



# Progressively Coupling Network for Brain MRI Registration in Few-Shot Situation

Zuopeng Tan<sup>1</sup>, Hengyu Zhang<sup>1</sup>, Feng Tian<sup>2,3</sup>, Lihe Zhang<sup>1(✉)</sup>, Weibing Sun<sup>2,3</sup>,  
and Huchuan Lu<sup>1</sup>

<sup>1</sup> School of Information and Communication Engineering, Dalian University of Technology, Dalian, China

{zuopengtan,skysthelimits}@mail.dlut.edu.cn,  
{zhanglihe,lhchuan}@dlut.edu.cn

<sup>2</sup> Department of Urology, Affiliated Zhongshan Hospital of Dalian University, Dalian, China

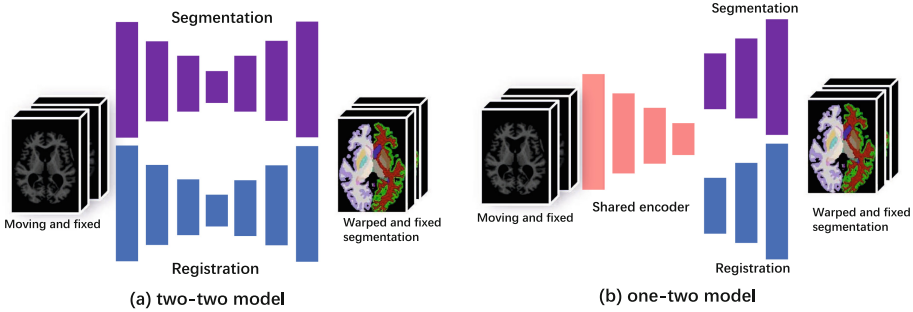
<sup>3</sup> Key Laboratory of Microenvironment Regulation and Immunotherapy of Urinary Tumors in Liaoning Province, Dalian, China

**Abstract.** Segmentation-assisted registration models can leverage few available labels in exchange for large performance gains by their complementarity. Recent related works independently build the prediction branches of deformation field and segmentation label without any information interaction except for the joint supervision. They ignore underlying relationship between the two tasks, thereby failing to fully exploit their complementary nature. To this end, we propose a ProGressively Coupling Network (PGCNet) that relies on segmentation to regularize the correct projecting of registration. Our overall framework is a multi-task learning paradigm in which features are extracted by one shared encoder and then separate prediction branches are built for segmentation and registration. In the prediction phase, we utilize the bidirectional deformation fields as bridges to warp the features of moving and fixed images to each other's segmentation branches, thereby progressively and interactively supplementing additional context information at multiple levels for their segmentation. By establishing the entangled correspondence, segmentation supervision can indirectly regularize registration stream to accurately project semantic layout for segmentation branches. In addition, we design the position correlation calculation for registration to easier capture the spatial correlation of the images from the shared features. Experimental results on public 3D brain MRI datasets show that our work performs favorably against the state-of-the-art methods.

## 1 Introduction

Deformable medical image registration has many applications in clinic, including but not limited to clinical case tracking, surgical navigation. In recent years, many advanced learning-based medical image registration methods [2, 6, 7, 15] have been proposed. Due to the limitations of accessible labels, it is more common

to employ unsupervised ways for optimization. Although unsupervised methods [5, 12, 14, 17, 19] can guide the registration network to optimize by maximizing the image-wise similarity between warped and fixed images, the lack of guidance from regions of interest (ROIs) makes the registration performance fall into a bottleneck. Moreover, medical image labeling usually requires professional medical staff, and it's laborious to get large-scale training labels.



**Fig. 1.** Two common joint registration and segmentation frameworks. (a) Two-two model: two encoders and two decoders. (b) One-two model: one shared encoder and two separate decoders.

Integrating segmentation into registration can considerably compensate for the lack of image labels via their complementarity [11, 23]. Specifically, warping segmentation result of moving image to align to that of fixed image by deformation field can provide additional supervision of ROIs for registration task. Meanwhile, as a way of data augmentation, the unlabelled images can participate in the segmentation optimization to prevent the overfitting of segmentation in few-shot situation [20, 26].

Recently, several joint learning methods have been proposed. These methods can be generally divided into two categories in terms of model structure (see Fig. 1), *two-two model* (i.e., two encoders and two decoders) [10, 11, 23] and *one-two model* (i.e., one encoder and two decoders) [8, 18, 27]. The *two-two models* employ two completely independent models for segmentation and registration, and achieve mutual learning by joint loss. DeepAtlas [23] alternately trains registration and segmentation networks to achieve mutual improvement in brain and knee images. He et al. [10] feed the segmentation masks into the registration network to provide the internal texture information of the regions of interest (ROIs) so as to avoid incorrect deformation of internal areas.

In comparison, the *one-two models* are more in line with the general paradigm of multi-task learning, which can be regarded as a process of inductive bias [21], i.e., utilizing the shared encoder to induct the commonality of each task, and then using the respective decoder for preference prediction. They can effectively reduce the risk of overfitting and reduce the parameters of the network through the way of sharing parameters [3]. However, existing *one-two models* all focus

on loss design to depict joint learning [8, 18, 27]. There is not any explicit feature interaction between two task streams. It is well known that sensible interactions help model capture extra key features or regularize networks to optimize in desired directions [22, 24, 25, 28]. In addition, the semantic information of moving and fixed images are interchangeable and mutually exploitable. As it happens, the deformation field in registration is able to interconvert the context information of moving and fixed images and the converted features can be employed as additional contextual information for their segmentation.

Based on the framework of *one-two model*, we propose a progressively coupling network (PGCNet) as shown in Fig. 2. Specifically, progressive field estimators are designed to calculate multi-level deformation fields from coarse to fine. By warping the moving features and computing the correlation between warped and fixed features at each level, the deformation field is refined progressively. Different from common correlation calculation [13, 14], we introduce learnable absolute position coordinates to provide spatial information for registration, which can better measure coordinate correspondence of voxels. In order to closely couple the registration and segmentation branches, we infuse the warped contexts into the segmentation branches. In this way, segmentation supervision sets an extra attention focusing for the registration decoder, which drives the deformation field to project semantic layout from moving image to fixed one. This semantic projection helps improve the registration accuracy. Meanwhile, the registration branch provides more sufficient semantic information for segmentation branch.

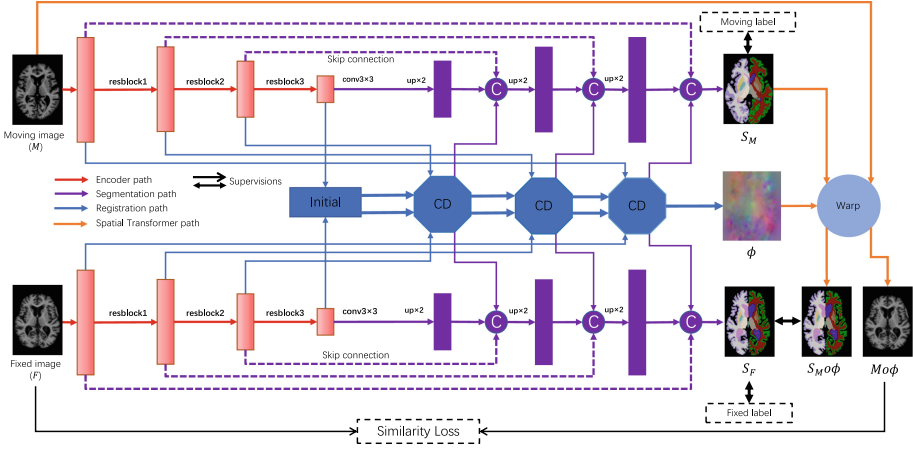
The main contributions can be summarized as follows: **1)** We propose a novel registration learning framework to establish the entangled relationship between registration and segmentation by progressively coupling the moving and fixed segmentation, which promotes both registration and segmentation performance in few-shot situation. **2)** To effectively measure the coordinate correspondence of voxels, we design the position correlation calculation, which provides spatial coordinate information for field estimator while measuring feature similarity. **3)** Experimental results on the general brain MRI datasets, OASIS and IXI, show that our method outperforms the state-of-the-art methods.

## 2 Method

Our model aims to learn registration and segmentation tasks simultaneously. Let  $M$ ,  $F$  be moving and fixed images, respectively. We parameterize our model as a function with parameter  $\theta$ ,  $f_\theta(M, F) = \phi, S_M, S_F$ , where  $\phi$  is the deformation field,  $S_M$  and  $S_F$  represent the segmentation results of  $M$  and  $F$ . Registration task completes the transformation from moving images to fixed ones, and segmentation task generates the segmentation maps of moving and fixed images.

### 2.1 Overall Network Structure

The overall pipeline of our method is shown in Fig. 2, which mainly consists of two parts, shared encoder and pyramid decoder. The parameters of the encoder



**Fig. 2.** The pipeline of progressively coupling network (PGCNet), which is an encoder-decoder structure. The red arrows denote the shared encoder path. The blue and purple arrows indicate the information flow of registration and segmentation, respectively, and the spatial transformation path is represented as brown. The structure of Coupling Decoder (the blue octagon) is shown in Fig. 3 (Color figure online).

(rose red block) and segmentation decoder (purple block) are shared. The shared encoder of our model is a 3D residual convolution network, which has 4 feature levels. Except for the input level, we use 3, 5, and 5 residual blocks at the other levels, respectively. A residual block contains two convolutional layers with pre-activation [9] structure and shortcut connection, to extract features, respectively. The numbers of channels at the four levels are 8, 16, 32, and 64, respectively, and each convolution layer is followed by Instance Normalization and Leaky-ReLU with a negative slope of 0.2 except for the last output. We first calculate the correlation of the deepest level features and then estimate the initial deformation field based on the correlation matrix. In the segmentation branch, we employ the commonly used skip connection structure, and concatenate the complementary features from the registration branch at each level, which are warped by bidirectional deformation fields.

## 2.2 Coupling Decoder

As shown in Fig. 3, there are four inputs in our coupling decoder: Field ( $\varphi$ ), Inv-Field ( $\varphi^{-}$ ), Moving Context ( $F_M$ ) and Fixed Context ( $F_F$ ). Among them, Field ( $\varphi$ ) and Inv-Field ( $\varphi^{-}$ ) are up-sampled by trilinear interpolation with the factor of 2 from the front level. Moving Context ( $F_M$ ) and Fixed Context ( $F_F$ ) are the context information from the corresponding level. We use  $\varphi$  to warp  $F_M$  and input the warped features ( $F_M \circ \varphi$ ) into the segmentation decoder of fixed image as additional complementary information (coupled features). In this way, the segmentation network obtains richer semantic information. The same

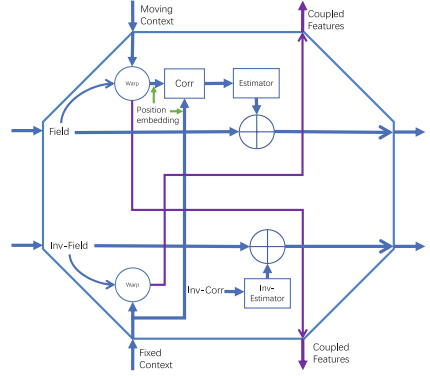
operation is done in the segmentation branch of moving image as well. Relying on the projecting ability of the deformation field, we align and combine context information of moving and fixed images. Thus, registration and segmentation branches establish an entangled correspondence to constrain each other to learn a well generalized model. Generally speaking, if the fixed image is precisely segmented, the coupled features, which are projected from moving image to fixed one, usually are well matched. That is, the coupling process can constrain the deformation field to become accurate.

### Position Correlation Calculation for Registration.

In order to measure the similarity between each voxel and its neighbors, the warped ( $F_M \circ \varphi$ ) and fixed ( $F_F$ ) context features will conduct correlation calculation to obtain a cost volume. Note that the inverse correlation is calculated between the warped ( $F_F \circ \varphi^{-}$ ) and moving ( $F_M$ ) context features, which is not shown in Fig. 3. The spatial correlation between adjacent voxels in moving and fixed images is the key to determining the deformation field. According to the cost volume, we can judge the displacement between moving and fixed images at the current level. Different from existing works [13, 14], we embed the position information into feature volumes before correlation calculation. Segmentation and registration tasks have different semantic expression requirements. The position coding can reduce the semantic confusion brought about by the shared encoder and enforce the decoder to understand spatial relationships of voxels. Let  $E$  be position coding map and  $r$  be the displacement radius, the correlation calculation is formulated as:

$$Corr(F_1, F_2, E) = \sum_{i,j,k}^r \langle (F_1 + E)_{x,y,z} \cdot (F_2 + E)_{x+i,y+j,z+k} \rangle, \quad (1)$$

where  $F_1 = F_M \circ \varphi$ ,  $F_2 = F_F$ .  $E$  is initialized as  $\mathbf{0}$ .  $x$ ,  $y$  and  $z$  represent the coordinate indexes of feature map in three directions, respectively. We move feature map point by point along  $x$ ,  $y$  and  $z$  directions with radius  $r$  and do dot product to generate a correlation map. The dimension of correlation matrix (cost volume) is  $(2r + 1)^3$ , and we set  $r = 1$  in this work. Then, Field and Inv-Field estimation modules utilize three residual blocks to compute the increment of forward and inverse displacement fields according to the correlation and inverse correlation matrices, respectively. The parameters of the two modules are shared.



**Fig. 3.** Coupling decoder. Spatial transformer utilizes deformation field to warp context feature. Correlation calculation gets cost volume. Field estimation module predicts the increment of deformation field.

### 2.3 Loss Function

The loss functions of our framework comprise three parts, which regularize registration, segmentation, and joint optimization, respectively.

The registration loss has similarity and smoothness penalty terms to align the moving and fixed images and ensure the smoothness of the deformation field. We use local normalized cross-correlation [2] with window size  $\omega$  ( $\omega = 9$ ) as similarity function. Let  $F$  and  $W$  represent the fixed and warped ( $M \circ \phi$ ) scans.  $L_2$  loss of deformation field gradient is set as regularization function. The registration loss can be defined as:

$$L_{reg} = LNCC_{\omega}(F, W) + \lambda_1 \sum \|\nabla \phi\|_2^2, \quad (2)$$

where  $\lambda_1$  is a balance hyperparameter and is set as 1 in the experiments.

For segmentation, we use the combination of weighted cross entropy and Dice loss, which is formulated as:

$$L_{wce}(S, S^*) = -\frac{1}{NP} \sum_{n=0}^N \sum_{p=0}^P w_n S_{n,p}^* \log S_{n,p}, \quad (3)$$

$$L_{dice}(S, S^*) = -\frac{1}{N} \sum_{n=0}^N \frac{\sum_{p=0}^P S_{n,p}^* \cdot S_{n,p}}{\sum_{p=0}^P S_{n,p}^* + \sum_{p=0}^P S_{n,p}}, \quad (4)$$

where  $N$  represents the total number of semantic classes,  $n$  indicates the channel of the corresponding category, and  $P$  is the number of voxels in each channel.  $S$  is the segmentation result and  $S^*$  is the voxel-wise manual segmentation label. In Eq. 3,  $w_n$  indicates the inverse proportion of voxels in category  $n$  to all voxels. For these two segmentation loss functions, their weights are fixed to the same value which means  $L_{seg} = L_{wce} + L_{dice}$ .

The joint loss function is defined as  $L_{dice}(S_F, S_M \circ \phi)$ . It regularizes registration network to provide additional training data for segmentation task. Meanwhile, registration network can pay more attention to the ROI regions.

## 3 Experiments and Results

### 3.1 Experimental Settings

**Datasets and Preprocessing.** We validate our method by using 414 T1-weighted brain MRI scans from OASIS [16] dataset and 576 T1-weighted brain MRI scans from IXI<sup>1</sup> dataset. These scans are preprocessed, which includes skull dissection, spatial normalization. All MRI scans are resampled with the same isotropic voxels of  $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ , and we center crop the scans from OASIS dataset to  $160 \times 192 \times 144$  and the scans from IXI dataset to  $160 \times 160 \times 192$ , which crops excess background to reduce the memory of GPU.

<sup>1</sup> IXI dataset is available in <https://brain-development.org/ixi-dataset/>.

The subcortical structures in OASIS and IXI are labeled as 35 and 44 categories by FreeSurfer for evaluation. We randomly select 5 scans as atlas from OASIS dataset, then the remaining scans are randomly divided into 255, 22, 132 for training, validation and testing. We do not use the labels of the training scans, only 5 labels of atlas are used to simulate a few-shot scenario. A total of 1,275, 110, 660 pairs of scans are used for training, validation and testing. Similarly, the IXI dataset is randomly split into 397, 58, 115, and 5 labeled scans are randomly taken as atlas for few-shot situation. The training, validation and testing include 1,985, 290, 575 pairs of scans, respectively.

**Implementation.** All trainings use Adam optimizer with learning rate  $1e^{-4}$  and the batch size is set as 1. We first train the network by optimizing  $L_{reg} + L_{seg}$  for 5,000 iterations and then by optimizing  $L_{reg} + L_{seg} + L_{joint}$  for 80,000 iterations. All experiments are performed on 1 Nvidia RTX 3090 GPU.

**Baseline Methods.** We compare our method with a series of registration models. SYN [1] and LDDM [4] are two traditional registration algorithms. Voxel-morph [2] and transmorph [5] are the learning-based methods, which directly predict the deformation field. While LapIRN [17] and ULAE [19] are two progressive methods. We also evaluate these two methods under the semi-supervised setting with auxiliary segmentation labels. The semi-supervised loss function is the same as ours. DeepAtlas [23], UResNet [8], and PC-Reg [10] are joint learning methods, and we adopt the same training strategy as theirs.

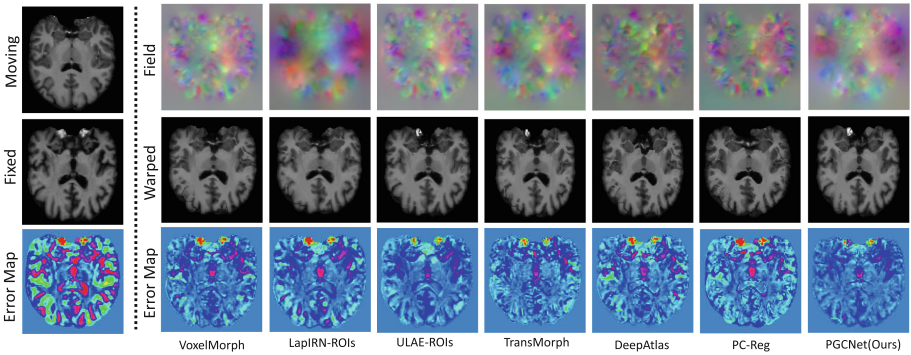
### 3.2 Experimental Results

**Registration Performance.** As shown in Table 1, PGCNet achieves competitive registration accuracy, 87.72%  $DSC_{mean}$  and 1.59  $HD95$  on OASIS dataset and 78.08%  $DSC_{mean}$  and 3.19  $HD95$  on IXI dataset. For VoxelMorph and TransMorph, which directly estimate the deformation field, their  $DSC_{mean}$  are relatively low. Coarse-to-fine registration methods, LapIRN and ULAE, can meet the complex transformation needs between images through multiple deformations, but they still have poor performance due to the lack of ROIs supervision information. By introducing additional ROIs supervision information, their performance is improved by 1.1% and 1.8%, respectively. Their improvement is weak because of the limited semi-supervised information. Our method outperforms UResNet, DeepAtlas, and PC-Reg by 12.7%, 5.7%, and 1.7%, respectively, in joint learning models. The poor performance of UResNet can be attributed to its approach of concatenating moving and fixed images to extract features, while only predicting the moving segmentation result. Figure 4 presents visual comparison of various methods. It can be observed that our method exhibits lowest registration error in the third row.

**Ablation Study.** We conduct a series of ablation experiments on OASIS dataset to verify the effectiveness of each module, as shown in Table 2. We take the

**Table 1.** Quantitative results of different registration methods on OASIS and IXI datasets. Average Dice similarity coefficient ( $DSC_{mean}$ ) and 95% percentile of the Hausdorff distance ( $HD95$ ) are used to measure the registration accuracy of the models, and the average proportion of folding voxels in the deformation fields ( $|J_\phi| < 0$ ) indicates the topological reversibility of the deformation field.

	OASIS			IXI		
	$DSC_{mean} \uparrow$	$HD95 \downarrow$	$ J_\phi  < 0 \downarrow$	$DSC_{mean} \uparrow$	$HD95 \downarrow$	$ J_\phi  < 0 \downarrow$
SYN [1]	76.08	2.22	<b>0</b>	67.21	3.76	<b>0</b>
LDDM [4]	76.07	2.23	<b>0</b>	66.94	3.90	<b>0</b>
VoxelMorph [2]	77.51	2.20	0.012	67.894	3.86	0.014
VoxelMorph-diff [7]	77.93	2.08	<b>0</b>	67.67	3.78	<b>0</b>
TransMorph [5]	79.94	2.04	0.010	69.58	3.72	0.014
LapIRN [17]	79.68	2.02	0.026	69.32	3.55	0.028
ULAE [19]	82.17	1.80	0.006	71.20	3.52	0.006
LapIRN-ROIs [17]	80.54	1.94	0.025	70.55	3.51	0.027
ULAE-ROIs [19]	83.62	1.70	0.006	72.47	3.42	0.007
UResNet [8]	77.83	2.16	0.020	67.76	3.70	0.019
DeepAtlas [23]	82.99	1.85	0.018	73.64	3.36	0.023
PC-Reg [10]	86.47	1.91	0.022	76.69	3.54	0.017
<b>PGCNet(ours)</b>	<b>87.72</b>	<b>1.59</b>	0.013	<b>78.08</b>	<b>3.19</b>	0.010



**Fig. 4.** Visualization results of different methods. In error maps, the redder color indicates the larger registration error, while the bluer color represents the smaller error.



**Table 2.** Ablation for registration.

JTrain	Couple	PEmbedding	RegDice
			82.51
✓			86.29
✓	✓		87.03
✓	✓	✓	<b>87.72</b>

**Table 3.** Ablation for segmentation.

Couple	JTrain	SegDice
		81.45
✓		84.17
✓	✓	<b>86.15</b>

registration-only model as baseline, which has the same encoder and field estimation module as the final model, utilizing common correlation calculation. Joint training (JTrain) introduces semi-supervised loss for registration. And the coupling feature connection (Couple) improves the performance of registration by guiding the deformation field in advance to map semantic features. Position embedding (PEmbedding) assists the field estimator in understanding the correlation between voxels. In addition, we further analyse the factors that affect the performance of segmentation (see Table 3). Coupling feature connection provides more adequate semantic information, resulting in 3.3% improvement, while joint training provides additional trainable data, leading to 2.4% improvement.

## 4 Conclusion

We propose a progressively coupling network (PGCNet), which employs the deformation fields to couple the registration and segmentation branches. This is a novel mode of regularization for registration and segmentation, which promotes their performance in few-shot situation. In addition, position embedding provides additional spatial coordinate information for registration branch. The experimental results indicate that our work is superior to existing methods.

**Acknowledgements.** This work was supported by the National Natural Science Foundation of China #62276046, and the Liaoning Natural Science Foundation #2021-KF-12-10.

## References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**(1), 26–41 (2008)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9252–9260 (2018)
3. Baxter, J.: A Bayesian/information theoretic model of learning to learn via multiple task sampling. *Mach. Learn.* **28**(1), 7–39 (1997)
4. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows. *Int. J. Comput. Vision* **61**(2), 139–157 (2005)

5. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: TransMorph: transformer for unsupervised medical image registration. *Med. Image Anal.* **82**, 102615 (2022)
6. Chen, X., Xia, Y., Ravikumar, N., Frangi, A.F.: A deep discontinuity-preserving image registration network. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 46–55. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87202-1\\_5](https://doi.org/10.1007/978-3-030-87202-1_5)
7. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Med. Image Anal.* **57**, 226–236 (2019)
8. Estienne, T., et al.: U-ReSNet: ultimate coupling of registration and segmentation with deep nets. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 310–319. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32248-9\\_35](https://doi.org/10.1007/978-3-030-32248-9_35)
9. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46493-0\\_38](https://doi.org/10.1007/978-3-319-46493-0_38)
10. He, Y., et al.: Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching. *IEEE J. Biomed. Health Inform.* **26**(3), 1177–1187 (2022). <https://doi.org/10.1109/JBHI.2021.3095409>
11. He, Y., et al.: Deep complementary joint model for complex scene registration and few-shot segmentation on medical images. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12363, pp. 770–786. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58523-5\\_45](https://doi.org/10.1007/978-3-030-58523-5_45)
12. Hu, B., Zhou, S., Xiong, Z., Wu, F.: Recursive decomposition network for deformable image registration. *IEEE J. Biomed. Health Inform.* **26**(10), 5130–5141 (2022)
13. Jonschkowski, R., Stone, A., Barron, J.T., Gordon, A., Konolige, K., Angelova, A.: What matters in unsupervised optical flow. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12347, pp. 557–572. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58536-5\\_33](https://doi.org/10.1007/978-3-030-58536-5_33)
14. Kang, M., Hu, X., Huang, W., Scott, M.R., Reyes, M.: Dual-stream pyramid registration network. *Med. Image Anal.* **78**, 102379 (2022)
15. Lv, J., et al.: Joint progressive and coarse-to-fine registration of brain MRI via deformation field integration and non-rigid feature fusion. *IEEE Trans. Med. Imaging* **41**(10), 2788–2802 (2022). <https://doi.org/10.1109/TMI.2022.3170879>
16. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **19**(9), 1498–1507 (2007)
17. Mok, T.C.W., Chung, A.C.S.: Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 211–221. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-59716-0\\_21](https://doi.org/10.1007/978-3-030-59716-0_21)
18. Qin, C., et al.: Joint learning of motion estimation and segmentation for cardiac MR image sequences. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 472–480. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00934-2\\_53](https://doi.org/10.1007/978-3-030-00934-2_53)
19. Shu, Y., Wang, H., Xiao, B., Bi, X., Li, W.: Medical image registration based on uncoupled learning and accumulative enhancement. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 3–13. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87202-1\\_1](https://doi.org/10.1007/978-3-030-87202-1_1)

20. Vakalopoulou, M., et al.: AtlasNet: multi-atlas non-linear deep networks for medical image segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 658–666. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00937-3\\_75](https://doi.org/10.1007/978-3-030-00937-3_75)
21. Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D., Van Gool, L.: Multi-task learning for dense prediction tasks: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(7), 3614–3633 (2021)
22. Xu, D., Ouyang, W., Wang, X., Sebe, N.: PAD-Net: multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 675–684 (2018)
23. Xu, Z., Niethammer, M.: DeepAtlas: joint semi-supervised learning of image registration and segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 420–429. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32245-8\\_47](https://doi.org/10.1007/978-3-030-32245-8_47)
24. Zhang, Z., Cui, Z., Xu, C., Jie, Z., Li, X., Yang, J.: Joint task-recursive learning for semantic segmentation and depth estimation. In: *European Conference on Computer Vision*, pp. 235–251 (2018)
25. Zhang, Z., Cui, Z., Xu, C., Yan, Y., Sebe, N., Yang, J.: Pattern-affinitive propagation across depth, surface normal and semantic segmentation. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4106–4115 (2019)
26. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8543–8553 (2019)
27. Zhao, F., Wu, Z., Wang, L., Lin, W., Xia, S., Li, G.: A deep network for joint registration and parcellation of cortical surfaces. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12904, pp. 171–181. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87202-1\\_17](https://doi.org/10.1007/978-3-030-87202-1_17)
28. Zhao, X., Pang, Y., Zhang, L., Lu, H.: Joint learning of salient object detection, depth estimation and contour extraction. *IEEE Trans. Image Process.* **31**, 7350–7362 (2022)