Deep Complementary Joint Model for Complex Scene Registration and Few-shot Segmentation on Medical Images

Yuting He, Tiantian Li, Guanyu Yang, Youyong Kong, Yang Chen, Huazhong Shu, Jean-Louis Coatrieux, Jean-Louis Dillenseger, Shuo Li

Deep Complementary Joint Model for Complex Scene Registration and Few-shot Segmentation on Medical Images

Yuting He[1[0000−0003−0878−8915]], Tiantian Li[1], Guanyu Yang[1,2(✉)[0000−0003−3704−1722]], Youyong Kong[1,2], Yang Chen[1,2], Huazhong Shu[1,2], Jean-Louis Coatrieux[2,3], Jean-Louis Dillenseger[2,3], and Shuo Li[4[0000−0002−5184−3230]]

[1] Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, China
yang.list@seu.edu.cn (G. Yang(✉))
[2] Centre de Recherche en Information Biomédicale Sino-Français (CRIBs)
[3] Univ Rennes, Inserm, LTSI - UMR1099, Rennes, F-35000, France
[4] Dept. of Medical Biophysics, University of Western Ontario, London, ON, Canada
slishuo@gmail.com (S. Li)

Abstract. Deep learning-based medical image registration and segmentation joint models utilize the complementarity (augmentation data or weakly supervised data from registration, region constraints from segmentation) to bring mutual improvement in complex scene and few-shot situation. However, further adoption of the joint models are hindered: 1) the diversity of augmentation data is reduced limiting the further enhancement of segmentation, 2) misaligned regions in weakly supervised data disturb the training process, 3) lack of label-based region constraints in few-shot situation limits the registration performance. We propose a novel Deep Complementary Joint Model (DeepRS) for complex scene registration and few-shot segmentation. We embed a perturbation factor in the registration to increase the activity of deformation thus maintaining the augmentation data diversity. We take a pixel-wise discriminator to extract alignment confidence maps which highlight aligned regions in weakly supervised data so the misaligned regions' disturbance will be suppressed via weighting. The outputs from segmentation model are utilized to implement deep-based region constraints thus relieving the label requirements and bringing fine registration. Extensive experiments on the CT dataset of MM-WHS 2017 Challenge[42] show great advantages of our DeepRS that outperforms the existing state-of-the-art models.

1 Introduction

Deep learning-based medical image segmentation models and registration models [23, 11, 19] are limited in complex scene and few-shot situation. In complex scene which has complex but task-unconcerned backgrounds, the unsupervised registration models [2, 7] pay equal attention to all regions for overall alignment

so that the performance on regions of interest (ROIs) will be limited by background. In few-shot situation which lacks labels, the segmentation models [29, 24] will over-fit [40, 30] due to the lack of supervision information.

The registration and segmentation tasks has great complementarity which will bring mutual improvement in complex scene and few-shot situation. As shown in Fig. 1, the registration model provides diverse augmentation data (warped images and labels) or weakly supervised data (fixed images and warped labels) for segmentation model [40, 37] during the training process, thus reducing the requirement of labels and enhancing the segmentation generalization in few-shot situation. The segmentation model feeds back region constraints [22, 37, 6] so that additional attention on ROIs is paid for finer registration in complex scene.

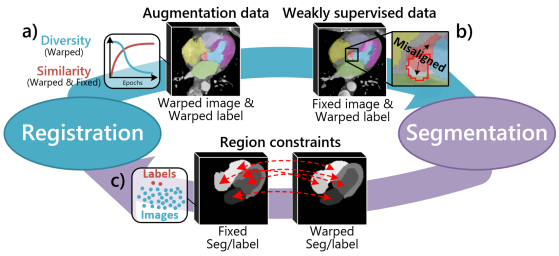

Fig. 1. The complementary topology and limitations of registration and segmentation tasks. Registration provides augmentation data and weakly supervised data for segmentation for higher generalization in few-shot situation, the segmentation feeds back region constraints for finer registration on ROIs in complex scene. a), b), c) illustrate the limitations in the utilization of this complementary topology.

Unfortunately, further exploiting of this complementary topology are hindered [40, 37, 6, 22] due to: **Limitation 1:** *Degradation of data augmentation capability* (Fig. 1(a)). During the training of registration model, it learns the deformation rule that matches real situation and generates diverse warped images as augmentation data to improve the segmentation generalization ability [37, 36]. However, the similarity between warped and fixed images increases and tends to become stable, and the diversity of warped images is gradually reduced as the similarity stabilizes. Therefore, in the later training stage of registration network, the identical warped images are generated in different epochs, resulting in the reduction of augmentation data diversity. Thus, the data augmentation ability of registration model is degraded and the further enhancement of segmentation will be limited. **Limitation 2:** *Misaligned regions in weakly supervised data* (Fig. 1(b)). The weakly supervised data enlarges the labeled dataset and provide additional supervision information for the segmentation model. However, large misaligned regions in these data will produce incorrect optimization targets and it will disturb the training process leading to serious mis-segmentation if used directly [37]. **Limitation 3:** *Lack of label-based region constraints* (Fig. 1(c)). Region constraints provide specific alignment information for regions bringing finer registration optimization. However, in few-shot situation, the label-based region constraints [37, 6, 22, 14] are lacked with few labels. Thus if in complex scene, the registration model [2, 36,

7] will take rough optimization and the complex backgrounds will limit the registration performance on ROIs.

Solution 1 for the degradation of data augmentation capability: we embed a random perturbation factor in the registration to increase the activity of deformation for sustainable data augmentation capability. The registration process is a displacement of structure information, and the adjustment of deformation degree is the sampling of the structure information on this displacement path [20, 12]. Therefore, our perturbation factor adjusts the deformation degree randomly to sample the structure information which is consistent with the real distribution to produce diverse and real augmentation data for the segmentation model.

Solution 2 to suppress the misaligned regions' disturbance: we extract alignment confidence maps from a pixel-wise discriminator to suppress the misaligned regions in weakly supervised data and utilize the supervision information in aligned regions. The pixel-wise discriminator, resulting in a generative adversarial network (GAN) [10] based registration model [7, 38, 8, 13], learns the similarity between warped and fixed images and outputs the alignment confidence maps that highlight the aligned regions [15, 26]. Thus, via these maps ,the misaligned regions will be suppressed and the supervision information in aligned regions will be utilized for higher segmentation generalization when calculating the weakly supervised loss function.

Solution 3 to cope with the lack of label-based region constraints: we build deep-based region constraints that calculate the loss value via the warped and fixed segmentations from the segmentation model so that fine registration optimization targets are available. Therefore, 1) label requirements of label-based region constraints are freed in few-shot situation, 2) different regions are independently optimized to avoid the misalignment of each region and 3) region attention on the ROIs is paid for finer registration.

In this paper, we propose a *Deep Complementary Joint Model (DeepRS)* that minimizes background interference in complex scene for finer registration on ROIs, and greatly reduces the label requirements of segmentation in few-shot situation for higher generalization ability. In short, the contributions of our work are summarized as follows:

- To the best of our knowledge, we build a novel complementary topology of registration and segmentation for the first time, and propose the DeepRS model utilizing the data generation ability of registration for few-shot segmentation, and the label-free region constraint ability of segmentation for complex scene registration.
- We propose a deep structure sampling (DSS) block adding a random perturbation factor to the registration for sustainable data augmentation ability.
- We propose an alignment confidence map (ACM) method which efficiently utilizes the supervision information in weakly supervised data thus bringing powerful segmentation generalization.
- We propose a deep-based region constraint (DRC) strategy which frees up the label requirements of label-based methods achieving finer registration on ROIs.

2 Related Works

2.1 Registration and segmentation joint models

Registration and segmentation tasks have great complementarity, thus building a registration and segmentation joint model has the potential of mutual improvement. The registration provides augmentation data and weakly supervised data for the segmentation [40, 37, 35], and the segmentation feeds back additional region constraints [14, 22, 6]. Zhao *et al.* [40] took a pre-trained registration model to generate augmentation data for more powerful segmentation ability. Li *et al.* [22] made a hybrid framework that took the label-based region constraints from labels and segmentations for finer registration. Similarity, Xu *et al.* [37] designed a semi-supervised method that combined registration and segmentation models bringing the mutual improvement in knee and brain images.

However, these existing methods only took the advantage of partial complementarity which hardly gives full play to their potential. The convergence of the registration model limits the diversity of the augmentation data and prevents further enhancement of the segmentation model [40]. Misaligned regions in weakly supervised data disturb the training of segmentation models, and if used directly, it will lead to serious mis-segmentation [37]. In few-shot situation, label-based region constraints are lacked due to the small labeled dataset [22], thus with inaccurate optimization targets, complex backgrounds will limit registration performance on ROIs in complex scene.

2.2 Data augmentation

Data augmentation [30], generating bigger dataset, has the ability to improve learning models[31], especially in few-shot situation. Some data augmentation strategies (random cropping, mirroring, rotation, flipping, etc.) are often used for higher generation ability, while inappropriate strategy combinations will generate unreasonable data which will weaken the model performance [30]. Learning-based data augmentation strategies [4, 27, 16, 21] learn the augmentation methods from dataset for real augmentation data. Registration learns transformation rules of structure information from the images [40, 20, 12, 37] so that the augmentation images with real structure information are obtained.

Disappointingly, the registration-based augmentation ability will degrade due to the reduction of deformation diversity. As the registration model converges, the moving image is stably aligned onto the fixed image and the identical warped images in different epochs are generated, resulting in the reduction of augmentation data diversity and limiting the further improvement of segmentation.

2.3 Weakly-supervised learning

Weakly-supervised learning [14, 33, 18, 28, 15, 26] utilizing non-precisely labeled data is a strategy for labeled data limitation. It has three typical types according to the weakly supervised data types [41]: 1) incomplete supervision

where part of the dataset without labels[26, 15], 2) inexact supervision where data with coarse-grained labels [14, 18] and 3) inaccurate supervision where data with inaccurate labels [33, 32]. In registration and segmentation tasks, the warped labels and fixed images from registration model make up weakly supervised data leading to inaccurate supervision which will improve the segmentation performance with appropriate strategy. Unfortunately, if the weakly supervised data is used directly, the misaligned regions will brings inaccurate optimization target, thus disturbing the training process and lead to mis-segmentation.

2.4 Generative Adversarial Networks

Generative adversarial networks (GANs) [10, 39, 9], consisting of a generator G and a discriminator D, learns a similarity metric of the generated and real images. The discriminator learns to distinguish the real or generated images and the generator takes the adversarial loss from the discriminator to improve the authenticity of the generated image to deceive the discriminator. GAN-based registration models [7, 38, 8, 13] take global discriminator to learn image-wise similarity metric of warped and fixed images which can be used to evaluate the weakly supervised data in our task.

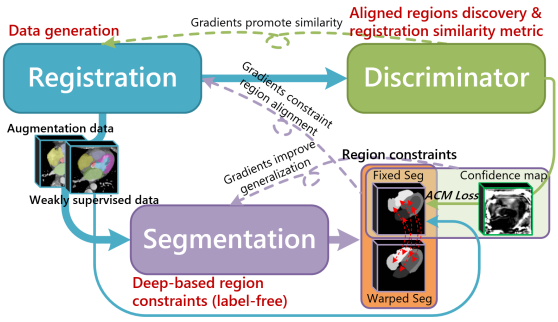

Fig. 2. The overview of our DeepRS. The data generation ability of the registration, the deep-based region constraint of the segmentation, the aligned regions discovery ability and the learned similarity metric of discriminator interact in the alternating training process.

However, the image-wise similarity has no ability to evaluate the regional similarity and in our segmentation task (pixel-wise), it will still introduce the error information in the weakly supervised data. Patch-GANs utilize pixel-wise discriminator [26, 15] consisting of a full convolution network to learn the pixel-wise similarity and output confidence maps which highlight task-beneficial regions. Thus, a patch-GAN is used in our model for alignment confidence maps to suppress the misaligned regions and utilize the supervision information in weakly supervised data.

3 Methodology

Our DeepRS model (Fig. 3, Fig. 2), which consists of registration, pixel-wise discriminator and segmentation models, leverages their complementarity for com-

plex scene registration and few-shot segmentation (Sec. 3.1) bringing mutual improvement. The registration generates diverse augmentation data via randomly adjusting the deformation field in a DSS block (Sec. 3.1) and provides weakly supervised data for the segmentation network to reduce the labeled data requirements in few-shot situation. The pixel-wise discriminator provides ACMs (Sec. 3.1) for the segmentation network for supervision information utilization in weakly supervised data. The segmentation network provides DRC (Sec. 3.1) for the registration network for finer registration on ROIs in complex scene. The joint strategy (Sec. 3.2) maximizes the complementarity via alternating training.

3.1 DeepRS for stronger registration and segmentation

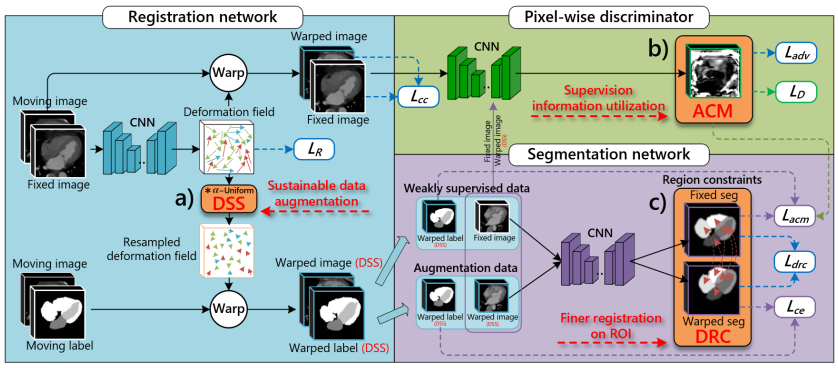

Fig. 3. In detail of our DeepRS model, we design a DSS block, a ACM method and a DRC strategy cleverly dealing with the limitations. a) The DSS block maintains the diversity of warped images bringing sustainable data augmentation ability. b) The ACM method utilizes the supervision information in weakly supervised data. c) The DRC strategy provides region attention on ROIs for finer registration.

The proposed DeepRS model leverages the complementarity of registration and segmentation tasks via the DSS block, ACM method and DRC strategy.

Deep structure sampling (DSS) for sustainable data augmentation DSS block generates diverse augmentation data sustainedly via embedding a random perturbation factor in the deformation field to increase the uncertainty of the warped images and labels. The registration process is the displacement of image structure information, and the perturbation of deformation degree realizes the sampling of information on this displacement path [20, 12]. Therefore, the DSS block brings two advantages: 1) Sustainable data augmentation. The perturbation factor controls the deformation degree so that the registration network is guaranteed to generate diverse augmentation data sustainedly. 2) Real

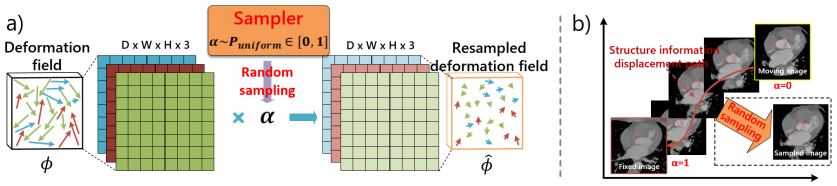

Fig. 4. The DSS block for sustainable data augmentation. a) A perturbation factor $\alpha \in [0, 1]$ from uniform distribution adjusts the deformation field making the sampling process. b) Illustration of the sampling that registration makes the structure information displacement and our DSS samples the information on its displacement path.

distribution. Sampling structure information from its displacement path generates the augmentation data more matching real distribution than other manual augmentation methods.

As shown in Fig. 4(a), a deformation field $\varnothing$ from registration network is multiplied by a random perturbation factor α from uniform distribution $p_{uniform} \in [0, 1]$ to obtain an resampled deformation field $\hat{\varnothing} = \varnothing \times \alpha \backsim p_{uniform} \in [0, 1]$. Therefore, the warped images and labels deformed by it will still have great diversity, even if the registration network has converged. Fig. 4(b) illustrates that as α increases, the warped images gradually approximate the fixed images since its structure information approaches the fixed image. It is evident that the randomly sampled deformations are non-rigid, yet produce realistically-looking images.

Alignment confidence map (ACM) for supervision information utilization ACM method utilizes the supervision information of aligned regions and suppresses the misaligned regions in weakly supervised data to improve the segmentation generalization ability. The ACM maps from the pixel-wise discriminator evaluate the pixel-wise similarity between warped and fixed images and will highlight the aligned regions. Thus, these maps will be taken to weight the loss of weakly supervised data to utilize its supervision information in aligned regions, as illustrated in Equ. 1:

$$\mathcal{L}_{acm} = -D(W(x_m, \hat{\varnothing}), x_f) W(y_m, \hat{\varnothing}) \log S(x_f) \tag{1}$$

where x_m, y_m, x_f and $\hat{\varnothing}$ are the moving image, moving label, fixed image and resampled deformation field from DSS block. As shown in Fig. 3, $W(\cdot, \cdot)$ is the 'warp' block which deforms the moving images and labels to the fixed images for warped images and labels following the spatial transformation layer in [2]. The pixel-wise discriminator $D(\cdot, \cdot)$ measures the similarity between warped and fixed images for the ACMs to weight the cross-entropy loss between warped labels and fixed segmentation (seg-) masks $S(x_f)$. Therefore, the loss value in misaligned region will get low weight and the disturbance will be suppressed.

The contribution of the weakly supervised data is increasing during the training. In early training stage, the powerful discriminator outputs weak maps, so that the loss from weakly supervised data is suppressed greatly and the optimization target of the segmentation network is dominated by the loss $\mathcal{L}_{ce}$ from augmentation data. As the training progresses, the registration network defeats the discriminator and obtains high responsive maps, thus increasing the contribution of ACM loss $\mathcal{L}_{acm}$, so that the segmentation generalization ability will be further enhanced.

Deep-based region constraint (DRC) for finer registration on ROIs DRC strategy guides the attention on the ROIs for finer registration via constraints between the fixed and warped seg-masks from the segmentation network. This deep-based region constraint takes the alignment of the corresponding regions in warped and fixed images as the optimization target, so that 1) label requirements of label-based region constraints is freed in few-shot situation, 2) different regions are independently optimized to avoid the misalignment between each other and 3) additional region attention on the ROIs is paid for finer registration.

As shown in Fig. 3(c), the warped image and the fixed image are input into the segmentation network respectively for the warped and the fixed seg-masks firstly. Then a mean square error loss between these two seg-masks is calculated as is illustrated in Equ. 2:

$$\mathcal{L}_{drc} = -(S(W(x_m, \hat{\varnothing})) - S(x_f))^2 \tag{2}$$

where x_m, x_f and $\hat{\varnothing}_n$ are the moving image, fixed image and deformation field from the DSS block. $W(\cdot, \cdot)$ is the deformation process in registration network and $S(\cdot)$ is the segmentation network. Each ROI is calculated in different channels obtaining independent fine optimization, while the task unconcerned regions are calculated in a background channel together. Thus, fine registration on ROIs is available and inter-regional error registration is avoided.

3.2 Joint learning strategy exerts complementarity

The registration network, segmentation network and pixel-wise discriminator in our DeepRS model (Fig. 3) are trained by different loss function combinations to coordinate the training process and achieve mutual improvement.

Registration network The registration network is optimized by four different targets. An adversarial loss $\mathcal{L}_{adv}$ [7] from the pixel-wise discriminator provides the similarity metric between warped and fixed images. The DRC loss $\mathcal{L}_{drc}$ from the segmentation network brings registration attention on ROIs. A local cross-correlation (CC) [2] $\mathcal{L}_{cc}$ maintains the stability of the training process, and a smooth loss [2] $\mathcal{L}_R$ penalizes local spatial variations in deformation field. Therefore, the total loss function $\mathcal{L}_{reg}$ is:

$$\mathcal{L}_{reg} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{drc}\mathcal{L}_{drc} + \lambda_{cc}\mathcal{L}_{cc} + \lambda_R\mathcal{L}_R \tag{3}$$

Segmentation network The loss function of the segmentation network $\mathcal{L}_{seg}$ consists of two components. One is the ACM loss $\mathcal{L}_{acm}$ that adds the weakly supervised data to the training for higher segmentation generalization ability. The other is cross-entropy loss $\mathcal{L}_{ce}$ between the warped images and labels that maintains the right optimization target:

$$\mathcal{L}_{seg} = \lambda_{acm}\mathcal{L}_{acm} + \lambda_{ce}\mathcal{L}_{ce} \tag{4}$$

Pixel-wise discriminator The training strategy of pixel-wise discriminator follows [7]: well-registered image pairs consisting of reference images x_r and fixed images x_f as positive cases and misaligned images consisting of warped images x_w and fixed images x_f as negative cases. The reference image x_r is a fusion of a moving image x_m and a fixed image x_f according to the formula $x_r = \beta * x_m + (1 - \beta) * x_f$. Thus, the loss for the discriminator $\mathcal{L}_D$ is:

$$\mathcal{L}_D = -\log(D(x_r, x_f)) - \log(1 - D(x_w, x_f)) \tag{5}$$

4 Experiments

Extensive experimental results show that our DeepRS model enhances the performance of complex scene registration and few-shot segmentation tasks on cardiac CT data which has complex task-unconcerned backgrounds.

4.1 Evaluation settings

Dataset We validated the superiority of our DeepRS model on the whole heart registration and segmentation tasks on the CT dataset of *MM-WHS 2017 Challenge* [42] which has complex backgrounds (lung, rib cage, etc.). This dataset consists of 20 labeled and 40 unlabeled CT images. Our experiments aim to register and segment seven cardiac structures including the ascending aorta, left atrial cavity (LA), left ventricular cavity(LV), myocardium of the left ventricle (Myo), pulmonary artery (PA), right atrial cavity (RA) and right ventricular cavity (RV). We first crop the rectangular regions containing the hearts for affine transformation and resample them to $128 \times 128 \times 96$. Then the labeled images are randomly split into 5 parts, 1 part (4 images) is used in training set as the moving images for few-shot situation and the remaining 4 parts (16 images) in testing set resulting in 5-folds evaluation. We put 40 unlabeled images as the fixed images in the training set leading to 160 data pairs together with the moving images, and the 16 images in the testing set are paired separately leading to 240 data pairs. Following the [2], we use Elastix[5] to perform affine transformation so that our model only needs to pay attention to the deformation registration process.

[5] https://www.elastix.org/

Implementation The segmentation network and pixel-wise discriminator follow the same 3D U-Net [3] structure. The registration network follows the VoxelMorph-2 [2] structure. We use RMSprop [34] to train the registration network and the discriminator for stable process [1], and Adam [17] to train the segmentation network for fast convergence. These models share the same learning rate of $2e^{-4}$ and training batch size of 1 due to the limitation of memory. According to extensive experiments, we finally set $\lambda_{adv} = 1$, $\lambda_{drc} = 10$, $\lambda_{cc} = 1$, $\lambda_R = 1$, $\lambda_{acm} = 1$ and $\lambda_{ce} = 1$. The models were implemented via Keras[6] with a Tensorlow[7] backend and were trained on a single NVIDIA TitanX GPU with 12 GB memory.

Comparison settings The comparison demonstrates the advancement on segmentation and registration of our DeepRS model. We compare our model's segmentation performance with three general segmentation networks (3D U-Net[3], V-Net[25], 3D FCN[24]) to illustrate the enhancement brought by registration. The 3D U-Net augmented by manual strategies (random rotate in $[-10°, 10°]$, random mirroring and random flipping) is compared with to show the advantages of registration-based data augmentation. We also compare two unsupervised registration models (VoxelMorph-2[2], Adv-Reg[7]) to illustrate the superiority of our deep-based region constraints in complex scene. In addition, two registration and segmentation joint models (DeepAtlas[37], HybridCNN[22]) are compared with to demonstrate the superiority of the DeepRS brought by our DSS block, ACM method and DRC strategy. What's more, our proposed DeepRS is also evaluated on different data amount to illustrate its excellent generalization ability in few-shot segmentation. Finally, an ablation study is used to analyse the contributions of each our innovation.

Evaluation metric We evaluate the registration and segmentation methods with dice coefficient [5]. The dice coefficient ([%]) is a metric that measures the coincidence degree between two sets according to $Dice(G, P) = \frac{2|G \bigcap P|}{|G|+|P|}$ where the G is the ground truth and the P is the predicted mask. It is suitable to evaluate the agreement between the predicted segmentation/registration and the ground truth. The Dice coefficients of the corresponding seven cardiac structures are calculated, and presented as $mean \pm std$.

4.2 Results

Extensive experimental results on cardiac CT dataset show that with merely 4 training labels, our proposed DeepRS appears to be a strong superiority both in the quantitative comparison and in visual. The experiments on different label amounts illustrate that our DeepRS greatly reduces the label dependence of the segmentation model.

[6] https://github.com/keras-team/keras
[7] https://github.com/tensorflow/tensorflow

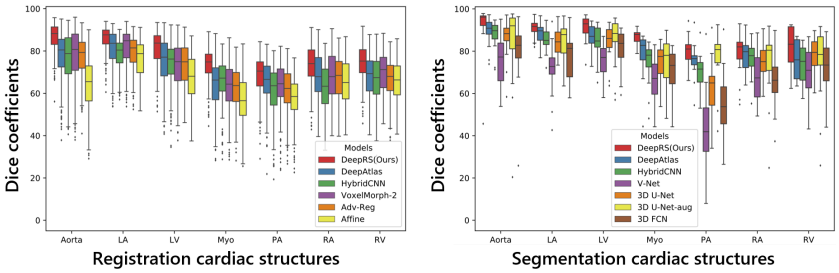

Fig. 5. Our DeepRS achieves excellent dice coefficients on each structure. The box plots shows the proposed DeepRS (red box) model achieves the state-of-the-art performance in complex scene registration (*Left*) and few-shot segmentation (*Right*).

Quantitative comparison As shown in Tab. 1, our DeepRS model achieves the state-of-the-art performance in both registration and segmentation tasks whose mean dice coefficients of all cardiac structures are 77.6% and 85.7%. Fig. 5 illustrates that the proposed DeepRS achieves excellent dice coefficients on each structure in complex scene registration and few-shot segmentation.

On the registration task, the registration network gets the deep-based region constraints from the segmentation network, bringing finer registration on ROIs in complex scene than other registration models. VoxelMorph-2 lacks region constraints, thus the dice is 5.9% lower than ours. The Adv-Reg takes a GAN whose training process is unstable to learn a similarity metric and gets worse results than VoxelMorph-2. DeepAtlas utilizes weakly supervised data directly, thus the misaligned regions disturbs the training process of the segmentation model finally in turn disturbing the registration performance (71.3%). HybridCNN lacks label-based region constraints in our few-shot situation thus the influence of the misaligned regions are more pronounced (69.2%).

Table 1. The proposed DeepRS model achieves the state-of-the-art performance both in registration (R) and segmentation (S) tasks on cardiac CT data.

Method	R-Dice	S-Dice
Affine only	64.6±10.7	-
VoxelMorph-2[2]	71.7±10.6	-
Adv-Reg[7]	68.8±10.7	-
3D U-Net[3]	-	78.8±9.2
3D U-Net-aug[3]	-	80.0±12.0
3D FCN[24]	-	71.4±11.3
V-Net[25]	-	69.8±10.9
DeepAtlas[37]	71.3±10.5	81.8±7.5
HybridCNN[22]	69.2±10.3	78.8±7.9
DeepRS(Ours)	**77.6±7.9**	**85.7±7.7**

On the segmentation task, the segmentation network in our DeepRS model effectively utilizes the augmentation data and weakly supervised data from the registration network, thus achieving much higher dice coefficient than 3D U-Net,

3D FCN and V-Net. Although the 3D U-Net augmented by manual strategies has get 1.2% dice improvement compared with non-augmentation, our DeepRS model has even greater advantage by 5.7%. Due to the influence of the misaligned regions in weakly supervised data, the HybridCNN only gets 78.8% dice. Similarly, the DeepAtlas takes the augmentation data from registration, but the misaligned regions still limits the enhancement which make it get only 81.8% dice.

Visual superiority Visually, our DeepRS model brings higher segmentation generalization ability with few labels, and achieves finer registration performance on ROIs in complex scene.

As illustrated in Fig. 6, our DeepRS brings finer registration on ROIs making 5 structures in moving image look more similar to these structures in fixed image. The HybridCNN uses weakly supervised data directly and lacks label-based region constraints. Therefore the misaligned regions interrupt the segmentation training process and in turn weaken the registration performance bringing serious region correspondence errors. The Adv-Reg is optimized by the unstable GAN making the warped image messy and rough in detail.

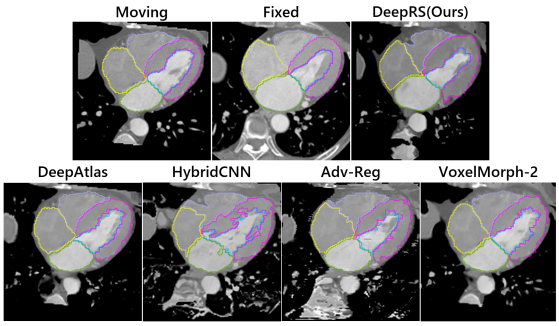

Fig. 6. Our DeepRS gets finer registration on ROIs. The example slices from 3D CT image show the overlaid boundaries of the LV (green), RA (yellow), RV (purple), LV (blue) and Myo (pink). Our model makes these structures in moving image alike structures in fixed image.

As shown in Fig. 7, our DeepRS model brings much higher segmentation generalization ability trained on merely 4 labeled images. Case 1 shows the excellent generalization ability and the yellow boxes show the performance in detail. Our DeepRS has achieved fine segmentation, while the 3D U-Net, 3D FCN, 3D U-Net-aug and V-Net have many mis-segmentation regions. The HybridCNN and DeepAtlas has more mis-segmentation regions than others due to the misaligned regions in weakly supervised data. Case 2 shows the fine segmentation capability in another perspective and sample. The 3D U-Net ,3D FCN and V-Net are limited by small dataset leading to various serious mis-segmentation.

DeepRS for few-shot segmentation In few-shot situation, the segmentation (S) network in our DeepRS model achieves higher mean dice coefficients of all structures than 3D U-Net as illustrated in Fig. 8. The effectiveness of our DeepRS

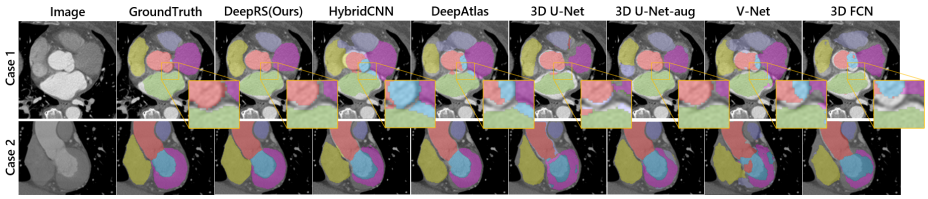

Fig. 7. Our DeepRS brings higher segmentation generalization ability trained on 4 labeled images. Yellow boxes show the excellent generalization ability in detail. The example slices from 3D CT image show the regions of Aorta (red), RA (yellow), RV (purple), Myo (pink), LV (green) and LV (blue).

is evaluated on randomly-sampled labeled data whose amount is 1, 4, 7 and 10 respectively. 3D U-Net is used for comparison and the mean dice coefficients of all structures are calculated. As the labeled data decreases, the superiority of our DeepRS on segmentation task becomes more prominent. When only one label is available, our segmentation performance is 18.1% higher than 3D U-Net.

4.3 Ablation study

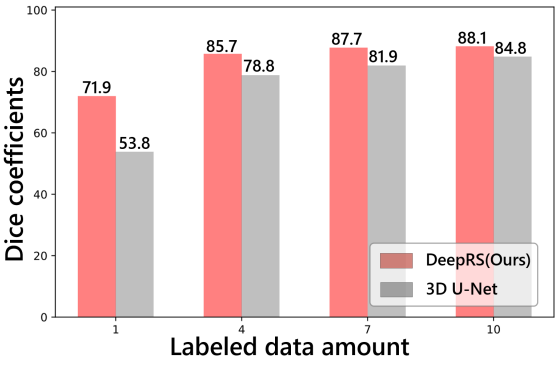

Fig. 8. Especially in few-shot situation, the segmentation network in our DeepRS model achieves much higher mean dice coefficients of all structures than 3D U-Net[3].

As shown in Tab. 2, an ablation study illustrates each great advantage brought by our innovations. The directly joint model only utilizes the registration's data augmentation ability thus the segmentation gets 80.5% dice and the registration gets 72.9% dice. Our DSS block embeds a random perturbation factor in the registration to maintain the diversity of augmentation data (**Solution 1**), thus bringing 3.4% segmentation dice growth. The ACM method adds the supervision information in weakly supervised data (**Solution 2**) to segmentation network so that it gets 3.6% segmentation dice improvement. The DRC strategy builds deep-based region constraints instead of label-based methods (**Solution 3**) via the warped and fixed segmentations increasing the direct joint model by 3% reg-

istration dice. We find that the segmentation and registration models achieve further promotion in our final DeepRS model owing to their complementarity, thus finally achieving 77.6% registration dice and 85.7% segmentation dice which are increased by 4.7% and 5.2% respectively.

5 Conclusion

This paper presents a *Deep Complementary Joint Model(DeepRS)* for complex scene registration and few-shot segmentation. Our proposed *DSS block* adjusts deformation fields randomly via a perturbation factor, thus increasing the activity of the warped images and labels and achieving sustainable data augmentation capability. Our proposed *ACM method* efficiently utilizes the supervision information in weakly supervised data via alignment confidence maps from a pixel-wise discriminator bringing higher segmentation generalization. Our proposed *DRC strategy* constructs label-free loss between the warp and fixed images from the segmentation model resulting in finer registration on ROIs. We train our proposed DeepRS model on the cardiac CT dataset which has complex background with few labels with merely 4 labels and shows great advantages in registration and segmentation tasks compared to existing methods.

Table 2. The ablation study analyses the contributions of our innovations.

R	S	DSS	ACM	DRC	R-Dice	S-Dice
✓					72.2±10.3	-
	✓				-	78.8±9.2
✓	✓				72.9±10.4	80.5±10.2
✓	✓	✓			72.9±9.6	83.9±8.3
✓	✓		✓		72.5±10.1	84.1±8.3
✓	✓			✓	75.9±9.1	82.5±9.2
✓	✓	✓	✓	✓	**77.6±7.9**	**85.7±7.7**

Our work greatly reduces the requirement of a large labeled dataset and provides the fine optimization targets, thus the registration and segmentation accuracy are improved and the cost is greatly saved. Especially, our DeepRS model has great potential in some situations where the labeling is difficult, the scene is complex or the dataset is small.

Acknowledgments This research was supported by the National Natural Science Foundation under grants (61828101,31571001,31800825), the Short-Term Recruitment Program of Foreign Experts (WQ20163200398), and Southeast University-Nanjing Medical University Cooperative Research Project (2242019K3DN08). We thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations in this paper.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. arXiv preprint arXiv:1701.07875 (2017)

2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9252–9260 (2018)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
4. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 113–123 (2019)
5. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
6. Estienne, T., Vakalopoulou, M., Christodoulidis, S., Battistela, E., Lerousseau, M., Carre, A., Klausner, G., Sun, R., Robert, C., Mougiakakou, S., et al.: U-resnet: Ultimate coupling of registration and segmentation with deep nets. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 310–319. Springer (2019)
7. Fan, J., Cao, X., Wang, Q., Yap, P.T., Shen, D.: Adversarial learning for mono-or multi-modal registration. Medical Image Analysis p. 101545 (2019)
8. Fan, J., Cao, X., Xue, Z., Yap, P.T., Shen, D.: Adversarial similarity network for evaluating image alignment in deep learning based registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 739–746. Springer (2018)
9. Ge, R., Yang, G., Xu, C., Chen, Y., Luo, L., Li, S.: Stereo-correlation and noise-distribution aware resvoxgan for dense slices reconstruction and noise reduction in thick low-dose ct. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 328–338. Springer (2019)
10. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
11. Haskins, G., Kruger, U., Yan, P.: Deep learning in medical image registration: A survey. arXiv preprint arXiv:1903.02026 (2019)
12. Hauberg, S., Freifeld, O., Larsen, A.B.L., Fisher, J., Hansen, L.: Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation. In: Artificial Intelligence and Statistics. pp. 342–350 (2016)
13. Hu, Y., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Vercauteren, T., Noble, J.A., Barratt, D.C.: Adversarial deformation regularization for training image registration neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 774–782. Springer (2018)
14. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. Medical image analysis **49**, 1–13 (2018)
15. Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H.: Adversarial learning for semi-supervised semantic segmentation. arXiv preprint arXiv:1802.07934 (2018)
16. Jackson, P.T., Atapour-Abarghouei, A., Bonner, S., Breckon, T., Obara, B.: Style augmentation: Data augmentation via style randomization. arXiv preprint arXiv:1809.05375 (2018)

17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
18. Kolesnikov, A., Lampert, C.H.: Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In: European Conference on Computer Vision. pp. 695–711. Springer (2016)
19. Lateef, F., Ruichek, Y.: Survey on semantic segmentation using deep learning techniques. Neurocomputing **338**, 321–348 (2019)
20. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2), 236–250 (2005)
21. Lemley, J., Bazrafkan, S., Corcoran, P.: Smart augmentation learning an optimal data augmentation strategy. Ieee Access **5**, 5858–5869 (2017)
22. Li, B., Niessen, W.J., Klein, S., de Groot, M., Ikram, M.A., Vernooij, M.W., Bron, E.E.: 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. pp. 645–653. Springer (2019)
23. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
24. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
25. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
26. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 370–378. Springer (2018)
27. Nielsen, C., Okoniewski, M.: Gan data augmentation through active learning inspired sample acquisition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 109–112 (2019)
28. Papandreou, G., Chen, L.C., Murphy, K.P., Yuille, A.L.: Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In: Proceedings of the IEEE international conference on computer vision. pp. 1742–1750 (2015)
29. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
30. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data **6**(1), 60 (2019)
31. Sun, C., Shrivastava, A., Singh, S., Gupta, A.: Revisiting unreasonable effectiveness of data in deep learning era. In: The IEEE International Conference on Computer Vision (ICCV) (Oct 2017)
32. Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., Schroers, C.: Normalized cut loss for weakly-supervised cnn segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1818–1827 (2018)
33. Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C., Boykov, Y.: On regularized losses for weakly-supervised cnn segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 507–522 (2018)

34. Tieleman, T., Hinton, G.: Lecture 6.5-rmsprop, coursera: Neural networks for machine learning. University of Toronto, Technical Report (2012)
35. Vakalopoulou, M., Chassagnon, G., Bus, N., Marini, R., Zacharaki, E.I., Revel, M.P., Paragios, N.: Atlasnet: Multi-atlas non-linear deep networks for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 658–666. Springer (2018)
36. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis **52**, 128–143 (2019)
37. Xu, Z., Niethammer, M.: Deepatlas: Joint semi-supervised learning of image registration and segmentation. arXiv preprint arXiv:1904.08465 (2019)
38. Yan, P., Xu, S., Rastinehad, A.R., Wood, B.J.: Adversarial image registration with application for mr and trus image fusion. In: International Workshop on Machine Learning in Medical Imaging. pp. 197–204. Springer (2018)
39. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical Image Analysis p. 101552 (2019)
40. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8543–8553 (2019)
41. Zhou, Z.H.: A brief introduction to weakly supervised learning. National Science Review **5**(1), 44–53 (2017)
42. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. Medical image analysis **31**, 77–87 (2016)