

Preserving Tumor Volumes for Unsupervised Medical Image Registration

Qihua Dong¹, Hao Du¹, Ying Song², Yan Xu^{3*}, Jing Liao^{1*}

¹Department of Computer Science, City University of Hong Kong

²National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital,
Chinese Academy of Medical Sciences and Peking Union Medical College

³School of Biological Science and Medical Engineering, Beihang University

dongqh078@gmail.com, haodu8-c@my.cityu.edu.hk, jingliao@cityu.edu.hk;

songying1770@hotmail.com

xuyan04@gmail.com

Abstract

Medical image registration is a critical task that estimates the spatial correspondence between pairs of images. However, current traditional and deep-learning-based methods rely on similarity measures to generate a deforming field, which often results in disproportionate volume changes in dissimilar regions, especially in tumor regions. These changes can significantly alter the tumor size and underlying anatomy, which limits the practical use of image registration in clinical diagnosis. To address this issue, we have formulated image registration with tumors as a constraint problem that preserves tumor volumes while maximizing image similarity in other normal regions. Our proposed strategy involves a two-stage process. In the first stage, we use similarity-based registration to identify potential tumor regions by their volume change, generating a soft tumor mask accordingly. In the second stage, we propose a volume-preserving registration with a novel adaptive volume-preserving loss that penalizes the change in size adaptively based on the masks calculated from the previous stage. Our approach balances image similarity and volume preservation in different regions, i.e., normal and tumor regions, by using soft tumor masks to adjust the imposition of volume-preserving loss on each one. This ensures that the tumor volume is preserved during the registration process. We have evaluated our strategy on various datasets and network architectures, demonstrating that our method successfully preserves the tumor volume while achieving comparable registration results with state-of-the-art methods. Our codes is available at: <https://dddraxxx.github.io/Volume-Preserving-Registration/>.

*Corresponding authors

1. Introduction

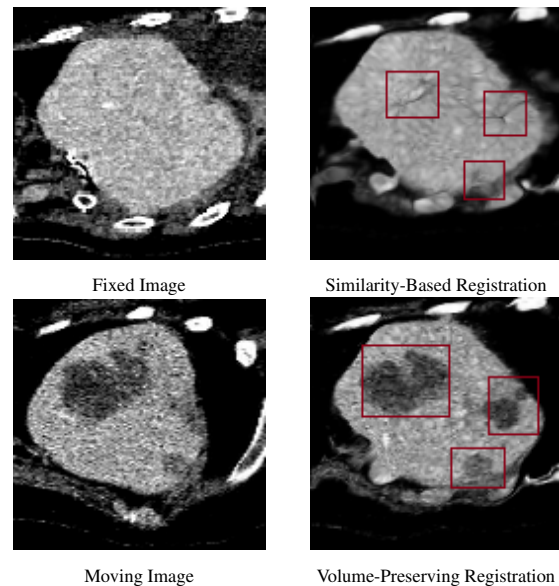


Figure 1. Red boxes represent the location of tumors in the moving image before warped. In image registration with tumors, similarity-based registration typically leads to significant alterations in tumor size while our volume-preserving registration is capable of preserving tumor anatomy.

Deformable image registration is a fundamental task that estimates non-linear spatial correspondences between two images. It is useful for medical image studies that involve 3D images of organs or tissues, such as MR brain scans and CT liver scans. Recently, a wide range of deep-learning-based methods are proposed in the field, with better performance and obvious speed-ups in inference time compared to the traditional registration methods [5, 38, 32, 14, 15].

Currently, most of the learning-based methods [5, 38] train a registration model to achieve high similarity in either

image intensity [38, 5] or anatomy label maps [14]. While this is useful in various medical applications, such as atlas-based segmentation and image fusion [8], problem emerges, however, when studying images with tumors [4, 12, 24]. Tracking tumor growth is a core task in cancer treatment, which can be used to evaluate the outcomes of radiotherapy and chemotherapy, and plan optimal postoperative treatment [12]. The process requires registration to align the anatomy of images from different periods of the patient, while preserving tumor properties *e.g.*, the size. Traditional models using regular registration have struggled to accomplish this task [12]. We have observed that this remains an issue for current learning-based methods, and a viable solution has yet to be proposed.

As shown in Figure 1, the volume of tumors significantly reduces when using popular unsupervised registration networks to align tumor images with atlas images. The issue is widespread since mainstream registration networks focus on optimizing the similarity between image pairs while ignoring tumor regions. Tumor regions often lack corresponding parts, and even for images of the same patient, the size and location of tumors may vary greatly. This can happen because tumors can change shape, size, location, or disappear over time. Therefore, in deformable settings, the size of tumors will change disproportionately to maximize the image similarity. This problem may be more severe when the registration network has better warping ability, leading to better performance in evaluations from previous works, as confirmed by our experiments in Section 4. The disproportionate change of tumor volume in deformable registration is lethal for clinical evaluation of tumor growth and greatly hinders the application of registration in clinics.

In this paper, we re-formulate the deformable registration problem as a constraint problem that preserves the properties of tumors in the original images, *i.e.*, the shape and the size of the tumor, while maximizing image similarity in non-tumor regions. For instance, in tumor image registration with atlas, besides the alignment of anatomical structures, the tumor should have similar morphing behaviours as its surrounding organs, *i.e.*, the size of it should change proportionally to the size of the organ. However, this poses two major challenges for our strategy. Firstly, due to the limited availability of annotated data, most existing registration methods resort to unsupervised learning. Therefore, a key question is how to identify tumor regions in the unsupervised setting. Second, after the tumor regions have been identified, the challenge still remains of preserving tumor volumes while simultaneously maximizing similarity in other regions. Finding a way to balance these two competing objectives is crucial for achieving effective tumor image registration.

To tackle the challenges, we propose a novel two-stage strategy that operates in an unsupervised manner. In the first

stage, we leverage a similarity-based registration network to identify regions that undergo excessive volume change during registration. Such regions are indicative of the presence of tumors, as tumors tend to exhibit larger changes in volume due to a lack of correspondence between image pairs. The outcome of the first stage is a soft tumor mask that distinguishes normal and possible tumor regions. In the second stage, we introduce an adaptive volume-preserving loss to train a volume-preserving registration network. Based on the soft tumor mask obtained from the first stage, this loss is designed to adaptively penalize volume changes in different regions of the image. Possible tumor regions are assigned larger loss weights, and similarity loss weights are adjusted accordingly in the opposite direction. The application of the adaptive volume-preserving loss effectively balances similarity and volume preservation in different parts of the image.

It should be noted that the soft masks from stage one do not need to be highly accurate to achieve desirable results, as our stage two is robust, which is demonstrated by our ablation experiments in section 4. In addition to evaluating the robustness of our proposed strategy on various datasets and frameworks, including CNN and transformer, as well as different types of imaging data such as brain and abdominal, MRI and CT images, we have also introduced the Square Tumor Size Ratio (STSR) metric to measure the preservation of tumor volumes. The results demonstrate that our methods achieve comparable warping performance metrics such as Dice coefficient and average landmark distances while effectively preserving tumor properties as measured by the STSR.

To conclude, our contribution includes: (1) We have re-formulated image registration with tumors as a constraint problem that preserves tumor volumes while maximizes image similarity in other normal regions. (2) We have designed a two-stage process that the first stage performs regular registration to estimate tumor regions, and the second stage performs volume-preserving registration on tumor regions based on the adaptive volume-preserving loss. (3) We have proposed the STSR metric to measure the preservation of tumor volumes in registration, and have evaluated the effectiveness of our strategy on different network architectures and various datasets.

2. Related Work

2.1. Learning-Based Deformable Image Registration

Compared to conventional registration methods, deep networks have achieved remarkable registration accuracy and speed using various network structures, *i.e.*, CNN and Transformer, in applications like atlas-image registration [29, 7, 16, 6, 32, 38, 5, 14, 15, 8]. They generally

adopts unsupervised learning, utilizing image similarity on either image intensities [29, 7, 6, 38] or synthesized label maps [14, 15] as well as smoothness regularization to supervise the registration network [7, 6, 14, 15, 38]. Advancements are often made through the introduction of novel regularization such as cycle consistency [16] and geometry preservation [1], or through innovative network structures like multi-scale architectures [28, 23]. However, existing methods do not adequately consider cases where tumors exist. Tumors often have no corresponding part between image pairs, and therefore, optimizing for similarity can excessively change their volumes, which limits the application of registration networks in critical clinical applications like cancer treatment. To the best of our knowledge, we are the first to study this problem with learning-based methods and propose a metric to measure volume preservation specifically for tumors.

2.2. Deformable Image Registration with Tumors

Conventional registration models often resolve registering images with tumor by (1) excluding the similarity measure on tumors [11, 18], which requires either massive manual segmentation [11] or initial seed to explicitly model tumor-growth [12, 18]. (2) jointly segmenting and registering images [9], which is time-consuming since it need to iteratively detect dissimilarity in image intensities and perform resection and retraction. (3) reconstructing quasi-normal images from pathological ones [13, 19], which is also slow since the reconstruction need to be iteratively improved.

Recently, learning-based methods for registration with tumors [24, 21, 33] are proposed for the release of BraTS-Reg dataset [4]. They aim to improve the spatial correspondence for non-tumor regions, thus either directly annotates landmarks to supervise registration [21] or introducing invertibility of the registration to mask non-inverted regions and ensure high correspondence in others [24, 33]. However, these methods assume the absence of tumors in one image of the registered pair, which may not be applicable in practical applications where we need to compare two tumor images from one patient to track the tumor growth. In this scenario, invertibility in tumors is still possible and thus, it is essential to consider the preservation of tumor properties during registration.

2.3. Unsupervised Tumor Segmentation

To ensure the preservation of tumor volumes during registration, it is crucial to first identify the locations of tumors. Current deep learning methods have made remarkable progress in unsupervised tumor segmentation, as demonstrated by recent studies [34, 35, 20]. These methods typically use GAN models or simulate tumors on healthy images to generate synthetic data, which is then used to

train the segmentation network. Although the well-trained network can achieve high segmentation accuracy, it often requires complex training strategies and additional models or testing steps to achieve the desired performance, as noted in recent research [35, 20]. However, in our approach, we do not rely on sophisticated segmentation methods to accurately segment tumors. Instead, our experiments, as presented in Section 4, demonstrate that our tumor mask estimation is sufficient for our strategy to learn a volume-preserving registration for tumors.

2.4. Volume-Preserving Deformable Image Registration

Volume-preserving registration has been studied extensively to improve the accuracy of anatomical registration. In pulmonary CT image registration, for example, tissue intensity changes between inspiration and expiration phases can be utilized as prior knowledge, and adding a volume-preserving constraint has been shown to be effective [31, 37]. In other cases, such as when soft tissues are incompressible or when images come from different modalities, a volume-preserving constraint can also serve as a helpful regularization [25, 30]. Our work, instead, focus on using volume-preservation to maintain the properties of tumors during image registration. The proposed strategy imposes this constraint adaptively on different parts of the images in order to preserve tumor volumes and ensure anatomical accuracy, while previous works have either preserved volumes for the entire images [25, 30] or can use simple methods to determine the regions to be preserved [31, 37].

3. Methodology

3.1. Preliminaries

Problem Setting. Given a moving image I_m and fixed image I_f defined over d -dimensional ($d=3$ in this paper) space Ω , deformable image registration aims to find a dense deformation field $\phi : \Omega \rightarrow \Omega$. This field warps I_m , meaning it projects every pixel position x in warped image I_w to I_m , to achieve alignment with I_f . In this paper, the task is to construct a registration network F_θ that takes input I_m, I_f to predict the deformation field ϕ .

$$\phi = F_\theta(I_m, I_f) \quad (1)$$

$$I_w(x) = I_m(\phi(x) + x) \quad (2)$$

$$I_w \sim I_f \quad (3)$$

Evaluation Criteria. In general, a successful deformable image registration involves accurately aligning the anatomical structures while also ensuring a smooth deformation field. This is often achieved through the use of similarity-based objectives and regularization methods that promote smoothness. However, in the context of image registration with tumors, simply aligning the anatomy is not

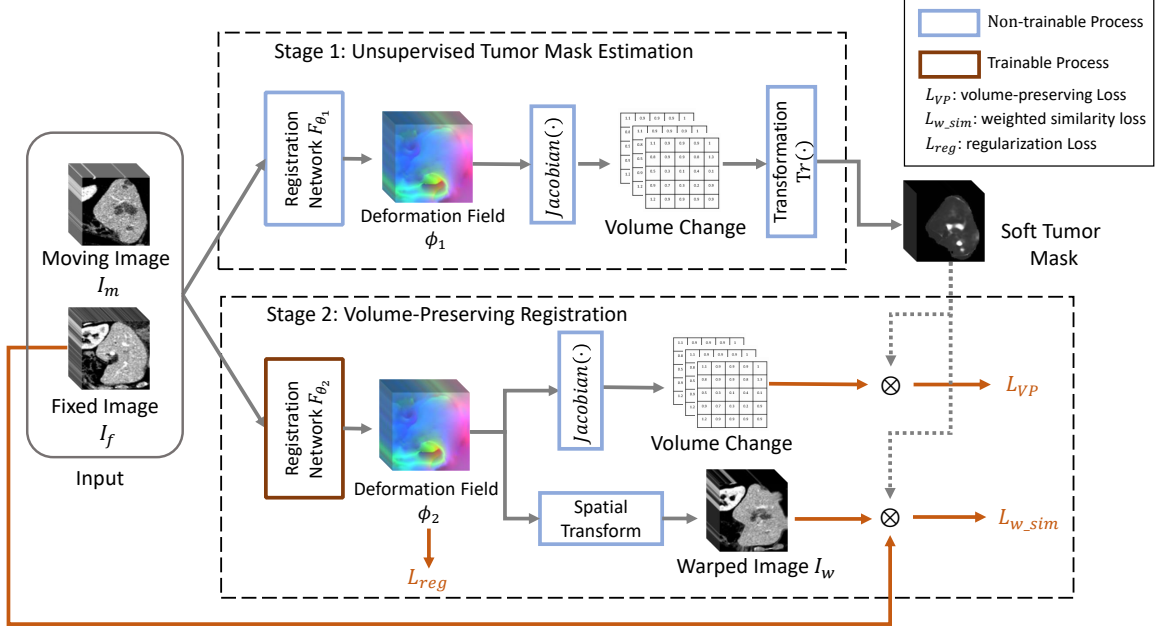


Figure 2. Main Framework. Our strategy consists of two stages. In the first stage, a soft mask indicating tumor regions is estimated by analyzing the Jacobian matrix of the deformation field obtained by a registration network pre-trained on similarity loss. In the second stage, the soft mask is used to guide the calculation of both volume-preserving and similarity losses for training the volume-preserving registration network. This deformation field estimated by the stage two ensures that both the image similarity and the preservation of tumor sizes.

sufficient. Preserving the volume of the tumors is crucial for accurately tracking tumor growth, as previously discussed. To be specific, given the moving image I_m with the tumor T_m and its surrounding organ O_m and the warped moving image I_w with T_w and O_w , the morphing of tumor T_m should be similar to that of O_m . Thus in respect of the volume preservation, the change of tumor size $|T_m|$ should be proportionate to the change of the containing organ $|O_m|$, *i.e.*:

$$TSR(I_m) = \frac{|T_m|}{|O_m|} \quad (4)$$

$$TSR(I_m) \approx TSR(I_w) \quad (5)$$

We therefore define the tumor size ratio (TSR) as $\frac{|T|}{|O|}$, which should be preserved after registration, and the re-formulated registration objective is now maximizing image similarity while preserving the volumes of tumors, *i.e.*, the TSR.

3.2. Overall Strategy

The proposed strategy consists of two stages, as shown in Figure 2. The first stage estimates potential tumor regions and generates a soft mask for them. This is done by analyzing volume changes, specifically the Jacobian matrix, in the deformation field obtained by an existing similarity-based registration network. As tumor regions tend to undergo excessive volume change during registration, they can

be identified through this analysis. In the second stage, we perform volume-preserving registration for tumors using an adaptive volume-preserving loss. This loss is designed to guide the volume-preserving process effectively and is associated with the tumor masks estimated in the first stage. Through the assignment of distinct weights to diverse regions, including both normal and tumor regions, it is possible to preserve tumor volumes while also ensuring a high degree of similarity in other regions.

Importantly, our strategy can be applied to nearly any learning-based registration network without introducing any additional modules.

3.2.1 Stage 1: Unsupervised Tumor Mask Estimation

Following previous works [5, 6, 38], we first train a similarity-based registration network F_{θ_1} with parameters θ_1 and fix its parameters. As shown in Figure 1, tumors in medical images can be detected by analyzing the change in volume through similarity-based registration. Our goal is to utilize this characteristic to facilitate the unsupervised estimation of the tumor mask for volume preservation during registration.

To determine the volume change at voxel x , we calculate the determinant of the Jacobian matrix of the deformation field (J_θ) with model parameter θ . Our objective is to ensure that the volume change in each point of the tumor region is

similar to the volume change of the organ, which is an indicator of the organ’s size change (as shown in Equation 6). To measure the distance between the two, we calculate the number of times one ratio exceeds another, denoted as D_θ , using Equations 8 and 7 where $D'_\theta(x)$ represents the relative volume change of voxel x with respect to organ areas.

$$\frac{|T_w|}{|O_w|} / \frac{|T_m|}{|O_m|} \approx 1 \Rightarrow \frac{|T_w|}{|T_m|} / \frac{|O_w|}{|O_m|} \approx 1, \text{ from Eq. 5, } (6)$$

$$D'_\theta(x) = |J_{\theta(x)}| / \frac{|O_w|}{|O_m|}, x \in \Omega, (7)$$

$$D_\theta(x) = \max(D'_\theta(x), 1/D'_\theta(x)), (8)$$

In practice, we estimate the organ masks O_m and O_w by warping a reference image with the ground truth (GT) organ segmentation. We randomly select the reference image from the training dataset, and the choice of reference image does not significantly impact the results because registration for organs achieves high accuracy, as evidenced by the dice value in table 1. By warping the reference image, we obtain the organ segmentation for the moving image O_m . We then calculate the organ segmentation for the warped image O_w .

To obtain a soft tumor mask STM from the distance D_{θ_1} defined in 8, we use a transformation function Tr . This function can be formulated as the following equation:

$$STM(x) = Tr(D_{\theta_1}(x)), (9)$$

Here, $D_{\theta_1}(x)$ represents the distance value at pixel x , and Tr transforms values from the range of $[0, \infty]$ to the range of $[0, 1]$. A value of 1 indicates the presence of a tumor, while a value of 0 indicates a normal region. In practical implementation, we use $Tr(x) = Sigm(5 \cdot (x - 1.5))$ since the sigmoid function is widely adopted for converting values to the range of $[0, 1]$ and to our experience, the size of tumor regions averagely changes around 1.5 times. However, the choice of transformation functions is not exclusive, as demonstrated in our experiments in Section 4. We also set the regions outside the organ mask to 0.

We observed that some normal regions, particularly those close to the organ boundary, may also exhibit some volume changes after registration. As a result, these regions may be incorrectly included in the tumor mask, as shown in Figure 3 (a). To address this issue, we implemented a pre-registration step to first align the edges and exclude volume changes caused by edge alignment. Specifically, in the pre-registration step, we applied a *bilateral filter* to the moving image, which preserves the edges while smoothing the interior regions. We then used the registration network to register the filtered moving image with the fixed image in order to align their edges. After this, an unfiltered warped image was obtained by spatially transforming the unfiltered moving image with the deformation field. The second registration process was then performed on the unfiltered warped

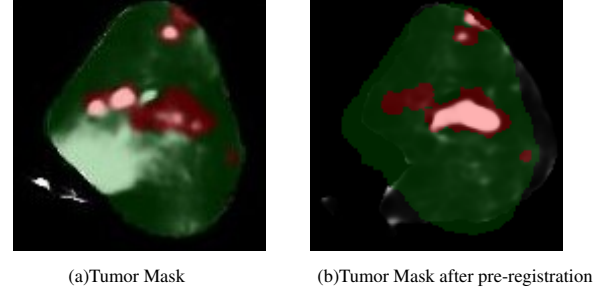


Figure 3. Visualization of the soft tumor mask. (a) is the estimated soft tumor mask without pre-registration and (b) is the one with pre-registration. The gray colors with intensity variations denote the soft mask, while the green and red colors represent the ground truth segmentation for the liver and tumors, respectively.

image, which focused on the registration of interior regions and was able to accurately estimate a tumor mask, as shown in Figure 3 (b). Therefore, this pre-registration approach effectively selects the tumor regions while filtering out the non-tumor regions.

3.2.2 Stage 2: Volume-Preserving Registration for Tumors

After obtaining the soft tumor mask STM , we can adaptively select tumor regions to maintain the size ratio. Here, we define the volume-preserving loss directly as the distance D_{θ_2} , as shown in Equation 8. By minimizing the distance, we can achieve smaller volume changes, indicating better preservation of tumor volume. The volume-preserving loss is multiplied by the soft tumor mask. This ensures that the volume-preserving constraints are adaptively applied to the tumor regions and not to the normal regions, which can be formulated as:

$$L_{VP} = \frac{1}{|I_m|} \sum_{x \in \Omega} D_{\theta_2}(x) \cdot STM(x), (10)$$

Here, in the second stage, the registration network takes the soft tumor mask as input, which serves as an indicator for the regions that need to be preserved in terms of volume. We follow the similarity and registration loss formulation in [38], but we use a weighted similarity loss to adapt to the volume-preserving loss, where Cov is the covariance between two images I_1, I_2 , \hat{I} is the mean of I , and Cov' is the adapted version utilizing STM weight:

$$Cov' [I_1, I_2] = \frac{\sum (I_1 - \hat{I}_1)(I_2 - \hat{I}_2)(1 - STM)}{\sum (1 - STM)}, (11)$$

$$L_{w.sim}(x) = \frac{Cov' [I_w, I_f]}{\sqrt{Cov[I_w, I_w]Cov [I_f, I_f]}}, (12)$$

Therefore, the training loss of the second stage is:

$$L = L_{w_sim} + \alpha_1 \cdot L_{VP} + \alpha_2 \cdot L_{reg}, \quad (13)$$

where we use the smoothness regularization from VoxelMorph [5] as L_{reg} , and α_1 and α_2 as hyperparameters to control the relative importance of the volume-preserving loss and the regularization loss.

4. Experiments

4.1. Experimental Settings

Implementation. Following the implementation of Recursive Cascaded Network (RCN) [38], we employed PyTorch and utilized the same similarity loss and regularization losses (for affine and deformable, respectively) for baseline methods. The weight of the volume-preserving losses is set to 0.1 while the weights of similarity loss is 1 and regularization loss is 0.1. Models were trained on one NVIDIA GeForce RTX 3090 GPU with a batch size of 4. During the training stage, we used the Adam optimizer [36] and ran for a total of 5 epochs, with 20000 iterations in each epoch. The learning rate was set to 10^{-4} .

Baselines. We re-implemented our strategy on four SOTA learning-based models that have distinct network architectures: 1) recursive cascaded network with VTN base network (VTN) [38], which stacks CNN networks to achieve higher registration accuracy; 2) VoxelMorph [5], which is a widely used CNN network that utilize stationary vector field to register images; 3) TransMorph [6], which is a well-performed hybrid Transformer-ConvNet network; 4) LapIRN [23], which is a Laplacian Pyramid network designed to operate in a coarse-to-fine manner. We fit these learning-based networks into our strategy with minor efforts and achieve good performance. Additionally, we compare our proposed method with two state-of-the-art traditional methods for deformable image registration, SyN [2] (integrated in ANTs [3] with the affine stage) and B-spline [26] (integrated in Elastix [17] with the affine stage) following [38].

Evaluation Metrics. The evaluation of the proposed strategy is composed of five metrics: Dice score, landmark distance to measure the anatomical alignment, and Square Tumor Size Ratio metric to measure the volume preservation of tumors. More specifically, the Dice score is calculated using the following formula:

$$Dice(I_1, I_2) = \frac{2|I_1 \cap I_2|}{|I_1| + |I_2|}, \quad (14)$$

Additionally, we also use landmark annotations to measure the anatomical alignment between the warped image and

the fixed image. We calculate the average distance between the fixed image’s landmarks and the warped landmarks of the moving image, as also introduced in RCN [38]. To quantify the diffeomorphism and smoothness of the deformation fields, we followed the methodology of previous works such as [7, 6, 39]. Here, two metrics were employed: the average percentage of voxels with non-positive Jacobian determinant ($|J_\theta| < 0(\%)$) in the deformation fields, and the standard deviation of the Jacobian determinant ($Std.|J_\theta|$).

To measure the volume preservation of tumors, we propose Square Tumor Size Ratio metric (STSR), which is calculated from the change of tumor size ratio (TSR) defined in Equation 4:

$$STSR = max\left(\frac{TSR(I_m)}{TSR(I_w)}, \left(\frac{TSR(I_w)}{TSR(I_m)}\right)^2\right), \quad (15)$$

Here, I_m, I_w refers to the moving image and the warped moving image. The square term is used to account for larger changes in tumor volumes, which can have a more significant impact on the accuracy of tumor growth estimation in clinical applications. By utilizing the STSR metric, we are able to accurately evaluate the performance of our proposed strategy.

Preprocessing. In our study, we follow standard preprocessing steps as outlined in VTN [39] and VoxelMorph [5]. The raw scans are resampled into $128 \times 128 \times 128$ voxels after cropping any unnecessary areas around the target object.

4.2. Datasets

Training. To train our model, we utilized two publicly available datasets: (1) LiTS17 [10]: The LiTS dataset comprises 201 CT scans that were acquired using various CT scanners and devices. The resolution of the images in this dataset ranges from 0.56mm to 1.0mm in the axial direction and from 0.45mm to 6.0mm in the z direction. The number of slices in the z direction ranges from 42 to 1026. (2) BraTS20 [22]: It consists of 369 MRI scans of brain tumors, each with ground truth segmentation of four different tumor components. For training our neural network, we utilized the T1ce modality and treated all four components as tumors, while considering the entire brain as the surrounding organ.

Testing. For testing, we utilized a set of 10 cases of CT liver scans with metastatic tumors, and 10 MRI brain scans with metastatic tumors. The modality of MRI brain scans is T1ce. Each case included both pre- and post-treatment images of the same patients, resulting in a total of 20 scans for each set. Following [27], the segmentation of the scans, which contains the organ (liver or brain) and the tumor inside, were carefully annotated by three experts in the field.

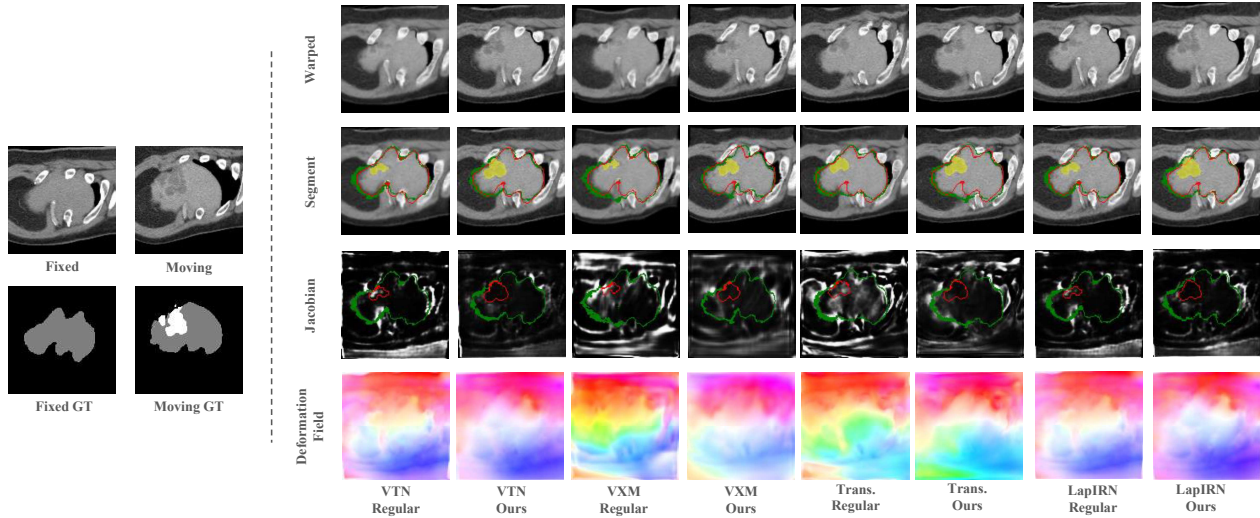


Figure 4. Qualitative comparison of different volume-preserving (VP) methods trained on the Liver Tumor Segmentation (LiTS) dataset. The left side of the figure shows two sets of images: Fixed and Ground Truth (GT), and Moving and GT. The first row of the figure displays the warped moving image, while the second row illustrates the organ outlines in green and red for the moving and fixed images, respectively. The yellow overlay highlights the tumors. Our proposed volume-preserving (VP) method ensures the preservation of tumor volume while aligning the images, as demonstrated by reduced number of visible changes in tumor size. In the third row, the Jacobian matrix of the deformation field is visualized. The green and red lines represent the organ and tumor outlines, respectively. The white areas indicate a large Jacobian, which corresponds to a more significant change in volume. The method without volume-preserving loss demonstrates a larger white area in the tumor, indicating a greater volume change of tumor volume. The last row of the figure displays the deformation field. Due to space limitations, qualitative results for BraTS20 dataset are provided in the supplemental material.

Network	Methods	LiTs17					BraTS20				
		Dice \uparrow	Lm. Dist \downarrow	$ J_\theta < 0(\%) \downarrow$	Std. $ J_\theta \downarrow$	STSR \downarrow	Dice \uparrow	Lm. Dist \downarrow	$ J_\theta < 0(\%) \downarrow$	Std. $ J_\theta \downarrow$	STSR \downarrow
SyN	-	0.850	12.36	0.192	0.221	1.564	0.9560	4.01	0.276	0.226	1.512
B-Spline	-	0.853	14.22	0.003	0.288	1.330	0.9601	4.51	0.001	0.171	1.622
VTN	Regular	0.912	10.46	3.545	2.048	2.330	0.9784	2.98	0.396	0.725	2.263
	Ours	0.908	10.77	2.956	1.124	1.260	0.9738	3.10	0.596	0.285	1.416
VXM	Regular	0.860	12.62	4.191	1.699	1.738	0.9701	3.56	0.024	0.448	1.929
	Ours	0.857	12.43	1.681	0.691	1.241	0.9741	3.60	0.030	0.226	1.463
TransMorph	Regular	0.859	13.35	6.377	1.906	1.785	0.9710	4.01	0.164	0.468	1.868
	Ours	0.856	13.19	4.084	0.827	1.302	0.9758	3.98	0.019	0.194	1.319
LapIRN	Regular	0.8933	10.23	5.301	1.788	2.360	0.9750	3.40	3.020	0.301	1.889
	Ours	0.8893	10.55	4.120	0.988	1.645	0.9743	3.36	0.040	0.105	1.224

Table 1. Quantitative comparison with state-of-the-art registration methods. Our strategy was applied to two datasets, namely LiTS17 and BraTS20, and three network architectures. In addition, five metrics were employed for the evaluation process. The results of the comparison indicate that our proposed strategy can be well generalized to other state-of-the-art methods to preserve tumor volumes while achieving comparable or better performance on other metrics.

Landmarks were also annotated by three experts in the field to evaluate the registration accuracy of our strategy.

4.3. Main Results

Quantitative Comparison. In Table 1, we present a comprehensive comparison of our proposed strategy with two traditional methods and three learning-based methods with different network architectures (CNN, cascaded CNN and hybrid transformer-CNN). We compare the methods using a similarity-based manner (regular) following the settings

in [38], as well as our volume-preserving manners. Our results demonstrate that our proposed strategy outperforms all other methods in terms of the STSR metrics, while achieving comparable results in Dice and Landmark Dist, and often better results in folding detection ($|J_\theta| < 0$) and smoothness quality ($Std. |J_\theta|$). Furthermore, we observe that registration networks that achieve better performance in terms of previous metrics such as Dice tend to have worse performance in preserving tumor volumes (STSR). This observation highlights the need to balance these two objectives

#	Network	Unsupervised	Adaptive	Region	LiTs17				
					Dice \uparrow	Lm. Dist \downarrow	$ J_\theta < 0(\%) \downarrow$	Std. $ J_\theta \downarrow$	STSR \downarrow
1		✓	–	–	0.912	10.46	3.545	2.048	2.330
2		×	–	–	0.914	10.22	3.241	1.165	1.648
3	VTN	×	×	Tumor	0.911	10.74	2.883	1.017	1.205
4		✓	×	Organ	0.904	11.07	3.194	1.015	1.640
5		✓	✓	Organ	0.908	10.77	2.956	1.124	1.260
6		✓	–	–	0.860	12.62	4.191	1.699	1.738
7		×	–	–	0.863	12.22	1.603	0.691	1.417
8	VXM	×	×	Tumor	0.859	12.38	1.567	0.698	1.291
9		✓	×	Organ	0.856	12.78	1.635	0.675	1.407
10		✓	✓	Organ	0.857	12.43	1.681	0.691	1.241
11		✓	–	–	0.859	13.35	6.337	1.906	1.785
12		×	–	–	0.863	13.10	4.923	1.540	1.540
13	TransMorph	×	×	Tumor	0.858	13.21	4.401	0.877	1.227
14		✓	×	Organ	0.849	13.02	4.651	0.845	1.454
15		✓	✓	Organ	0.856	13.19	4.084	0.827	1.302

Table 2. Effectiveness of adaptive volume-preserving loss in unsupervised registration frameworks on LiTs17 dataset. Three networks, namely VTN in [38], VXM in [5], and TransMorph in [6], were applied. The methods used in the study were regular unsupervised (first row), weighted similarity loss only based on the ground truth (GT) mask (second row), volume-preserving loss based on GT(third row), volume-preserving loss on the whole organ based on estimation (fourth row), and on the tumor based on estimation (ours).

in order to achieve competitive results.

Qualitative Comparison. Figure 5 presents a qualitative comparison of our proposed strategy with different state-of-the-art methods. Comparing the “VTN/VXM/Trans./LapIRN regular” column and the “VTN/VXM/Trans./LapIRN ours” column, it is evident that our method outperforms other methods in terms of volume preservation in the warped image (1st row) and the corresponding deformation field (4th row). Furthermore, in terms of visualization of the registration accuracy and the overlay of the tumor (2nd row), our proposed method achieves better volume preservation with comparable accuracy compared to other methods. Lastly, the Jacobian determinant of the deformation field (3rd row) reveals that regular methods tend to have a larger Jacobian in tumor regions, indicating greater volume change, compared to our method. This highlights the effectiveness of our volume-preserving approach in preserving tumor volumes while aligning the images.

4.4. Ablation study

Volume-Preserving Losses. In Table 2, we present the results of our evaluation of the effectiveness of various versions of volume-preserving losses. For each sub-table, the first row (#1, #6, and #11) represents using only the similarity loss, while the second row (#2, #7, and #12) utilizes the ground truth as a weighted similarity loss. The third row (#3, #8, and #13) uses the ground truth binary mask as a tumor mask, while the fourth row (#4, #9, and #14) uses the mask of the entire organ as a tumor mask. The fifth row (#5, #10, and #15) represents our proposed strategy.

By comparing the first and second rows of each sub-table with the fifth row, we observe that our proposed volume-

preserving loss can effectively preserve details and achieve superior performance across various metrics. The improvement from the third and fourth rows to the fifth row indicates the importance of adaptiveness in preserving the volume of the organ and tumor, respectively. It should be noted that we do not require pixel-wise ground truth for tumor volume preservation.

Robustness to Tumor Mask Estimation. In Figure 5, we showcase the robustness of our proposed strategy in the presence of noisy ground truth segmentation masks. To evaluate this, we randomly select points within organs to ensure that tumors have a certain Dice score with respect to the ground truth segmentation of tumors. As depicted in the figure, despite the variations in the noisy ground truth segmentation masks, our strategy’s performance remains consistent. Specifically, when the tumor mask has a Dice score above a certain threshold, such as 0.15, the estimated tumor mask consistently achieves a Dice score of approximately 0.17. In contrast, using the whole organ as mask in volume-preserving loss causes degradation in performance as it only has a dice score of roughly 0.05. This highlights the necessity and efficiency of our tumor mask estimation in stage one.

5. Conclusions

In summary, this work addressed the problem of preserving tumor volumes while maximizing similarity during learning-based registration, which is crucial for tracking tumor growth. We proposed a two-stage unsupervised strategy that effectively preserves the size ratio of tumors, without requiring extra networks. Our approach involves generating a soft tumor mask in the first stage using a similarity-based registration network and incorporating

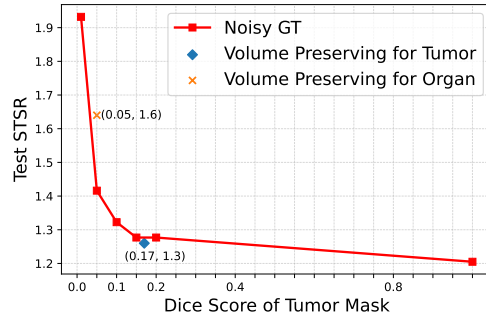


Figure 5. Robustness of our tumor mask estimation. We evaluated the robustness of our tumor mask estimation by running our strategy using RCN on the LiTS datasets with different noisy ground truth (GT) masks as tumor masks, each with specific Dice scores with tumors. We plotted the results on a graph with the Dice score on the x-axis and the STSR in the test set on the y-axis (where smaller values are better).

an adaptive volume-preserving loss and a weighted similarity loss in the second stage to improve registration performance. To evaluate our strategy’s performance, we introduced a new metric, the Square Tumor Size Ratio (STSR), which measures the preservation of tumor volume. Our proposed strategy was evaluated on various networks and demonstrated superior performance in tumor volume preservation, while achieving comparable results in other metrics. This highlights the effectiveness of our approach in tracking tumor growth and its ability to generalize well to different registration networks. Additionally, we noted that the volume preservation approach could be extended to other lesions besides tumors.

Acknowledgement: This work was supported by the HKSAR Innovation and Technology Commission (ITC) under ITF Project MHP/109/19, the National Natural Science Foundation in China under Grant 62022010, the Beijing Natural Science Foundation Haidian District Joint Fund in China under Grant L222032 and the Fundamental Research Funds for the Central Universities of China from the State Key Laboratory of Software Development Environment in Beihang University in China.

References

- [1] Moab Arar, Yiftach Ginger, Dov Danon, Amit H Bermano, and Daniel Cohen-Or. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13410–13419, 2020. [3](#)
- [2] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008. [6](#)
- [3] Brian B Avants, Nick Tustison, Gang Song, et al. Advanced normalization tools (ants). *Insight j*, 2(365):1–35, 2009. [6](#)
- [4] Bhakti Baheti, Diana Waldmannstetter, Satrajit Chakrabarty, Hamed Akbari, Michel Bilello, Benedikt Wiestler, Julian Schwarting, Evan Calabrese, Jeffrey Rudie, Syed Abidi, et al. The brain tumor sequence registration challenge: establishing correspondence between pre-operative and follow-up mri scans of diffuse glioma patients. *arXiv preprint arXiv:2112.06979*, 2021. [2](#), [3](#)
- [5] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9252–9260, 2018. [1](#), [2](#), [4](#), [6](#), [8](#)
- [6] Junyu Chen, Eric C Frey, Yufan He, William P Segars, Ye Li, and Yong Du. Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis*, 82:102615, 2022. [2](#), [3](#), [4](#), [6](#), [8](#)
- [7] Jiashun Chen, Donghuan Lu, Yu Zhang, Dong Wei, Munan Ning, Xinyu Shi, Zhe Xu, and Yefeng Zheng. Deformer: Towards displacement field learning for unsupervised medical image registration. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 141–151. Springer, 2022. [2](#), [3](#), [6](#)
- [8] Xiang Chen, Yan Xia, Nishant Ravikumar, and Alejandro F Frangi. Joint segmentation and discontinuity-preserving deformable registration: Application to cardiac cine-mr images. *arXiv preprint arXiv:2211.13828*, 2022. [2](#)
- [9] Nicha Chitphakdithai and James S. Duncan. Non-rigid registration with missing correspondences in preoperative and postresection brain images. [3](#)
- [10] Patrick Christ. Lits – liver tumor segmentation challenge (lits17). 2017. [6](#)
- [11] Olivier Clatz, Hervé Delingette, Ion-Florin Talos, Alexandra J. Golby, Ron Kikinis, Ferenc A. Jolesz, Nicholas Ayache, and Simon K. Warfield. Robust nonrigid registration to capture brain shift from intraoperative mri. *IEEE Transactions on Medical Imaging*, 2005. [3](#)
- [12] Ahmed Elazab, Yousry M Abdulazeem, Ahmed M Anter, Qingmao Hu, Tianfu Wang, and Baiying Lei. Macroscopic cerebral tumor growth modeling from medical images: A review. *IEEE Access*, 6:30663–30679, 2018. [2](#), [3](#)
- [13] Xu Han, Xiao Yang, Stephen R. Aylward, Roland Kwitt, and Marc Niethammer. Efficient registration of pathological images: A joint pca/image-reconstruction approach. *arXiv: Computer Vision and Pattern Recognition*, 2017. [3](#)
- [14] Malte Hoffmann, Benjamin Billot, Douglas N Greve, Juan Eugenio Iglesias, Bruce Fischl, and Adrian V Dalca. Synthmorph: learning contrast-invariant registration without acquired images. *IEEE transactions on medical imaging*, 41(3):543–558, 2021. [1](#), [2](#), [3](#)
- [15] Malte Hoffmann, Andrew Hoopes, Douglas N Greve, Bruce Fischl, and Adrian V Dalca. Anatomy-aware and acquisition-agnostic joint registration with synthmorph. *arXiv preprint arXiv:2301.11329*, 2023. [1](#), [2](#), [3](#)
- [16] Boah Kim, Dong Hwan Kim, Seong Ho Park, Jieun Kim, June-Goo Lee, and Jong Chul Ye. Cyclemorph: cycle consistent unsupervised deformable image registration. *Medical image analysis*, 71:102036, 2021. [2](#), [3](#)
- [17] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2009. [6](#)
- [18] Dongjin Kwon, Marc Niethammer, Hamed Akbari, Michel Bilello, Christos Davatzikos, and Kilian M. Pohl. Portr: Pre-operative and post-recurrence brain tumor registration. *IEEE Transactions on Medical Imaging*, 2014. [3](#)
- [19] Dongjin Kwon, Ke Zeng, Michel Bilello, and Christos Davatzikos. Estimating patient specific templates for pre-operative and follow-up brain tumor registration. *Lecture Notes in Computer Science*, 2015. [3](#)
- [20] Fei Lyu, Mang Ye, Andy J Ma, Terry Cheuk-Fung Yip, Grace Lai-Hung Wong, and Pong C Yuen. Learning from synthetic ct images via test-time training for liver tumor segmentation. *IEEE transactions on medical imaging*, 41(9):2510–2520, 2022. [3](#)
- [21] Mingyuan Meng, Lei Bi, Dagan Feng, and Jinman Kim. Brain tumor sequence registration with non-iterative coarse-to-fine networks and dual deep supervision. *arXiv preprint arXiv:2211.07876*, 2022. [3](#)
- [22] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. [6](#)
- [23] Tony CW Mok and Albert CS Chung. Large deformation diffeomorphic image registration with laplacian pyramid networks. [3](#), [6](#)
- [24] Tony CW Mok and Albert CS Chung. Unsupervised deformable image registration with absent correspondences in pre-operative and post-recurrence brain tumor mri scans. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 25–35. Springer, 2022. [2](#), [3](#)
- [25] Torsten Rohlfing, Calvin R Maurer, David A Bluemke, and Michael A Jacobs. An alternating-constraints algorithm for volume-preserving non-rigid registration of contrast-

- enhanced mr breast images. In *Biomedical Image Registration: Second International Workshop, WBIR 2003, Philadelphia, PA, USA, June 23-24, 2003. Revised Papers 2*, pages 291–300. Springer, 2003. 3
- [26] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999. 6
- [27] Siyuan Shan, Wen Yan, Xiaoqing Guo, Eric I Chang, Yubo Fan, Yan Xu, et al. Unsupervised end-to-end learning for deformable medical image registration. *arXiv preprint arXiv:1711.08608*, 2017. 6
- [28] Shuwei Shao, Zhongcai Pei, Weihai Chen, Wentao Zhu, Xingming Wu, and Baochang Zhang. A multi-scale unsupervised learning for deformable image registration. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–10, 2022. 3
- [29] Jiacheng Shi, Yuting He, Youyong Kong, Jean-Louis Coatrieux, Huazhong Shu, Guanyu Yang, and Shuo Li. Xmorpher: Full transformer for deformable medical image registration via cross attention. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 217–226. Springer, 2022. 2, 3
- [30] Christine Tanner, Julia A. Schnabel, Andreas Degenhard, Andy D. Castellano-Smith, Carmel Hayes, Martin O. Leach, D. R. Hose, Derek L. G. Hill, and David J. Hawkes. Validation of volume-preserving non-rigid registration: Application to contrast-enhanced mr-mammography. *Lecture Notes in Computer Science*, 2002. 3
- [31] Di Wang, Yue Pan, Oguz C. Durumeric, Joseph M. Reinhardt, Eric A. Hoffman, Joyce D. Schroeder, and Gary E. Christensen. Plosl: Population learning followed by one shot learning pulmonary image registration using tissue volume preserving and vesselness constraints. *Medical Image Analysis*, 2022. 3
- [32] Yibo Wang, Wen Qian, Mengqi Li, and Xuming Zhang. A transformer-based network for deformable medical image registration. In *Artificial Intelligence: Second CAAI International Conference, CICA 2022, Beijing, China, August 27–28, 2022, Revised Selected Papers, Part I*, pages 502–513. Springer, 2022. 1, 2
- [33] Marek Wodzinski, Artur Jurgas, Niccolo Marini, Manfredo Atzori, and Henning Muller. Unsupervised method for intra-patient registration of brain magnetic resonance images based on objective function weighting by inverse consistency: Contribution to the brats-reg challenge. *arXiv preprint arXiv:2211.07386*, 2022. 3
- [34] Qingsong Yao, Li Xiao, Peihang Liu, and S Kevin Zhou. Label-free segmentation of covid-19 lesions in lung ct. *IEEE transactions on medical imaging*, 40(10):2808–2819, 2021. 3
- [35] Xiaoman Zhang, Weidi Xie, Chaoqin Huang, Ya Zhang, Xin Chen, Qi Tian, and Yanfeng Wang. Self-supervised tumor segmentation with sim2real adaptation. *IEEE Journal of Biomedical and Health Informatics*, 2023. 3
- [36] Zijun Zhang. Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pages 1–2. Ieee, 2018. 6
- [37] Bowen Zhao, G. Christensen, Joo Hyun Song, Yue Pan, Sarah E. Gerard, Joseph M. Reinhardt, Kaifang Du, Taylor J. Patton, John E. Bayouth, and Geoffrey D. Hugo. Tissue-volume preserving deformable image registration for 4dct pulmonary images. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 481–489, 2016. 3
- [38] Shengyu Zhao, Yue Dong, Eric I Chang, Yan Xu, et al. Recursive cascaded networks for unsupervised medical image registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10600–10610, 2019. 1, 2, 3, 4, 5, 6, 7, 8
- [39] Shengyu Zhao, Tingfung Lau, Ji Luo, I Eric, Chao Chang, and Yan Xu. Unsupervised 3d end-to-end medical image registration with volume tweening network. *IEEE journal of biomedical and health informatics*, 24(5):1394–1404, 2019. 6