Double-Uncertainty Guided Spatial and Temporal Consistency Regularization Weighting for Learning-based Abdominal Registration

Zhe Xu[1], **Jie Luo**[3], **Donghuan Lu**[2], **Jiangpeng Yan**[4], **Sarah Frisken**[3], **Jayender Jagadeesan**[3], **William M. Wells III**[3], **Xiu Li**[5], **Yefeng Zheng**[2], **Raymond Kai-yu Tong**[1]

[1]Department of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, China

[2]Tencent Healthcare Co., Jarvis Lab, Shenzhen, China

[3]Brigham and Women's Hospital, Harvard Medical School, Boston, USA

[4]Department of Automation, Tsinghua University, Beijing, China

[5]Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

Abstract

In order to tackle the difficulty associated with the ill-posed nature of the image registration problem, regularization is often used 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: (i) Besides the laborious grid search for the optimal fixed weight, the regularization strength of a specific image pair should be associated with the content of the images, thus the "one value fits all" training scheme is not ideal; (ii) Only spatially regularizing the transformation may neglect some informative clues related to the ill-posedness. In this study, we propose a mean-teacher based registration framework, which incorporates an additional temporal consistency regularization term by encouraging the teacher model's prediction to be consistent with that of the student model. More importantly, instead of searching for a fixed weight, the teacher enables automatically adjusting the weights of the spatial regularization and the temporal consistency regularization by taking advantage of the transformation uncertainty and appearance uncertainty. Extensive experiments on the challenging abdominal CT-MRI registration show that our training strategy can promisingly advance the original learning-based method in terms of efficient hyperparameter tuning and a better tradeoff between accuracy and smoothness.

kytong@cuhk.edu.hk, caleblu@tencent.com.

1   Introduction

Recently, learning-based multimodal abdominal registration has greatly advanced percutaneous nephrolithotomy due to their substantial improvement in computational efficiency and accuracy [2, 24, 10, 21, 25, 23]. In the training stage, given a set of paired image data, the neural network optimizes the cost function

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

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, the second regularization term $\mathcal{L}_{reg}$ is used to constrain its solution space. The tradeoff between registration accuracy and transformation smoothness is controlled by a weighting coefficient λ.

In classical iterative registration methods, the weight λ is manually tuned for each image pair. Yet, in most learning-based approaches [2], the weight λ is commonly set to a fixed value for all image pairs throughout the training stage, assuming that they require the same regularization strength. Besides the notoriously time-consuming grid search for the so-called "optimal" fixed weight, this "one value fits all" scheme (Fig. 1(c)) may be suboptimal because the regularization strength of a specific image pair should be associated with their content and misalignment degree, especially for the challenging abdominal registration with various large deformation patterns. In these regards, HyperMorph [11] estimated the effect of hyperparameter values on deformations with the additionally trained hypernetworks [7]. Being more parameter-efficient, recently, Mok et al. [16] introduced conditional instance normalization [6] into the backbone network and used an extra distributed mapping network to implicitly control the regularization by normalizing and shifting feature statistics with their affine parameters. As such, it enables optimizing the model with adaptive regularization weights during training and reducing the human efforts in hyperparameter tuning. We also focus on this underexploited topic, yet, propose an explicit and substantially different alternative without changing any components of the backbone.

On the other hand, besides the traditional regularization terms, e.g., smoothness and bending energy [4], task-specific regularizations have been further proposed, e.g., population-level statistics [3] and biomechanical models [12, 17]. Inherently, all these methods still belong to the category of *spatial regularization*. Experimentally, however, we notice that the estimated solutions in unsupervised registration vary greatly at different training steps due to the ill-posedness. Only spatially regularizing the transformation at each training step may neglect some informative clues related to the ill-posedness. It might be advantageous to further exploit such temporal information across different training steps.

In this paper, we present a novel double-uncertainty guided spatial and temporal consistency regularization weighting strategy, in conjunction with the Mean-Teacher (MT) [20] based registration framework. Specifically, inspired by the consistency regularization [20, 22] in semi-supervised learning, this framework further incorporates an additional *temporal consistency regularization* term by encouraging the consistency between the predictions of

the student model and the temporal ensembling predictions of the teacher model. More importantly, instead of laboriously grid searching for the optimal fixed regularization weight, the self-ensembling teacher model takes advantage of the transformation uncertainty and the appearance uncertainty [15] derived by Monte Carlo dropout [14] to heuristically adjust the regularization weights for each image pair during training (Fig. 1 (c)). Extensive experiments on a challenging intra-patient abdominal CT-MRI dataset show that our training strategy can promisingly advance the original learning-based method in terms of efficient hyperparameter tuning and a better tradeoff between accuracy and smoothness.

2 Methods

The proposed framework, as depicted in Fig. 1 (a), is designed on top of the Mean-Teacher (MT) architecture [20] which is constructed by a student registration network (Student RegNet) and a weight-averaged teacher registration network (Teacher RegNet). This architecture was originally proposed for semi-supervised learning, showing superior performance in further exploiting unlabeled data. We appreciate the MT-like design for registration because the temporal ensembling strategy can (i) help us efficiently exploit the temporal information across different training steps in such a ill-posed problem (Sec. 2.1), and (ii) enable smoother uncertainty estimation [27] to heuristically adjust the regularization weights during training (Sec. 2.2).

2.1 Mean-Teacher based Temporal Consistency Regularization

Specifically, the student model is a typical registration network updated by back-propagation, while the teacher model uses the same network architecture as the student model but its weights are updated from that of the student model via Exponential Moving Average (EMA) strategy, allowing to exploit the temporal information across the adjacent training steps. Formally, denoting the weights of the teacher model and the student model at training step k as $\theta_k^{'}$ and θ_k, respectively, we update $\theta_k^{'}$ as:

$$\theta_k^{'} = \alpha\theta_{k-1}^{'} + (1-\alpha)\theta_k, \qquad (2)$$

where α is the EMA decay and empirically set to 0.99 [20]. To exemplify our paradigm, we adopt the same U-Net architecture used in VoxelMorph (VM) [2] as the backbone network. 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. Skip connections between encoder and decoder are also applied. Given the predicted ϕ_s and ϕ_t from the student and the teacher, respectively, the Spatial Transformation Network (STN) [13] is utilized to warp the moving image I_m into I_{ws} and I_{wt}, respectively. Following the common practice, the dissimilarity between I_{ws} and I_f can be measured as the similarity loss $\mathcal{L}_{sim}(I_{ws}, I_f)$ with a spatial regularization term $\mathcal{L}_\phi$ for ϕ_s. To exploit the informative temporal information related to the ill-posedness, we encourage the temporal ensemble prediction of the teacher model to be consistent with that of the student model 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 deformation field at the current training step, $\mathcal{L}_c$ penalizes

the difference between predictions across adjacent training steps. Thus, we call $\mathcal{L}_c$ *temporal consistency regularization*.

As such, the cost function Eqn. 1 can be reformulated as the following Eqn. 3, which is a combination of similarity loss $\mathcal{L}_{sim}$, spatial regularization loss $\mathcal{L}_\phi$ and temporal consistency regularization loss $\mathcal{L}_c$:

$$\mathcal{L} = \mathcal{L}_{sim} + \lambda_\phi \mathcal{L}_\phi + \lambda_c \mathcal{L}_c, \tag{3}$$

where λ_ϕ and λ_c are the uncertainty guided adaptive tradeoff weights, as elaborated in the following Sec. 2.2. In order to handle multimodal abdominal registration, we use the Modality Independent Neighborhood Descriptor (MIND) [8] based dissimilarity metric for $\mathcal{L}_{sim}$. $\mathcal{L}_c$ is measured by mean squared error (MSE). Following the benchmark method [2], the choice of $\mathcal{L}_\phi$ is the generic L2-norm of the deformation field gradients for smooth transformation.

2.2 Double-Uncertainty Guided Adaptive Weighting

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

Double-Uncertainty Estimation.—Besides predicting the deformation ϕ_t, the teacher model can also serve as our uncertainty estimation branch since the model weight-averaged strategy can improve the stability of the predictions [20] that enables smoother model uncertainty estimation [27]. Particularly, we adopt the well-known Monte Carlo dropout [14] for Bayesian approximation due to its superior robustness [18]. We repetitively perform N stochastic forward passes on the teacher model with random dropout. After this step, we can obtain a set of voxel-wise predicted deformation fields $\{\phi_{t_i}\}_{i=1}^{N}$ with a set of warped images $\{I_{wt_i}\}_{i=1}^{N}$. We propose to use *the absolute value of the ratio of standard deviation to the mean*, which can characterize the normalized volatility of the predictions, to represent the uncertainty [19]. More specifically, the proposed registration uncertainties can be categorized into the transformation uncertainty and appearance uncertainty [15]. Formulating $\mu_{\phi_t}^c = \frac{1}{N}\sum_{i=1}^{N}\phi_{t_i}^c$ 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 cth channel of the deformation field (i.e., x, y, z displacements) and i denotes the ith forward pass, the transformation uncertainty map $U_{\phi_t} \in \mathbb{R}^{H \times W \times D \times 3}$ and the appearance uncertainty map $U_{I_{wt}} \in \mathbb{R}^{H \times W \times D}$ can be calculated as:

$$\begin{cases} \sigma_{\phi_t}^c = \sqrt{\dfrac{1}{N-1}\sum_{i=1}^{N}\left(\phi_{t_i}^c - \mu_{\phi_t}^c\right)^2} \quad \text{and} \quad U_{\phi_t}^c = \left|\dfrac{\sigma_{\phi_t}^c}{\mu_{\phi_t}^c}\right|, \\[3mm] \sigma_{I_{wt}} = \sqrt{\dfrac{1}{N-1}\sum_{i=1}^{N}\left(I_{wt_i} - \mu_{I_{wt}}\right)^2} \quad \text{and} \quad U_{I_{wt}} = \left|\dfrac{\sigma_{I_{wt}}}{\mu_{I_{wt}}}\right|. \end{cases} \tag{4}$$

With the guidance of U_{ϕ_t} and $U_{I_{wt}}$, we then propose to heuristically assign the weights λ_ϕ and λ_c of the spatial regularization $\mathcal{L}_\phi$ and the temporal consistency regularization $\mathcal{L}_c$ for each image pair during training.

Adaptive Weighting.—Firstly, for the typical spatial regularization $\mathcal{L}_\phi$, considering that unreliable predictions often correlate with biologically-implausible deformations [26], we assume that stronger spatial regularization can be given when the network tends to produce more uncertain predictions. As for the temporal consistency regularization $\mathcal{L}_c$, we notice that more uncertain predictions can be characterized as that this image pair is harder-to-align. Particularly, the most recent work [22] in semi-supervised segmentation combats with the assumption in [27] and experimentally reveals an interesting finding that emphasizing the unsupervised teacher-student consistency on those unreliable (often challenging) regions can provide more informative and productive clues for network training. Herein, we follow this intuition, i.e., more uncertain (difficult) samples should receive more attention (higher λ_c) for the consistency regularization $\mathcal{L}_c$. Formally, for each training step s, we update λ_ϕ and λ_c as follows:

$$\lambda_\phi(s) = k_1 \cdot \frac{\sum_v \mathbb{I}(U_{\phi_t}(s) > \tau_1)}{V_{U_{\phi_t}}} \quad \text{and} \quad \lambda_c(s) = k_2 \cdot \frac{\sum_v \mathbb{I}(U_{I_{wt}}(s) > \tau_2)}{V_{U_{I_{wt}}}}, \tag{5}$$

where $\mathbb{I}(\cdot)$ is the indicator function; v denotes the v-th voxel; $V_{U_{\phi_t}}$ and $V_{U_{I_{wt}}}$ 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 predictions. Noteworthily, the proposed strategy can work with any learning-based image registration architectures without increasing the number of trainable parameters. Besides, as shown in Fig. 1 (b), only the student model is utilized at the inference stage, which can ensure the computational efficiency.

3 Experiments and Results

Datasets.

We focus on the challenging application of abdominal CT-MRI multimodal registration for improving the accuracy of percutaneous nephrolithotomy. Under institutional review board approval, a 50-pair intra-patient abdominal CT-MRI dataset was collected from our partnership hospital with radiologist-examined segmentation masks for the region-of-interests (ROIs), including liver, kidney and spleen. We randomly divided the dataset into three groups for training (35 cases), validation (5 cases) and testing (10 cases), respectively. After sequential preprocessing steps including resampling, affine pre-alignment, intensity normalization and cropping, the images were processed into sub-volumes of $176 \times 176 \times 128$ voxels at the 1 mm isotropic resolution.

Implementation 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 so that each step contains an image pair. We set N

= 6 for the uncertainty estimation. We empirically set k_1 to 5 since the deformation fields with maximum $\lambda_\phi = 5$ are diffeomorphic in most cases. The scalar value k_2 is set to 1. Thresholds τ_1 and τ_2 are set to 10% and 1%, respectively. 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, respectively.

Comparison Study.

Table 1 and Fig. 2 present the quantitative and qualitative comparisons, respectively. We compare with several baselines, 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 [5]. As an alternative for adaptive weighting, we also include the recent conditional registration network [16] with the same scalar value 5 (denoted as VM (CIR-DM)). Fairly, we use the MIND-based dissimilarity metric in all learning-based methods. Although DIF-VM preserves better diffeomorphism properties, we find that its results are often suboptimal. Thus, we adopt VM as our backbone. For simplicity, we denote our adaptive spatial and temporal consistency regularization weighting strategy as VM (AS+ATC). Instead of training multiple models for finding the optimal fixed weight, only one model needs to be trained in both VM (CIR-DM) [16] and our VM (AS+ATC). Experimentally, training each VM model from scratch requires an average of 9.2h for this task. For the typical grid search scheme, five individual VM models are trained with varying fixed spatial regularization weights from 1 to 5, resulting in around 46h training time in total, wherein we observe that the overall best-performing VM model appears at $\lambda = 3$. As a comparison, the implicit control method, i.e., VM (CIR-DM), achieves comparable results with ~4.7x shorter total training time. Compared with VM (CIR-DM), our VM (AS+ATC) requires slightly longer training time due to the uncertainty estimation, yet, still resulting in ~4.2x faster than the grid search scheme. Distinct from VM (CIR-DM), we do not need to change any components in the network. Encouragingly, we find that VM (AS+ATC) further improves the registration accuracy in terms of Dice and ASD along with better properties of diffeomorphism and smoothness, implying that our strategy helps produce more desirable (smoother) solutions in this ill-posed problem. Visually, the structure boundaries registered by SyN and Deeds still have considerable disagreements, while the learning-based methods achieve more appealing boundary alignment. Besides, all trained models can infer an alignment in a second with a GPU.

Ablation Study.

To better understand our training strategy, we perform an ablation study with three variants: (i) VM (AS): removing the adaptive temporal consistency term; (ii) VM (S+TC): using empirical fixed weights $\lambda_\phi = 3$ and $\lambda_c = 0.5$; (iii) VM (AS+TC): adaptively adjusting λ_ϕ while λ_c is fixed as 0.5. The quantitative results are also presented in Table 1. Especially, similar to VM (CIR-DM), VM (AS) only adaptively adjusts λ_ϕ. We find that VM (AS) achieves even better performance compared with VM ($\lambda = 3$) and VM (CIR-DM),

highlighting that our uncertainty-guided weighting scheme enables more precise control of the spatial regularization strength during training. It can be also observed that both VM (S+TC) and VM (AS+TC) achieve better performance than VM (AS), demonstrating that additionally exploiting the temporal information can be rewarding. When we integrate these components into our synergistic training scheme, their better efficacy can be brought into play.

Visualized Uncertainty Map and Weighting Process.

Examples of two uncertainty maps U_{ϕ_t} and $U_{I_{wt}}$ are visualized in Fig. 3 (a) and (c), respectively. We observe that high uncertainty often occurs in hard-to-align ambiguous areas. Note that at the early stage, the weights are relatively small since limited transformation has been captured. As the training goes, the weights of $\mathcal{L}_\phi$ and $\mathcal{L}_c$ are adaptively modulated after each step (Fig. 3 (b) and (d)) assisted by the two uncertainties. Such scheme helps the model pursue a better tradeoff between accurate alignment and desirable diffeomorphism properties.

4 Conclusion

In this paper, we proposed a double-uncertainty guided spatial and temporal consistency regularization weighting strategy, assisted by a mean-teacher based registration framework. Besides temporal consistency regularization for further exploiting the temporal clues related to the ill-posedness, more importantly, the self-ensembling teacher model takes advantage of two estimated uncertainties to heuristically adjust the regularization weights for each image pair during training. Extensive experiments on abdominal CT-MRI registration showed that our strategy could promisingly advance the original learning-based method in terms of efficient hyperparameter tuning and a better tradeoff between accuracy and smoothness.

Acknowledgement

References

1. Avants BB, Epstein CL, Grossman M, Gee JC: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12 1, 26–41 (2008) [PubMed: 17659998]

2. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV: 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 SY, Kavan L, Whitaker RT: A cooperative autoen-coder 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) [PubMed: 9873908]

5. Dalca AV, Balakrishnan G, Guttag J, Sabuncu MR: Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 729–738. Springer (2018)

6. Dumoulin V, Shlens J, Kudlur M: A learned representation for artistic style. International Conference on Learning Representations (2016)

7. Ha D, Dai A, Le QV: Hypernetworks. arXiv preprint arXiv:1609.09106 (2016)

8. Heinrich MP, Jenkinson M, Bhushan M, Matin T, Gleeson FV, Brady M, Schnabel JA: MIND: Modality independent neighbourhood descriptor for multimodal deformable registration. Medical Image Analysis 16(7), 1423–1435 (2012) [PubMed: 22722056]

9. Heinrich MP, Jenkinson M, Brady M, Schnabel JA: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Transactions on Medical Imaging 32(7), 1239–1248 (2013) [PubMed: 23475350]

10. Hering A, Hansen L, Mok TC, Chung A, Siebert H, Häger S, Lange A, Kuckertz S, Heldmann S, Shao W, et al. : Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. arXiv preprint arXiv:2112.04489 (2021)

11. Hoopes A, Hoffmann M, Fischl B, Guttag J, Dalca AV: Hypermorph: Amortized hyperparameter learning for image registration. In: International Conference on Information Processing in Medical Imaging. pp. 3–17. Springer (2021)

12. Hu Y, Gibson E, Ghavami N, Bonmati E, Moore CM, Emberton M, Vercauteren T, Noble JA, Barratt DC: 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. Kendall A, Gal Y: What uncertainties do we need in Bayesian deep learning for computer vision? arXiv preprint arXiv:1703.04977 (2017)

15. 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)

16. Mok TC, Chung A: Conditional deformable image registration with convolutional neural network. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 35–45. Springer (2021)

17. Qin C, Wang S, Chen C, Qiu H, Bai W, Rueckert D: Biomechanics-informed neural networks for myocardial motion tracking in mri. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 296–306. Springer (2020)

18. Qu Y, Mo S, Niu J: DAT: Training deep networks robust to label-noise by matching the feature distributions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6821–6829 (2021)

19. Smith L, Gal Y: Understanding measures of uncertainty for adversarial example detection. arXiv preprint arXiv:1803.08533 (2018)

20. 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)

21. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I: A deep learning framework for unsupervised affine and deformable image registration. Medical Image Analysis 52, 128–143 (2019) [PubMed: 30579222]

22. Wu Y, Xu M, Ge Z, Cai J, Zhang L: Semi-supervised left atrium segmentation with mutual consistency training. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 297–306. Springer (2021)

23. Xu Z, Luo J, Yan J, Li X, Jayender J: F3RNet: Full-resolution residual registration network for deformable image registration. International Journal of Computer Assisted Radiology and Surgery 16(6), 923–932 (2021) [PubMed: 33939077]

24. 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)

25. Xu Z, Yan J, Luo J, Li X, Jagadeesan J: Unsupervised multimodal image registration with adaptative gradient Guidance. In: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1225–1229. IEEE (2021)

26. Xu Z, Yan J, Luo J, Wells W, Li X, Jagadeesan J: Unimodal cyclic regularization for training multimodal image registration networks. In: IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1660–1664. IEEE (2021)

27. Yu L, Wang S, Li X, Fu CW, Heng PA: Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 605–613. Springer (2019)
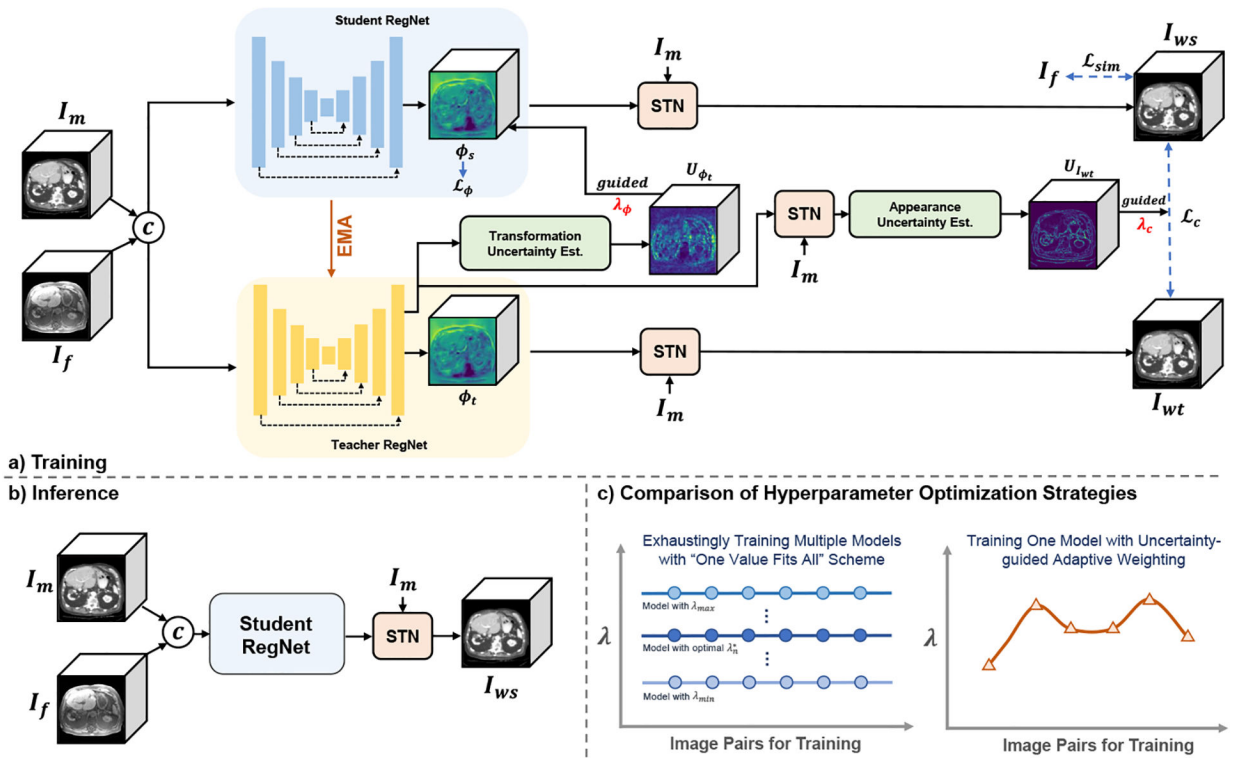
Fig. 1.

Illustration of the proposed framework with (a) training process, (b) inference process and (c) comparison of hyperparameter optimization strategies.
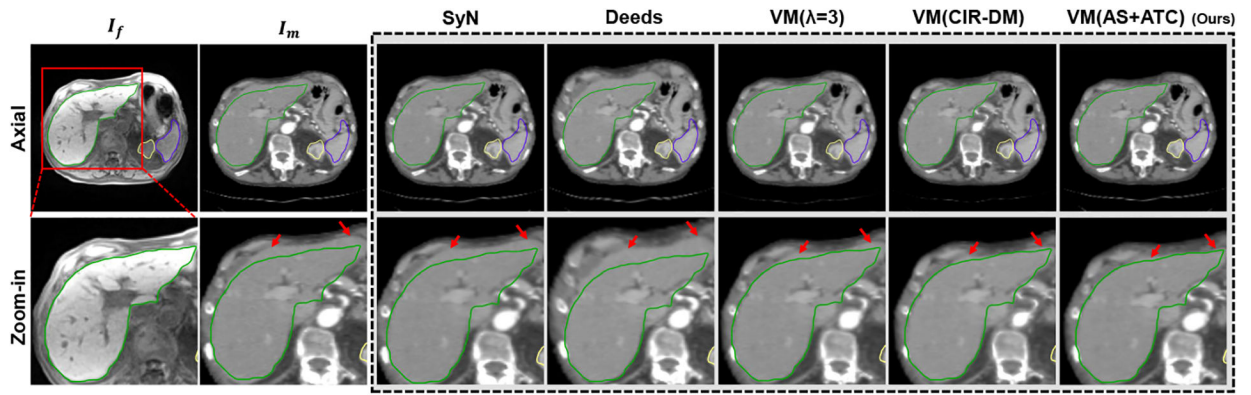
Fig. 2.
Exemplar axial slice of an abdominal CT-MRI registration case. The segmentation contours of the liver (green), kidney (yellow) and spleen (blue) extracted from the fixed abdominal MRI I_f are overlaid on all images. Better alignment drives structures closer to the fixed contours of I_f. The red arrows indicate the registration of interest around the organ boundary.
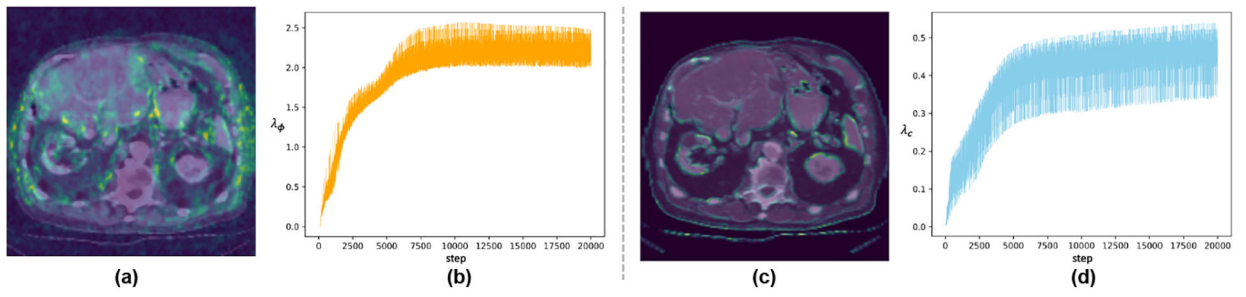
Fig. 3.

(a) and (c) are the examples of U_{ϕ_t} and $U_{I_{wt}}$ (overlay on the moving image), respectively, where brighter areas denote more uncertain regions. (b) and (d) show the adaptive weighting process across the training steps for λ_ϕ and λ_c, respectively.

Table 1.

Quantitative results for abdominal CT-MRI registration (mean±std).

Methods	Dice [%] ↑			ASD [voxel] ↓			$|J_\phi| \leq 0$	$\sigma(|J_\phi|)$
	Liver	Spleen	Kidney	Liver	Spleen	Kidney		
Initial	76.23±4.12	77.94±3.31	80.18±3.06	4.98±0.83	2.02±0.51	1.95±0.35	-	-
SyN	79.42±4.35	80.33±3.42	82.68±3.03	4.83±0.82	1.62±0.61	1.91±0.44	0.07%	0.42
Deeds	82.16±3.15	81.48±2.64	83.82±3.02	3.97±0.55	1.44±0.62	1.59±0.40	0.01%	0.28
DIF-VM	83.14±3.24	82.45±2.59	83.24±2.98	3.88±0.62	1.52±0.59	1.63±0.41	<0.001%	0.12
VM (λ = 1)	85.21±3.06	84.04±2.52	83.12±2.81	3.17±0.59	1.34±0.56	1.55±0.37	0.03%	0.19
VM (λ = 3) [†]	85.36±3.12	84.24±2.61	83.40±3.11	3.19±0.53	1.31±0.52	1.56±0.42	0.001%	0.14
VM (λ = 5)	84.28±2.93	83.71±2.40	82.96±2.77	3.83±0.68	1.48±0.59	1.65±0.44	<0.0001%	0.08
VM (CIR-DM)	85.29±3.39	84.17±2.59	83.01±3.06	3.32±0.41	1.33±0.47	1.49±0.46	0.002%	0.17
VM (AS+ATC) (ours)	**87.01±3.22**	**84.96±2.55**	**84.47±2.86**	**2.57±0.48**	1.24±0.49	**1.26±0.43**	<0.0005%	0.13
Ablation Study								
VM (AS)	86.02±3.02	84.75±2.59	83.44±2.96	3.21±0.49	1.33±0.50	1.45±0.49	0.0007%	0.12
VM (S+TC)	86.32±3.07	84.37±2.53	83.95±3.02	3.05±0.50	1.26±0.52	1.47±0.47	0.001%	0.15
VM (AS+TC)	86.49±3.13	84.87±2.68	84.03±2.89	2.88±0.53	1.20±0.44	1.32±0.40	0.0005%	0.14

Higher average Dice score (%) and lower ASD (mm) are better.

[†] indicates the best model via grid search.

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.