Double-Uncertainty Assisted Spatial and Temporal Regularization Weighting for Learning-based Registration

Preprint · July 2021

CITATIONS
0

READS
85

10 authors, including:

Zhe Xu
The Chinese University of Hong Kong
29 PUBLICATIONS **158** CITATIONS
SEE PROFILE

Jie Luo
Harvard Medical School
38 PUBLICATIONS **315** CITATIONS
SEE PROFILE

Jiangpeng Yan
Tsinghua University
43 PUBLICATIONS **198** CITATIONS
SEE PROFILE

Jayender Jagadeesan
Harvard Medical School
107 PUBLICATIONS **1,571** CITATIONS
SEE PROFILE

Double-Uncertainty Assisted Spatial and Temporal Regularization Weighting for Learning-based Registration

Zhe Xu[1⋆], Jie Luo[3], Donghuan Lu[2], Jiangpeng Yan[4], Jayender Jagadeesan[3], William Wells III[3], Sarah Frisken[3], Kai Ma[2], Yefeng Zheng[2], and Raymond Kai-yu Tong[1]

[1] The Chinese University of Hong Kong, Hong Kong, China
[2] Tencent Jarvis Lab, Shenzhen, China
[3] Brigham and Women's Hospital, Harvard Medical School, Boston, USA
[4] Tsinghua University, Beijing, China

Abstract. In order to tackle the difficulty associated with the ill-posed nature of the image registration problem, researchers use regularization to constrain the solution space. For most learning-based registration approaches, the regularization usually has a fixed weight and only constrains the spatial transformation. Such convention has two limitations: (1) The regularization strength of a specific image pair should be associated with the content of the images, thus the "one value fits all" scheme is not ideal; (2) Only spatially regularizing the transformation (but overlooking the temporal consistency of different estimations) may not be the best strategy to cope with the ill-posedness. In this study, we propose a mean-teacher based registration framework. This framework incorporates an additional *temporal regularization* term by encouraging the teacher model's temporal ensemble prediction to be consistent with that of the student model. At each training step, it also automatically adjusts the weights of the *spatial regularization* and the *temporal regularization* by taking account of the transformation uncertainty and appearance uncertainty derived from the perturbed teacher model. We perform experiments on multi- and uni-modal registration tasks, and the results show that our strategy outperforms the traditional and learning-based benchmark methods.

Keywords: Image Registration · Regularization · Registration Uncertainty.

1 Introduction

Recently, learning-based registration (LBR) has gained success in many clinical applications due to their substantial improvement in computational efficiency

⋆ Contributed to this work at Tencent Jarvis Lab.

and accuracy [2,29,28,26,30]. In the training stage of LBR, given a set of paired image data, the neural network optimizes the cost function

$$\mathcal{L}_{sim}\left(I_f, I_m \circ \phi\right) + \lambda \mathcal{L}_{reg}(\phi), \tag{1}$$

to learn a mapping function that can rapidly estimate the deformation field ϕ for a new pair of images. In the cost function, the first term $\mathcal{L}_{sim}$ quantifies the appearance dissimilarity between the fixed image I_f and the warped moving image $I_m \circ \phi$. Since image registration is an ill-posed problem, researchers also use the second regularization term $\mathcal{L}_{reg}$ to constrain its solution space. The strength of the regularization is controlled by a weighting coefficient λ.

Conventionally, the weight λ is set to a fixed value throughout the training stage, assuming that all image pairs in the dataset require the same level of regularization. However, this "one value fits all" scheme is suboptimal because the regularization strength of a specific image pair should be associated with the content of the images. In the traditional iterative image registration literature, there have been works on adaptively varying the weight of the regularization. Some approaches used a tissue segmentation map to divide the image into several regions, and assigned the value of λ for a specific region based on its tissue type [19,23]. Other approaches computed the gradient of image dissimilarity and use it to adjust λ [16,24]. One principal way to adaptively change the regularization weight was to model λ using Bayesian inference. The Bayesian inference scheme allowed λ to be inferred from the data without strong heuristics [21,20,18,7]. However, due to fundamental differences, the aforementioned adaptive strategies for iterative registration cannot be directly applied to learning-based registration frameworks.

As an ill-posed problem, image registration has no unique solution. Regularization preserves the smoothness and penalizes unnecessary complexity of the estimated transformation, and hence reduces the solution space of image registration. For LBR, researchers have proposed task-specific regularization, e.g., population-level statistics [3] and biomechanical model [12,17], to replace traditional linear elasticity [15] and thin-plate spline bending energy [4]. Like traditional regularization, these methods all belong to the category of applying spatial constraints to the transformation. Here we call them *spatial regularization*. In our experiments, however, we notice that the estimated solutions vary greatly at different time steps. It might be advantageous to develop a *temporal regularization* to further constrain the solution to such an ill-posed problem.

In this paper, we propose a double-uncertainty assisted spatial and temporal regularization weighting strategy, cooperated with our Mean-Teacher (MT) model [25] based registration framework. Specifically, this framework incorporates an additional *temporal regularization* term by encouraging the consistency between the predictions of the student model and the momentum teacher model. Furthermore, distinct from the "one value fits all" scheme, the framework automatically adjusts both the spatial and temporal regularization weights for each training sample by taking account of their transformation uncertainty and the appearance uncertainty [14] derived from the perturbed teacher model. We evaluate the proposed strategy on two datasets: (1) a clinically acquired multimodal

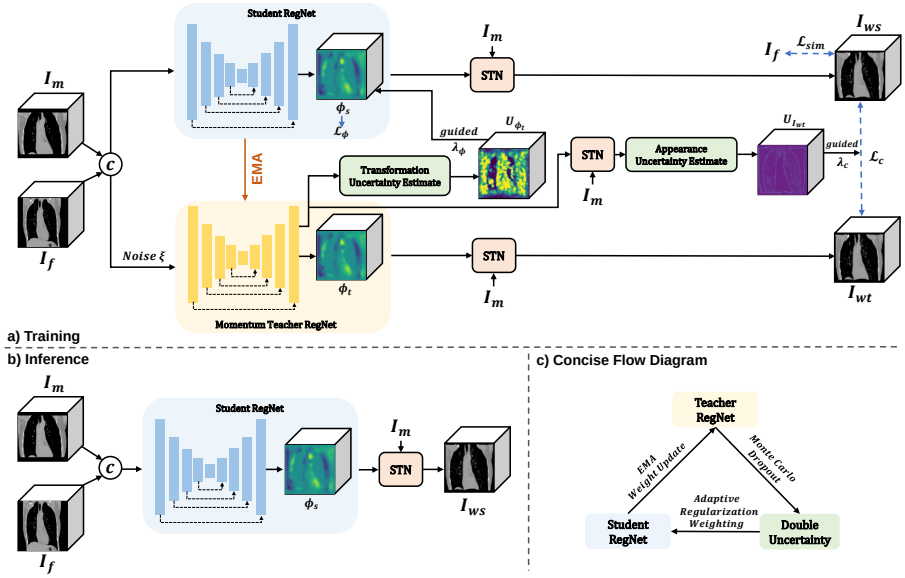

Fig. 1. Illustration of the proposed framework with training process (a), inference process (b) and a concise flow diagram (c).

CT-MRI dataset of the abdomen; (2) a unimodal CT dataset of the thorax taken at inspiration and expiration stages. Experimental results show that our strategy outperforms the traditional and learning-based benchmark methods.

2 Methods

The proposed Mean-Teacher registration network is shown in Fig. 1 with the illustrations of the training process (a), inference process (b), and a concise flow diagram (c). In this section, we give a detailed explanation of every step.

2.1 Mean-Teacher based Temporal Regularization

Inspired by the strategy of weight-averaged consistency [25], we propose a novel Mean-Teacher registration network, which is shown in Fig. 1(a), including two sub-models called the student registration network (Student RegNet) and the momentum teacher registration network (Momentum Teacher RegNet).

Student RegNet. The student model is a typical registration network and updated by back-propagation. Theoretically, the network can adopt various backbones. Here, we follow the same U-Net architecture used in VoxelMorph (VM) [2]. Concisely, the moving image I_m and the fixed image I_f are concatenated as

a single 2-channel 3D image input, and downsampled by four $3 \times 3 \times 3$ convolutions with stride of 2 as the encoder. Then, corresponding 32-filter convolutions and four upsampling layers are applied to form a decoder, followed by four convolutions to refine the 3-channel deformation field ϕ_s. Skip connections between encoder and decoder are also applied. Given the predicted ϕ_s, the network uses the Spatial Transformation Network (STN) [13] to warp the moving image I_m into I_{ws}, so that the dissimilarity $\mathcal{L}_{sim}$ between I_{ws} and I_f can be measured.

Momentum Teacher RegNet. The teacher model uses the same network architecture as the student model but with a different weight updating strategy as well as perturbed input. Unlike back-propagation for the student model, the weights of the teacher model are updated from that of the student model via Exponential Moving Average (EMA) strategy, allowing to leverage the temporal information between the adjacent training steps. Formally, denoting the weights of the teacher model and the student model at training step k as θ'_k and θ_k, respectively, we update θ'_k as: $\theta'_k = \alpha\theta'_{k-1} + (1-\alpha)\theta_k$, where α is a momentum coefficient, i.e., EMA decay. Recommended by [25], α is set to 0.99. The output of the teacher model is the second predicted deformation field ϕ_t. Then, the STN warps the I_m into I_{wt} using ϕ_t. Another key idea in the Mean-Teacher registration framework is adding perturbation, i.e., Gaussian noise ξ in this work, to the input of the teacher model, which can help the network capture more accurate and robust anatomical information in the presence of noise and artifacts [5,25].

Temporal Regularization. As mentioned above, I_{ws} and I_{wt} are the registration results predicted by the student model and the momentum teacher model, respectively. Following the same spirit of temporal ensembling [25,27], we encourage the temporal ensemble prediction of the teacher model to be consistent with that of the student model under perturbations by adding an appearance consistency constraint $\mathcal{L}_c(I_{ws}, I_{wt})$ to the training loss. In contrast to $\mathcal{L}_\phi$ which spatially constrains the transformation, $\mathcal{L}_c$ penalizes the difference between predictions calculated at adjacent training time steps. Thus, we call $\mathcal{L}_c$ *temporal regularization*.

2.2 Overall Training Loss

The overall loss is a weighted combination of dissimilarity loss $\mathcal{L}_{sim}$, spatial regularization loss $\mathcal{L}_\phi$ and temporal regularization loss $\mathcal{L}_c$, calculated by:

$$\mathcal{L} = \mathcal{L}_{sim} + \lambda_\phi \mathcal{L}_\phi + \lambda_c \mathcal{L}_c, \tag{2}$$

where λ_ϕ and λ_c are the uncertainty assisted adaptive trade-off weights, as elaborated in Sec. 2.3. In order to handle both multimodal and unimodal registration, we use the Modality Independent Neighborhood Descriptor (MIND) [8] based dissimilarity metric for $\mathcal{L}_{sim}$ and $\mathcal{L}_c$. Following the benchmark method [2], the choice of $\mathcal{L}_\phi$ is the generic L2-norm of the deformation field gradients.

2.3 Double-Uncertainty Assisted Adaptive Weighting

Distinct from the previous "one value fits all" strategy, we propose a double-uncertainty assisted adaptive weighting method to locate the uncertain samples and then adaptively adjust λ_ϕ and λ_c for each training step.

Double-Uncertainty Estimation. Besides deformation prediction, the teacher model can serve as an uncertainty estimation branch. First, following the strategy used in [27], we repetitively perform N stochastic forward passes on the teacher model under random Gaussian noise ξ for each input. After this step, we can obtain a set of voxel-wise predicted deformation fields $\{\phi_{t_i}\}_{i=1}^{N}$. With the clean input I_m used as the moving image in STN as shown in Fig. 1, a set of warped images $\{I_{wt_i}\}_{i=1}^{N}$ can also be obtained. We propose to use *the absolute value of the ratio of standard deviation to the mean*, which can characterize the normalized volatility of the predicted results, to represent the uncertainty [22]. More specifically, the registration uncertainties here can be categorized into the transformation uncertainty and appearance uncertainty [14]. Formulating $\mu^c_{\phi_t} = \frac{1}{N}\sum_{i=1}^{N}\phi^c_{t_i}$ and $\mu_{I_{wt}} = \frac{1}{N}\sum_{i=1}^{N} I_{wt_i}$ as the mean of the deformation fields and the warped images, respectively, where c represents the c^{th} channel of the deformation field (i.e., x, y, z displacements) and i denotes the i^{th} forward pass, the transformation uncertainty map $U_{\phi_t} \in \mathbb{R}^{H\times W\times D\times C}$ and the appearance uncertainty map $U_{I_{wt}} \in \mathbb{R}^{H\times W\times D}$ can be calculated as:

$$\begin{cases} \sigma^c_{\phi_t} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\phi^c_{t_i} - \mu^c_{\phi_t}\right)^2} \quad \text{and} \quad U^c_{\phi_t} = \left|\frac{\sigma^c_{\phi_t}}{\mu^c_{\phi_t}}\right|, \\ \sigma_{I_{wt}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I_{wt_i} - \mu_{I_{wt}}\right)^2} \quad \text{and} \quad U_{I_{wt}} = \left|\frac{\sigma_{I_{wt}}}{\mu_{I_{wt}}}\right|. \end{cases} \tag{3}$$

Adaptive Weighting. With the guidance of U_{ϕ_t} and $U_{I_{wt}}$, we propose to adaptively assign the weights λ_ϕ and λ_c for the *spatial regularization* $\mathcal{L}_\phi$ and the *temporal regularization* $\mathcal{L}_c$ for specific sample during network training. Intuitively, stronger spatial regularization should be given when the network tends to produce more uncertain predictions to reduce unnecessary transformation complexity. As for the temporal regularization along with $U_{I_{wt}}$ that characterizes sample-wise registration difficulty, more uncertain samples should receive more attention for the learning process. Thus, for each training step s, we update λ_ϕ and λ_c as follows:

$$\lambda_\phi(s) = k_1 \cdot \frac{\sum_v \mathbb{I}\left(U_{\phi_t}(s) > \tau_1\right)}{V_{U_{\phi_t}}} \quad \text{and} \quad \lambda_c(s) = k_2 \cdot \frac{\sum_v \mathbb{I}\left(U_{I_{wt}}(s) > \tau_2\right)}{V_{U_{I_{wt}}}}, \tag{4}$$

where $\mathbb{I}(\cdot)$ is the indicator function; v denotes the v-th voxel; $V_{U_{\phi_t}}$ and $V_{I_{wt}}$ represent the volume sizes of U_{ϕ_t} and $U_{I_{wt}}$, respectively; k_1 and k_2 are the empirical scalar values; and τ_1 and τ_2 are the thresholds to select the most uncertain targets.

Noteworthily, the proposed strategy can be applied to any learning-based registration architecture without increasing the number of trainable parameters

and the size of inference network, i.e., only the student model is used for inference as shown in Fig. 1 (b), ensuring the computational efficiency. Besides, this strategy can potentially reduce the burden of exhaustive grid search for the optimal fixed weight.

3 Experiments and Results

Datasets. We evaluate the proposed strategy on a clinically acquired CT-MRI abdominal dataset and a public thoracic CT dataset:
1) *Abdominal (ABD) CT-MRI Dataset*. Under IRB approved study, a 50-pair intra-patient abdominal CT-MRI dataset was obtained from a local hospital with ground-truth segmentation masks for the liver, kidney and spleen. We randomly divided the dataset into two groups for training (40 cases) and testing (10 cases). After sequential preprocessing steps including resampling, affine prealignment, intensity normalization and cropping, the images were processed into $176 \times 176 \times 128$ sub-volumes with $1mm$ isotropic voxels.
2) *Thoracic CT Dataset*. This dataset consists of paired inspiratory and expiratory thoracic CT images from 30 subjects with lung segmentation masks (20 cases for training [11] and 10 cases for testing [10]). After similar preprocessing, all images were cropped into $128 \times 128 \times 160$ sub-volumes. Note that in addition to the severe deformation between expiratory and inspiratory images, there is another challenge that the lungs are not fully visible in expiratory scans (shown in Fig. 2).

Implemenation and Evaluation Criteria. The proposed framework is implemented on PyTorch and trained on an NVIDIA Titan X (Pascal) GPU. We employ the Adam optimizer with a learning rate of 0.0001 with a decay factor of 0.9. The batch size is set to 1 due to the limitation of GPU memory. For the double-uncertainty estimation, we set $N = 10$, and the scalar values k_1 and k_2 are empirically set to 5 and 1, respectively, referring to the exploratory experiments using backbone network, i.e., the VM model [2]. Thresholds τ_1 and τ_2 are set to 10% and 1%. We adopt a series of evaluation metrics, including Average Surface Distance (ASD) and the average Dice score between the segmentation masks of warped images and fixed images. In addition, the average percentage of voxels with non-positive Jacobian determinant ($|J_\phi| \leq 0$) in the deformation fields and the standard deviation of the Jacobian determinant ($\sigma(|J_\phi|)$) are obtained to quantify the diffeomorphism and smoothness of the deformation fields.

Comparison Study. The proposed strategy is compared with other cutting edge approaches, including two traditional methods SyN [1] and Deeds [9] with five levels of discrete optimization, as well as the benchmark learning-based method VoxelMorph (VM) [2] and its probabilistic diffeomorphic version DIF-VM [6]. Consistently, we use the MIND-based dissimilarity metric in all learning-based methods to handle both multi- and uni-modal tasks. Although DIF-VM preserves better diffeomorphism properties, we found that its results are often

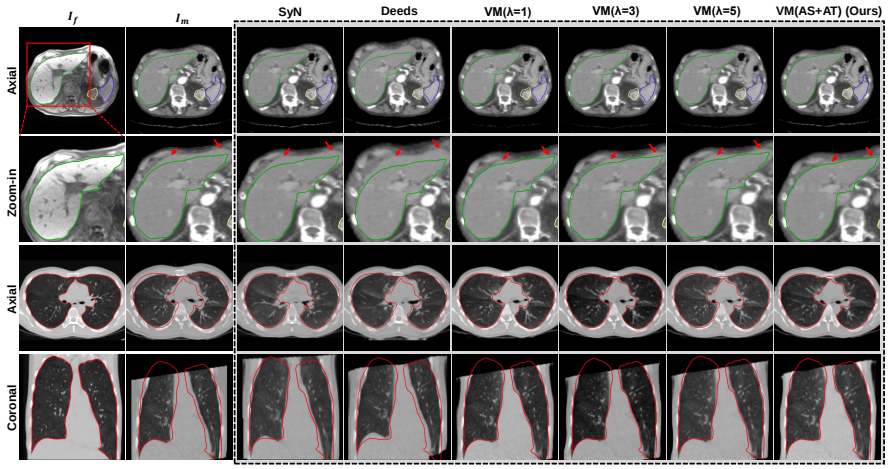

Fig. 2. Visualization of an abdominal CT-MRI case and a thoracic CT case. The segmentation contours of the liver (green), kidney (yellow) and spleen (blue) of the fixed abdominal MRI are presented in the abdominal image. The red contours illustrate the lung segmentation of the fixed inspiratory CT image. Better alignment drives structures to close to the fixed contours of I_f. The red arrows indicate the registration of interest at the organ boundary.

suboptimal, as shown in Table 1. Thus, we mainly focus on VM, which is also adopted as our backbone network. For comparison, five VM models were trained with varying fixed spatial regularization weights from 1 to 5. Wherein, we select three representative models with weights $\{1, 3, 5\}$ to present, as shown in Table 1 and Fig. 2. Our training strategy, i.e., adaptive spatial and temporal regularization weighting, is denoted as **VM(AS+AT)**. Visually, the structure boundaries registered by SyN and Deeds still have considerable disagreements, while the learning-based methods achieve more appealing boundary alignment for both cases. Quantitatively, we can observe that our training strategy further improves the registration performance of VM models in terms of Dice and ASD along with better properties of diffeomorphism and smoothness, implying that our strategy can effectively find a more desirable solution in such an ill-posed problem. In terms of efficiency, all learning-based methods can register a pair of images in a second with a GPU.

Ablation Study. To further verify the effectiveness of the proposed training strategy (VM(AS+AT)), we conduct an ablation study with two variants: a) the VM backbone with additional temporal regularization **(VM(S+T))**, and b) extended VM(S+T) by adaptively weighting the spatial regularization **(VM(AS+T))**. The quantitative results are also presented in Table 1. We can observe that the two variants also provide better results than the previous fixed-weight and spatial regularization-only strategy. Overall, VM(AS+AT) achieves

Table 1. Quantitative results for abdominal CT-MRI and thoracic CT registration. Higher average Dice score (%) and lower ASD (mm) are better. Best results are shown in bold. Average percentage of foldings ($|J_\phi| \leq 0$) and the standard deviation of the Jacobian determinant ($\sigma(|J_\phi|)$) are also given.

Methods	Abdominal CT-MRI								Thoracic CT			
	Dice			ASD			$\lvert J_\phi\rvert \leq 0$	$\sigma(\lvert J_\phi\rvert)$	Dice	ASD	$\lvert J_\phi\rvert \leq 0$	$\sigma(\lvert J_\phi\rvert)$
	Liver	Spleen	Kidney	Liver	Spleen	Kidney						
Initial	76.23	77.94	80.18	4.98	2.02	1.95	-	-	86.37	2.51	-	-
SyN	79.42	80.33	82.68	4.83	1.62	1.91	0.07%	0.40	86.67	2.27	0.09%	0.37
Deeds	82.16	81.48	83.82	3.97	1.44	1.59	0.01%	0.27	87.68	2.35	0.02%	0.14
DIF-VM	83.38	82.71	83.05	3.86	1.50	1.61	<0.001%	0.11	89.15	1.95	<0.001%	0.16
VM($\lambda = 1$)	85.79	**84.16**	83.51	3.04	1.32	1.53	0.02%	0.17	90.77	1.90	0.06%	0.27
VM($\lambda = 3$)	85.33	84.03	83.37	3.22	1.31	1.55	0.001%	0.12	90.45	1.94	<0.001%	0.12
VM($\lambda = 5$)	84.18	82.69	82.98	3.85	1.47	1.68	0%	0.08	89.84	2.02	<0.0005%	0.10
VM(AS+AT)(ours)	**86.96**	84.07	**84.01**	**2.67**	1.24	**1.39**	<0.0005%	0.16	**91.40**	**1.67**	<0.0005%	0.15
Ablation Study												
VM(S+T)	85.74	84.01	83.53	3.01	1.35	1.48	0.001%	0.13	90.86	1.84	<0.001%	0.14
VM(AS+T)	86.03	84.05	83.62	2.98	**1.22**	1.43	<0.0005%	0.14	91.18	1.73	<0.001%	0.16

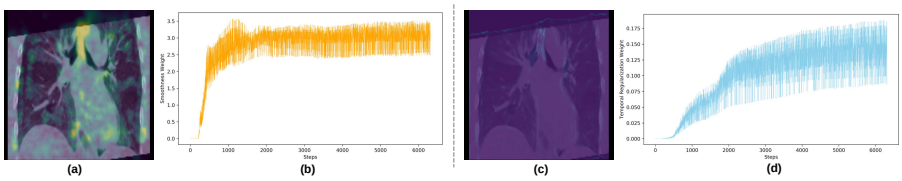

Fig. 3. (a) and (c) are the examples of U_{ϕ_t} and $U_{I_{wt}}$ (overlay on the input volumes) for thoracic CT registration, where brighter areas denote more uncertain regions. (b) and (d) show the adaptive weighting process across the training steps for spatial regularization and temporal regularization, respectively.

the best performance.

Visualized Uncertainty Map and Weighting Process. Examples of two uncertainty maps U_{ϕ_t} and $U_{I_{wt}}$ for thoracic CT registration are visualized in Fig. 3 (a) and (c), respectively. We can observe that high uncertainty often occurs in hard-to-align ambiguous areas. Assisted by the two uncertainties, the weights of the spatial regularization and temporal regularization are adaptively modulated after each step (shown in Fig. 3 (b) and (d), respectively) to help the model pursue a better trade-off between better image alignment (indicated by higher Dice and lower ASD scores) with more desirable diffeomorphism properties (<0.0005% foldings in the above two tasks).

4 Conclusion

In this paper, we proposed a double-uncertainty assisted spatial and temporal regularization weighting strategy, assisted by our Mean-Teacher registration framework. This framework incorporates an additional *temporal regularization* term by encouraging the momentum teacher model's prediction to be consistent with that of the student model under perturbations. Further, it makes use of

two registration uncertainties to adaptively adjust the weights of both *spatial regularization* and the proposed *temporal regularization* during training. Experiments on both multi- and uni-modal registration tasks showed that our strategy outperformed the traditional and learning-based benchmark methods.

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. Medical Image Analysis **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: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9252–9260 (2018)
3. Bhalodia, R., Elhabian, S.Y., Kavan, L., Whitaker, R.T.: A cooperative autoencoder for population-based regularization of CNN image registration. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 391–400. Springer (2019)
4. Bookstein, F.: Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Medical Image Analysis pp. 225–243 (1997)
5. Cui, W., Liu, Y., Li, Y., Guo, M., Li, Y., Li, X., Wang, T., Zeng, X., Ye, C.: Semi-supervised brain lesion segmentation with an adapted mean teacher model. In: International Conference on Information Processing in Medical Imaging. pp. 554–565. Springer (2019)
6. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 729–738. Springer (2018)
7. Folgoc, L., Delingette, H., Criminisi, A., Ayache, N.: Sparse Bayesian registration of medical images for self-tuning of parameters and spatially adaptive parametrization of displacements. Medical Image Analysis **36**, 79–97 (2017)
8. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, M., Schnabel, J.A.: MIND: Modality independent neighbourhood descriptor for multimodal deformable registration. Medical Image Analysis **16**(7), 1423–1435 (2012)
9. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Transactions on Medical Imaging **32**(7), 1239–1248 (2013)
10. Hering, A., Murphy, K., van Ginneken, B.: Learn2Reg Challenge: CT Lung Registration - Test Data (2020). https://doi.org/10.5281/zenodo.4048761
11. Hering, A., Murphy, K., van Ginneken, B.: Learn2Reg Challenge: CT Lung Registration - Training Data (2020). https://doi.org/10.5281/zenodo.3835682
12. 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)
13. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems. pp. 2017–2025 (2015)

14. Luo, J., Sedghi, A., Popuri, K., Cobzas, D., Zhang, M., Preiswerk, F., Toews, M., Golby, A., Sugiyama, M., Wells, W., Frisken, S.: On the applicability of registration uncertainty. In: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 107–115. Springer (2019)
15. Miller, M., Christensen, G., Amit, Y., Grenander, U.: Mathematical textbook of deformable neuroanatomies. In: PNAS. pp. 11944–11948 (1993)
16. Papiez, B., Heinrich, M., Risser, L., Schnabel, J.: Complex Lung Motion Estimation via Adaptive Bilateral Filtering of the Deformation Field. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 25–32. Springer (2013)
17. Qin, C., Wang, S., Chen, C., Qiu, H., Bai, W., Rueckert, D.: Biomechanics-informed Neural Networks for Myocardial Motion Tracking in MRI. arXiv preprint arXiv:2006.04725 (2020)
18. Risholm, P., Janoos, F., Norton, I., Golby, A., W.M., W.: Bayesian characterization of uncertainty in intra-subject non-rigid registration. Medical Image Analysis **17**(5), 538–555 (2013)
19. Schmah, T., Risser, L., Vialard, F.: Left-Invariant Metrics for Diffeomorphic Image Registration with Spatially-Varying Regularisation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 203–210. Springer (2013)
20. Simpson, I., Cardoso, M., Modat, M., Cash, D., Woolrich, M., Andersson, J., Schnabel, J., Ourselin, S.: Probabilistic non-linear registration with spatially adaptive regularisation. Medical Image Analysis **26**(1), 203–216 (2015)
21. Simpson, I., Schnabel, J., Groves, A., Andersson, J., Woolrich, M.: Probabilistic inference of regularisation in non-rigid registration. NeuroImage **59**(3), 2438–2451 (2012)
22. Smith, L., Gal, Y.: Understanding measures of uncertainty for adversarial example detection. arXiv preprint arXiv:1803.08533 (2018)
23. Staring, M., Klein, S., Pluim, J.: Nonrigid registration with tissue-dependent filtering of the deformation field. Physics in Medicine and Biology **52**(23), 6879 (2007)
24. Tang, L., Hamarneh, G., Abugharbieh, R.: Reliability-driven, spatially-adaptive regularization for deformable registration. In: International Workshop on Biomedical Image Registration. pp. 173–185. Springer (2010)
25. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems. pp. 1195–1204 (2017)
26. 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)
27. Wang, Y., Zhang, Y., Tian, J., Zhong, C., Shi, Z., Zhang, Y., He, Z.: Double-uncertainty weighted method for semi-supervised learning. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 542–551. Springer (2020)
28. Xu, Z., Luo, J., Yan, J., Pulya, R., Li, X., Wells, W., Jagadeesan, J.: Adversarial uni-and multi-modal stream networks for multimodal image registration. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 222–232. Springer (2020)
29. Xu, Z., Yan, J., Luo, J., Li, X., Jagadeesan, J.: Unsupervised multimodal image registration with adaptative gradient Guidance. In: ICASSP 2021-2021 IEEE In-

ternational Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1225–1229. IEEE (2021)
30. Xu, Z., Yan, J., Luo, J., Wells, W., Li, X., Jagadeesan, J.: Unimodal cyclic regularization for training multimodal image registration networks. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1660–1664. IEEE (2021)