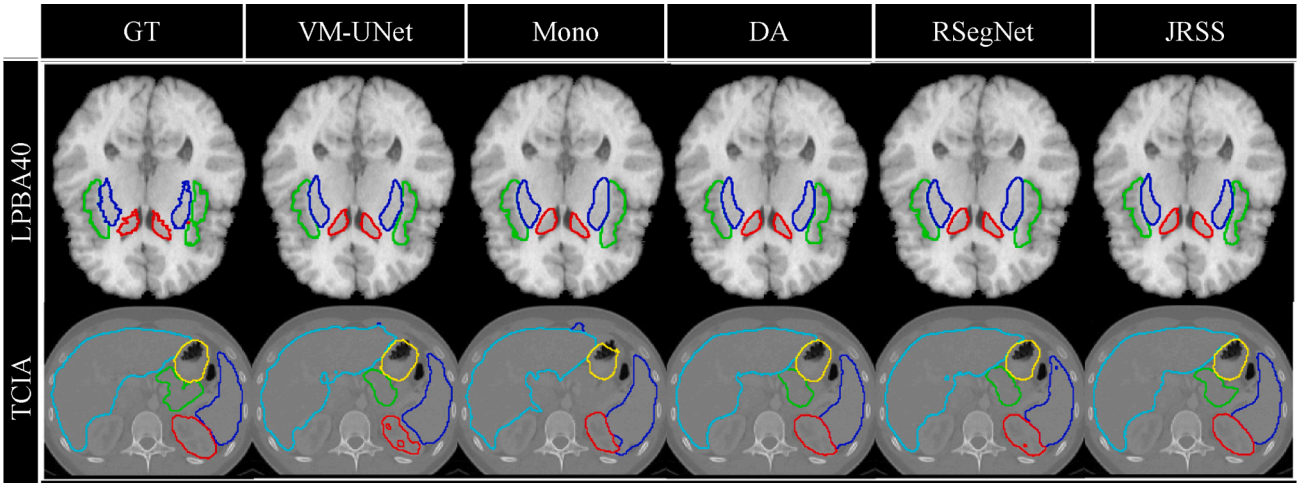



**Fig. 5.** Examples of segmentation results from brain MRI (top) and abdominal CT (bottom) (axial slices of 3D images taken) using 5 labeled training samples (N = 5). From left to right are the segmentation results of various models, and their boundaries of segmentation labels. It is obvious that the proposed method (JRSS) is closer to the ground truth (GT) than other methods in semi-supervised scenarios. Sections of the brain show the insular cortex (green), putamen (blue), and caudate (red) of the left and right brain; sections of the abdomen show the liver (blue), pancreas (green), stomach (yellow), left kidney (red) and spleen (purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

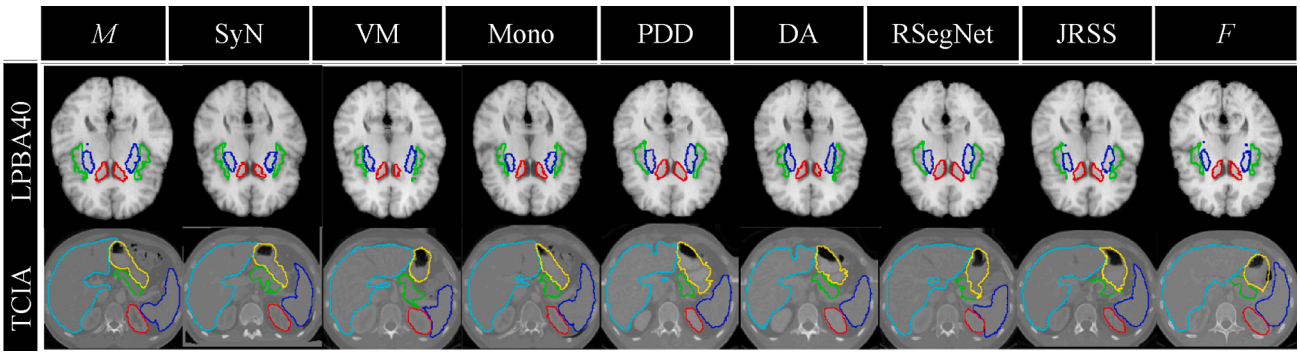



**Fig. 6.** Examples of registration results from brain MRI (top) and abdominal CT (bottom) using 5 labeled training samples (N = 5). The first and last columns are moving (M) and fixed (F) images respectively. The registration accuracy of JRSS in the semi-supervised scenarios is closer to the fixed image than other methods.

### 4.5. Ablation studies

1) **Improvement of Pseudo-labels**: The statistical results in Table 3 and Fig. 7 show that in the few-shot scenario, JRSS achieves a significant improvement in dual-task performance compared to the single-task network and the unsupervised scenario. The segmentation and registration accuracy increases with the number of labeled training samples, indicating that the quality of pseudo-labels has been enhanced, and the joint self-training strategy of JRSS enables a virtuous cycle. In the experiments, we found that when $N = 5$, the JRSS after two rounds of iterative training can generate qualified pseudo-labels for all unlabeled data, which makes the third round of joint self-training be carried out in a fully supervised manner. This also gives JRSS a huge advantage over single-task segmentation networks for semi-supervised learning in the same few-shot scenario, with an average of about 10 % higher DSC accuracy (see Table 3). The improvement of segmentation accuracy assists the improvement of registration. The unsupervised registration of the brain ($N = 0$) has achieved high accuracy, however, the JRSS for the more difficult TCIA abdominal registration has significantly improved the DSC accuracy compared to Mono-Net in the same weakly supervised setting (+5.6 % at $N = 1$; +8.0 % at $N = 5$, see Table 3). It is worth noting that when $N = 5$, the registration accuracy of JRSS on TCIA dataset is already better than that of fully supervised VoxelMorph; when $N = 7$, the segmentation accuracy of JRSS on LPBA40 dataset is also better than that of fully supervised RSegNet.

The segmentation and registration examples in Figs. 8 and 9 show that, with the increase of labeled training data, the accuracy of JRSS is improved, and the results are closer to ground-truth labels and fully supervised model (upper bound). In addition, under the same $N$, the results of JRSS have a significant advantage over Mono-Net. This verifies the superiority of JRSS in the improved dual-task performance in few-shot scenarios.

2) **Obelisk Kernel Improves Large Deformation Registration**: As shown in Table 3, we find that using only the traditional UNet-like combined with the STN like VoxelMorph may not be sufficiently capable of constraining the smoothness of the deformation field, especially in weakly supervised registration that requires segmentation label assistance, and when large deformations are registered. the Jacobian standard deviation (std(Jac)) of the Voxelmorph predicted deformation field varies significantly from 0.66 to 1.668 under the TCIA abdominal large deformation registration in the unsupervised and fully supervised cases. In contrast, using hybrid-obelisk-CNN as the backbone of Mono-Net and JRSS, the std(Jac) of their deformation field can still remain stable in the lower interval [0.671, 0.998] under weak supervision. In fact, our framework avoids the need to fine-tune the network structure. We choose hybrid-obelisk-CNN as backbone due to its efficient computational performance, especially under large deformation registration. So we can focus on verifying the effectiveness of JRSS. In fact, the backbone of JRSS can choose other networks to plug-and-play.

3) **Pseudo Label Threshold**: As shown in Fig. 10, We investigate the optimal choice of confidence threshold $\tau$ for different datasets and different numbers of labeled data. We experiment with the optimal pseudo-label screening threshold for JRSS in the range of $\tau \in [0.5, 0.8]$ according to the threshold settings of [21]. And get the best model under each threshold based on the accuracy on the validation set. Fig. 10 shows that JRSS is robust to the choice of $\tau$ in [0.6, 0.7] under different few-shot scenarios. For the TCIA dataset, a threshold of 0.6 shows the best segmentation pseudo-label quality; while for the LPBA40 dataset, it is 0.7; both decreasing and increasing have negative effects. The threshold controls the trade-off between the quality and quantity of

**Table 3**
Segmentation and registration evaluation results (Avg. (Std.)) under different supervised scenarios. Where $N = 0$ and $N = K$ denote unsupervised and fully supervised scenarios respectively.

| Labeled ($N$) | Methods | LPBA40 | | | TCIA | | |
|---|---|---|---|---|---|---|---|
| | | Segmentation | Registration | | Segmentation | Registration | |
| | | DSC ↑ | DSC ↑ | Std(Jac) ↓ | DSC ↑ | DSC ↑ | Std(Jac) ↓ |
| 0 | VoxelMorph | – | 0.711 (0.137) | 0.347 (0.107) | – | 0.428 (0.117) | 0.460 (0.122) |
| | Mono-Net | – | 0.713 (0.121) | 0.372 (0.117) | – | 0.440 (0.131) | 0.471 (0.127) |
| 1 | Mono-Net | 0.653 (0.133) | 0.720 (0.166) | 0.421 (0.203) | 0.609 (0.192) | 0.448 (0.157) | 0.641 (0.089) |
| | **(Our) JRSS** | **0.756** (0.128) | **0.741** (0.147) | 0.423 (0.188) | **0.691** (0.163) | **0.514** (0.158) | 0.663 (0.137) |
| 3 | Mono-Net | 0.668 (0.118) | 0.727 (0.129) | 0.508 (0.130) | 0.626 (0.132) | 0.452 (0.155) | 0.678 (0.141) |
| | **(Our) JRSS** | **0.765** (0.121) | **0.751** (0.120) | 0.510 (0.115) | **0.719** (0.143) | **0.520** (0.164) | 0.664 (0.167) |
| 5 | Mono-Net | 0.679 (0.138) | 0.731 (0.133) | 0.548 (0.131) | 0.644 (0.140) | 0.459 (0.179) | 0.691 (0.141) |
| | **(Our) JRSS** | **0.795** (0.126) | **0.759** (0.130) | 0.544 (0.105) | **0.753** (0.151) | **0.539** (0.174) | 0.684 (0.157) |
| 7 | Mono-Net | 0.698 (0.133) | 0.748 (0.122) | 0.649 (0.144) | 0.649 (0.120) | 0.466 (0.177) | 0.703 (0.143) |
| | **(Our) JRSS** | **0.805** (0.106) | **0.769** (0.128) | 0.634 (0.135) | **0.769** (0.141) | **0.547** (0.178) | 0.714 (0.147) |
| $K$ (full-sup) | RSegNet | 0.796 (0.063) | 0.781 (0.128) | 0.901 (0.111) | 0.787 (0.133) | 0.549 (0.143) | 1.121 (0.142) |
| | VoxelMorph | – | 0.774 (0.186) | 0.925 (0.128) | – | 0.532 (0.161) | 1.668 (0.132) |
| | Mono-Net | 0.834 (0.193) | 0.783 (0.255) | 0.839 (0.147) | 0.791 (0.142) | 0.555 (0.201) | 0.998 (0.146) |

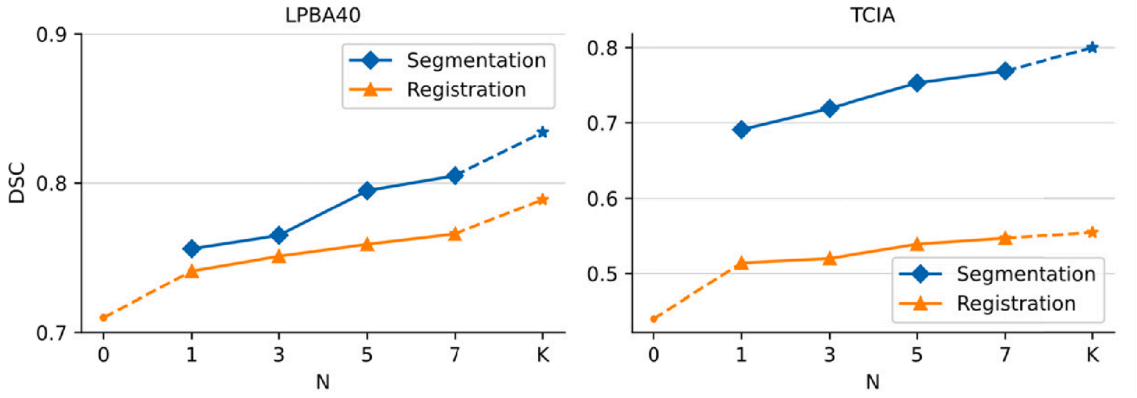



**Fig. 7.** The lines show the simultaneous improvement of the segmentation and registration accuracy of JRSS with the increase of labeled training samples (N).

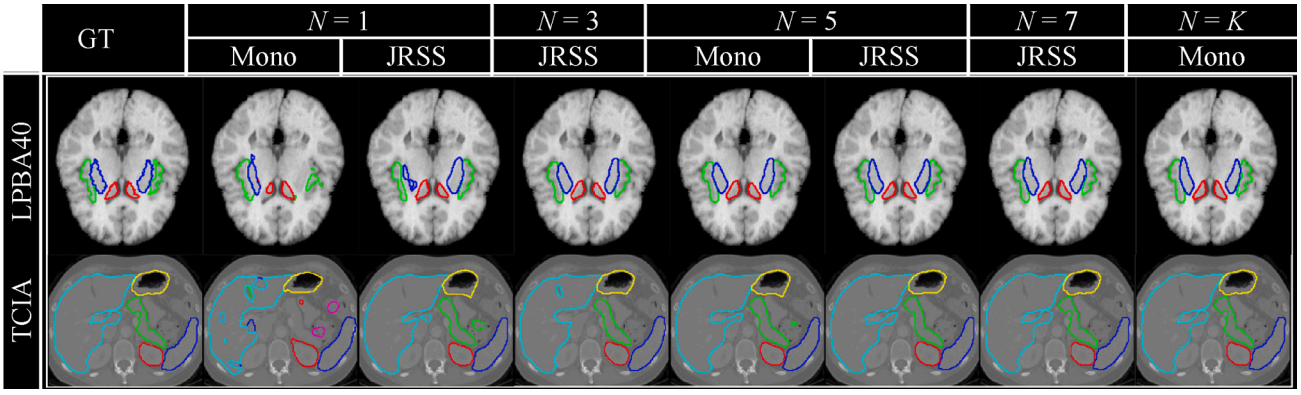



**Fig. 8.** Examples of segmentation results of JRSS with different numbers of labeled training samples (N).

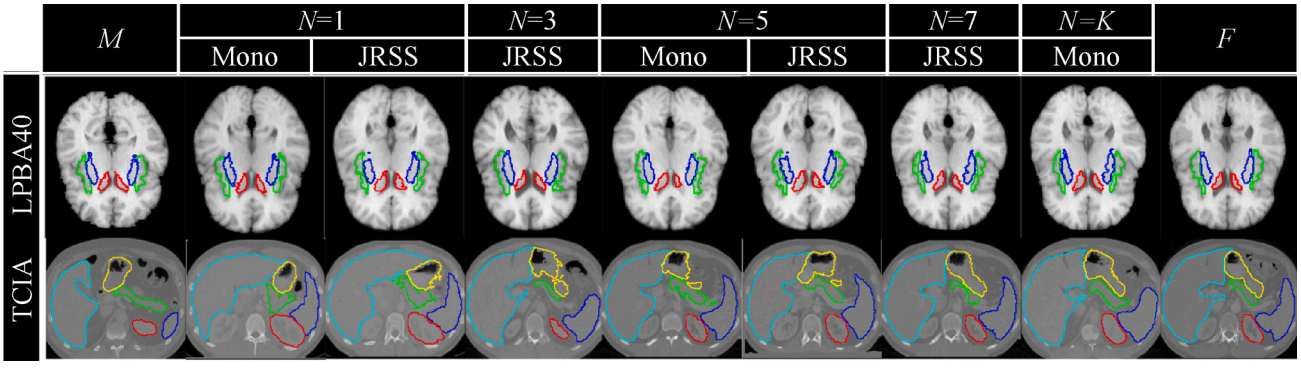



**Fig. 9.** Examples of registration results of JRSS with different numbers of labeled training samples (N).

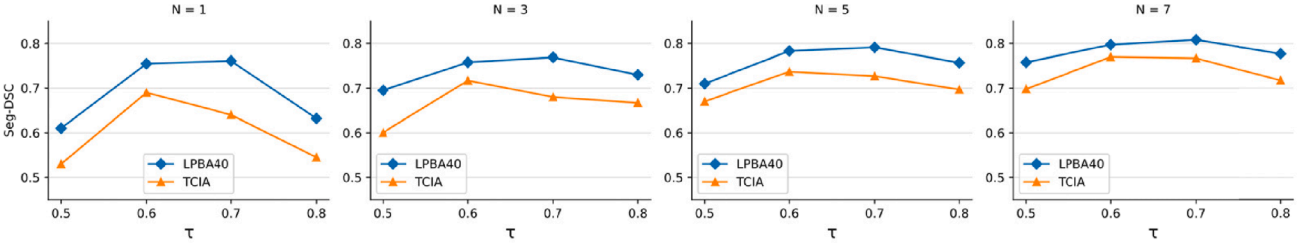



**Fig. 10.** The line graphs show the segmentation accuracy of JRSS on the test set under different few-shot scenarios and pseudo-label threshold τ.

pseudo-labels [37]. We find that, for some drastic and small structures in TCIA abdominal images, they will be ignored by the higher threshold, and more unlabeled data will not be able to obtain pseudo-labels; while some segmentation structures such as liver and spleen can still preserve precision even after lowering the threshold. In fact, we also get good results when we choose$\tau = 0.6$ as the common threshold for both datasets in our experiments. Therefore, the threshold setting gives priority to organs that are difficult to segment.

## 5. Discussion and conclusion

In this paper, we propose a joint registration and segmentation self-training framework, JRSS, to improve the dual-task performance of registration and segmentation of medical images in scenarios with few manual annotations. We ensure that the segmentation network is iteratively optimized during self-training through noise injection and uncertainty correction for pseudo-labels. The segmentation network trained from coarse to fine predicts more qualified pseudo-labels for unlabeled data, so unlabeled data and its pseudo-labels can be added as weakly supervised registration constraints based on unsupervised registration learning. The data warping provided by registration injects input noise into segmentation self-training and provides more reasonable data augmentation.

Based on analysis of all experimental results, we can find that our JRSS framework for joint registration and segmentation self-training improves dual tasks performance on separately learned single-task networks, and outperforms other JRS methods without pseudo-label screening and coarse-to-fine correction. Especially when dealing with large abdominal deformations and organs with blurred boundaries, registration is often more difficult, and segmentation can still maintain high performance (see Fig. 4 (c) and (d)). It can be seen that registration benefits from the low cost of segmentation. JRSS achieves state-of-the-art accuracy and generality when trained on datasets that only rely on few human segmentation labels (few-shot). We demonstrate that a self-training framework for joint registration and segmentation can complement the dual tasks in few-shot scenarios.

Despite the very promising results, there are many potential extensions to further improve our method. In future work, the pseudo-label threshold $\tau$ and the deformation field regularization scalar $\lambda_r$ can be used as learnable parameters, conditional image registration method [48] combined with the self-supervised learning paradigm can be used to search for the optimal solution in an automated manner. Multimodal dual-task learning can also be investigated by combining joint self-training with adversarial learning domain adaptation [33], which may further reduce the reliance on labeled data for model training.

## CRediT authorship contribution statement

**Huabang Shi:** Conceptualization, Methodology, Software, Validation, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Liyun Lu:** Validation, Visualization, Investigation, Writing – original draft, Writing – review & editing. **Mengxiao Yin:** Project administration, Resources. **Cheng Zhong:** Supervision, Resources, Funding acquisition. **Feng Yang:** Conceptualization, Writing – review & editing, Supervision, Project administration, Resources, Funding acquisition.

## 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.

## Data availability

Data will be made available on request.

## Acknowledgement

This work was partly supported by the National Natural Science Foundation of China (NSFC) (Grant No.: 61861004, 61862006).

## References

[1] G. Litjens, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[2] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J.V. Guttag, A.V. Dalca, VoxelMorph: a learning framework for deformable medical image registration, IEEE Trans. Med. Imag. 38 (8) (2019) 1788–1800.
[3] W. Walton, S. -J. Kim, L. Mullen, Automated registration for dual-view X-ray mammography using convolutional neural networks, IEEE Trans. Biomed. Eng., doi: 10.1109/TBME.2022.3173182.
[4] T. Estienne, et al., U-ReSNet: Ultimate coupling of registration and segmentation with deep nets, in: Proc. Int. Conf. Med. Image Comput. Comput.Assist. Intervention, 2019, pp. 310-319.
[5] T. Estienne, et al., Deep learning-based concurrent brain registration and tumor segmentation, Frontiers Comput Neurosci. 14 (3) (2020) pp.
[6] L. Mansilla, D.H. Milone, E. Ferrante, Learning deformable registration of medical images with anatomical constraints, Neural Networks 124 (2020) 269–279.
[7] Z. Xu, M. Niethammer, DeepAtlas: Joint semi-supervised learning of image registration and segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2019, pp. 420-429.
[8] L. Qiu, H. Ren, RSegNet: A joint learning framework for deformable registration and segmentation, IEEE Trans. Automat. Sci. Eng. 19 (3) (July 2022) 2499–2513, https://doi.org/10.1109/TASE.2021.3087868.
[9] A. Zhao, G. Balakrishnan, F. Durand, J.V. Guttag, A.V. Dalca, Data augmentation using learned transformations for one-shot medical image segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 8543–8553.
[10] A. Bitarafan, M. Nikdan, M.S. Baghshah, 3D image segmentation with sparse annotation by self-training and internal registration, IEEE J. Biomed. Health Inform. 25 (7) (2021) 2665–2672.
[11] Q. Xie, M. -T. Luong, E. Hovy, Q.V. Le, Self-training with noisy student improves imagenet classification, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2020, pp. 10684-10695.
[12] Z. Feng, Q. Zhou, G. Cheng, X. Tan, J. Shi, L. Ma, Semi-supervised semantic segmentation via dynamic self-training and class-balanced curriculum, 2020, arXiv: 2004.08514.
[13] Y. Ge, D. Chen, H. Li, Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification, 2020, arXiv: 2001.01526.
[14] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, in: Proc. Conf. Advances Neural Information Processing Systems, 2017, pp. 1195-1204.
[15] H.R. Roth, A. Farag, E.B. Turkbey, L. Lu, J. Liu, R.M. Summers, Data from pancreas-ct. the cancer imaging archive, 2016. https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU.
[16] D.W. Shattuck, et al., Construction of a 3d probabilistic atlas of human cortical structures, Neuroimage 39 (3) (2008) 1064–1080.
[17] H.R. Roth, et al., Deeporgan: multi-level deep convolutional 795 networks for automated pancreas segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, Springer, 2015, pp. 556–564.
[18] E. Gibson, et al., Automatic multiorgan segmentation on abdominal ct with dense v-networks, IEEE Trans. Med. Imag. 37 (8) (2018) 1822–1834.
[19] K.W. Clark, et al., The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository, J. Digit. Imaging 26 (6) (2013) 1045–1057.
[20] T. Estienne, et al., Deep learning based registration using spatial gradients and noisy segmentation labels, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2020, pp. 87–93.
[21] L. Yu, S. Wang, X. Li, C. Fu, P. Heng, Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2019, pp. 605-613.
[22] H. Zheng, Y. Zhang, L. Yang, C. Wang, D. Z. Chen, An annotation sparsification strategy for 3D medical image segmentation via representative selection and self-training, in: Proc. Int. Conf. Association Advancement Artificial Intelligence, 2020, 6925–6932.
[23] Y. Zhang, et al., 2017, Deep adversarial networks for biomedical image segmentation utilizing unannotated images, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2017, pp. 408–416.
[24] I. Radosavovic, P. Dollár, R. B. Girshick, G. Gkioxari, K. He, "Data distillation: towards omni-supervised learning, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2018, pp. 4119–4128.
[25] Z. Eaton-Rosen, F. Bragman, S. Ourselin, M.J. Cardoso, Improving data augmentation for medical image segmentation, in: Proc. Int. Conf. Medical Imaging Deep Learning, 2018.
[26] H. Yi, et al., Weakly-supervised convolutional neural networks for multimodal image registration, Med. Image Anal. 49 (2018) 1–13.
[27] M.P. Heinrich, Closing the gap between deep and conventional image registration using probabilistic dense displacement networks, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2019, pp. 50-58.

[28] W. Bai et al., Semi-supervised learning for network-based cardiac MR image segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2017, pp. 253–260.
[29] Y. Zhou, et al., Semi-supervised multi-organ segmentation via multi-planar co-training, 2018, arXiv:1804.02586.
[30] M. Elmahdy, J. Wolterink, H. Sokooti, I.Ғ sgum, M. Staring, Adversarial optimization for joint registration and segmentation in prostate ct radiotherapy, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2019, pp. 366-374.
[31] D. Mahapatra, Z. Ge, S. Sedai, R. Chakravorty, Joint registration and segmentation of X-ray images using generative adversarial networks, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2018, pp. 73–80.
[32] Z. Zhang, J. Li, Z. Zhong, Z. Jiao, X. Gao, A sparse annotation strategy based on attention-guided active learning for 3D medical image segmentation, 2019, arXiv: 1906.07367.
[33] H. Guan, M. Liu, Domain adaptation for medical image analysis: a survey, IEEE Trans. Biomed. Eng. 69 (3) (2022) 1173–1185.
[34] Y. He, et al., Deep Complementary joint model for complex scene registration and few-shot segmentation on medical images, in: Proc. European Conf. Comput. Vis., 2020, pp. 770-786.
[35] B. Li, et al., A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2019, pp. 645–653.
[36] L. Beljaards, M.S. Elmahdy, F.J. Verbeek, M. Staring, A cross-stitch architecture for joint registration and segmentation in adaptive radiotherapy, in: Proc. Int. Conf. Medical Imaging Deep Learning, 2020, pp. 62-74.
[37] K. Sohn, et al., FixMatch: simplifying semi-supervised learning with consistency and confidence, in: Proc. Conf. Advances Neural Information Processing Systems, 2020, pp. 596-608.
[38] Z. Zheng, Y. Yang, Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation, Int. J. Comput. Vision 129 (4) (2021) 1106–1120.
[39] M. Jaderberg, et al., Spatial transformer networks, in: Proc. Conf. Advances Neural Information Processing Systems, 2015, pp. 2017–2025.
[40] M.P. Heinrich, O. Oktay, N. Bouteldja, OBELISK-Net: Fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions, Med. Image Anal. 54 (2019) 1–9.
[41] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2016, pp. 424-432.
[42] J. Wu, H. Fan, X. Zhang, S. Lin, Z. Li, Semi-supervised semantic segmentation via entropy minimization, in: IEEE Int. Conf. Multimedia Expo (ICME), 2021, pp. 1-6.
[43] B.B. Avants, N. Tustison, G. Song, Advanced normalization tools (ANTs), Insight j 2 (365) (2009) 1–35.
[44] G. Haskins, U. Kruger, P. Yan, Deep learning in medical image registration: a survey, Mach. Vis. Appl. 31 (1) (2020) 1–18.
[45] A. Paszke, et al., PyTorch: an imperative style, high-performance deep learning library, in: Proc. Conf. Advances Neural Information Processing Systems, 2019, pp. 8024-8035.
[46] M. Rahimpour et al., Cross-modal distillation to improve MRI-based brain tumor segmentation with missing MRI sequences, in: IEEE Trans. Biomed. Eng., doi: 10.1109/TBME.2021.3137561.
[47] 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.
[48] T.C.W. Mok, A.C.S. Chung, Conditional deformable image registration with convolutional neural network, in Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention, 2021, pp. 35-45.
[49] Y. He, et al., Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching, IEEE J. Biomed. Health Inform. 26 (3) (2022) 1177–1187, https://doi.org/10.1109/JBHI.2021.3095409.