



# Anatomically constrained and attention-guided deep feature fusion for joint segmentation and deformable medical image registration

Hee Guan Khor <sup>a</sup>, Guochen Ning <sup>a</sup>, Yihua Sun <sup>a</sup>, Xu Lu <sup>b</sup>, Xinran Zhang <sup>a</sup>, Hongen Liao <sup>a,\*</sup>

<sup>a</sup> Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China

<sup>b</sup> Graduate School at Shenzhen, Tsinghua University, Shenzhen, 518055, Guangdong, China



## ARTICLE INFO

### Keywords:

Anatomical constraint  
Deep learning  
Deformable image registration  
Multi task learning

## ABSTRACT

The main objective of anatomically plausible results for deformable image registration is to improve model's registration accuracy by minimizing the difference between a pair of fixed and moving images. Since many anatomical features are closely related to each other, leveraging supervision from auxiliary tasks (such as supervised anatomical segmentation) has the potential to enhance the realism of the warped images after registration. In this work, we employ a Multi-Task Learning framework to formulate registration and segmentation as a joint issue, in which we utilize anatomical constraint from auxiliary supervised segmentation to enhance the realism of the predicted images. First, we propose a Cross-Task Attention Block to fuse the high-level feature from both the registration and segmentation network. With the help of initial anatomical segmentation, the registration network can benefit from learning the task-shared feature correlation and rapidly focusing on the parts that need deformation. On the other hand, the anatomical segmentation discrepancy from ground-truth fixed annotations and predicted segmentation maps of initial warped images are integrated into the loss function to guide the convergence of the registration network. Ideally, a good deformation field should be able to minimize the loss function of registration and segmentation. The voxel-wise anatomical constraint inferred from segmentation helps the registration network to reach a global optimum for both deformable and segmentation learning. Both networks can be employed independently during the testing phase, enabling only the registration output to be predicted when the segmentation labels are unavailable. Qualitative and quantitative results indicate that our proposed methodology significantly outperforms the previous state-of-the-art approaches on inter-patient brain MRI registration and pre- and intra-operative uterus MRI registration tasks within our specific experimental setup, which leads to state-of-the-art registration quality scores of 0.755 and 0.731 (i.e., by 0.8% and 0.5% increases) DSC for both tasks, respectively.

## 1. Introduction

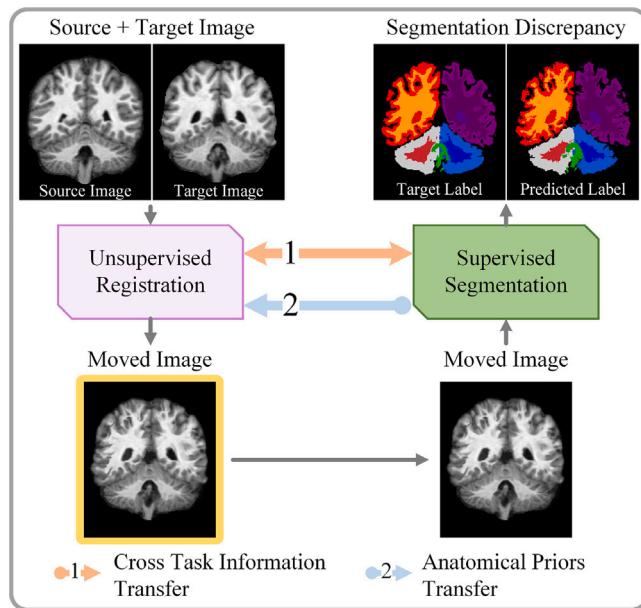
Deformable image registration (DIR) has been widely used in clinical diagnosis, prognosis, disease monitoring, and surgical navigation, with the goal of producing deformation fields that warps the source (moving) image and align it with the corresponding target (fixed) image. By evaluating voxel-wise similarities, traditional deformable registration approaches such as Symmetric Normalization (Syn) (Avants et al., 2008) and NiftyReg (Modat et al., 2010) typically formulate the deformable registration as a complex optimization problem that requires intensive computational power. The development of learning-based approaches, which enable them to take advantage of the powerful feature learning capability of deep Convolutional Neural Networks (CNN) (Kuckertz et al., 2020; Nielsen et al., 2019), has significantly advanced image registration tasks in recent deep learning methodologies. As a result, by enabling accurate alignment of a new pair

of volumes with less computing time, the deep learning-based image registration approaches achieved state-of-the-art registration accuracy. However, learning-based medical image registration methodologies typically require considerable supervised information, such as anatomical landmarks or ground truth deformation fields. Since acquiring a large-scale medical dataset with such comprehensive annotations is labor-intensive, this unavoidably limits the clinical feasibility of the supervised learning method.

Recently, unsupervised learning-based registration has the potential to increase the similarity between a pair of moving and fixed volumes by integrating prior knowledge (about image modalities, anatomy, and structure) into the registration process (Mansilla et al., 2020) or implementing smoothness constraints on the deformation fields (Balakrishnan et al., 2019). Furthermore, stationary velocity fields (SVFs)

\* Corresponding author.

E-mail address: [liao@tsinghua.edu.cn](mailto:liao@tsinghua.edu.cn) (H. Liao).



**Fig. 1.** We propose to use supervised anatomical segmentation (green) to improve deformable registration (purple) quality scores under the multi-task learning setup. We explicitly learn the task-shared feature correlation (orange) between registration and segmentation and use it to improve the realism of the warped images after final registration. We use the anatomical constraints (blue) approximated by segmentation prediction discrepancy of the initial moved and fixed image to further improve the realism of our final predicted warped moving images. The proposed anatomy-aware unsupervised registration method can largely improve the realism of the images after deformation.

with an effective scaling-and-squaring strategy are mainly used in learning-based diffeomorphic models (Dalca et al., 2019; Mok and Chung, 2020) since it is relatively simple to implement.

In recent years, motivated by the success of multi-task learning (MTL) (Elmahdy et al., 2021), auxiliary tasks (such as anatomical segmentation) have been increasingly used to help the information transfer. Auxiliary tasks have been demonstrated to be advantageous for the main registration task (Vandewinckel et al., 2020) since they are highly correlated with the global anatomical features of the organs (i.e., shape and topology). Existing works (Balakrishnan et al., 2019; Mansilla et al., 2020) primarily integrates pixel-level similarity measure defined on the warped moving- and fixed-segmentation masks as a weakly supervised approach to train an unsupervised registration network but do not take global anatomical priors into account as they are highly domain-specific. For diffeomorphic registration and to reduce the number of foldings, the scaling-and-squaring strategy (Li et al., 2021; Dalca et al., 2019) ought to theoretically guarantee diffeomorphic transformation with a sufficient number of integration steps. However, they observed that there were almost “no non-negative Jacobian voxels” in the works that were given, and that 0.023% to 0.151% of the voxels had a negative Jacobian determinant. This is due to two key aspects, as discussed in Qiu et al. (2021): (i) The velocity field could not be smooth enough. By increasing the regularization weight, this problem can be resolved. However, the registration accuracy usually degrades as a result; (ii) There were not enough integration steps provided. While doing so will increase the computational cost, it will also lead to fewer foldings. Inspire by the concept of anatomically constrained neural networks (Oktay et al., 2017) and MTL (Elmahdy et al., 2021), we propose leveraging the global non-linear representations of image anatomy acquired during the segmentation task to constrain the registration process.

The additional supervised anatomical segmentation can facilitate us to explicitly learn the correlation between tasks to improve the final deformable registration accuracy (i.e., alignment between a pair

of fixed and moving images in terms of Dice Similarity Coefficient (DSC) and Structural Similarity Index (SSIM) scores). The learning of the correlation is motivated by the fact that the correlation between tasks is more invariant across anatomical priors than the individual viewpoints. As discussed in the previous works (Huo et al., 2019), medical image registration and segmentation are complementary and interrelated (i.e., moving annotation labels warped by image registration algorithms are frequently used for image segmentation (Huo et al., 2019; Wang et al., 2014), and boundaries of anatomical structures can be used to guide the image registration method in addition to the intensity of the images (Elmahdy et al., 2019b,a), both of which are complementary), and it may be advantageous to use both simultaneously and collaboratively. These task-robust correlations between registration and segmentation have the potential to largely improve the target medical image registration accuracy in the presence of high anatomical variability.

Approaches that enable the training of deep segmentation networks with a small number of labeled data are therefore highly desirable in order to address the issue of having huge labeled datasets, which is often extremely labor and time consuming. To deal with a scarcity of labeled training data, data augmentation or semi-supervised strategies are widely used. However, many of the present methodologies’ augmentation strategies are either hand-engineered or involve time-consuming searches. Olut et al. (2020) proposed learning a statistical deformation model that could capture realistic anatomical variations from unlabeled data via deep registration models in order to create adversarial samples subject to a particular transformation model. Shen et al. (2020) also utilized a similar concept. Subspace-based shape models, despite their popularity and successes, have a number of significant drawbacks, such as the need for one-to-one shape correspondences across the training data, the inadequacy of the affine subspace representation, or the requirement for a large and diverse training population in order to infer flexible yet specific models with adequate generalization capabilities.

In this paper, we focus on enhancing the information flow between the supervised segmentation network and the unsupervised registration in the Joint Registration and Segmentation (JRS) framework. We leverage the anatomical connection between the two tasks to enhance the registration accuracy. To incorporate anatomical priors in the form of global constraints into the deformable learning process, we propose that such a correlation can be leveraged in two ways. On the one hand, we propose to explicitly learn the task feature correlation between registration and segmentation. This is achieved by using our newly designed Cross-Task Attention Block to model the interaction and complementarity between registration and segmentation features. The target registration task can leverage the global anatomical context flow acquired from segmentation to generate more realistic and anatomically plausible images after warping. The correlation, on the other hand, is being utilized to enhance the deformation field generated by the registration network and to enforce anatomical correspondences between the deformed moving and fixed images. We approximate the anatomical constraint by calculating the discrepancy between the ground-truth fixed annotation labels and the labels predicted by the segmentation network from the warped images after initial registration. Since registration and segmentation are coupled, we assume that the estimated anatomical constraint can be transferred from segmentation to registration. We propose to use this relation to guide the deformation field refinement on the target registration task. Combining the two ways of correlation exploitation leads to our proposed Anatomical Correlation-Aware Deformable Medical Image Registration (AC-DMiR) approach. In Fig. 1, we demonstrate the proposed AC-DMiR to utilize the correlation.

For training, the non-linear and compact representation of the anatomy correlated with medical images can be learned collaboratively between these tasks, however, the networks can be employed individually during the testing phase. This enables the prediction of only the

registration output, when the segmentation labels are not accessible during testing. We demonstrate the effectiveness of our proposed approach on the inter-patient brain Magnetic Resonance Imaging (MRI) registration and pre- and intra-operative uterus MRI registration tasks, on which we achieve new state-of-the-art registration quality scores. Our contributions are summarized as follows:

1. We propose a novel Joint Registration and Segmentation framework that effectively utilizes supervised anatomical segmentation available on initial warped moving images to improve the accuracy of final deformable registration.
2. Specifically, we explicitly learn the correlation between registration and segmentation tasks and maximize the information flow between the two tasks. Furthermore, we improve the anatomical plausibility of the final warped moving images by using the anatomical constraint approximated by segmentation map prediction discrepancy.
3. Furthermore, we compare AC-DMiR against various state-of-the-art registration algorithms. Extensive quantitative and qualitative results demonstrate that AC-DMiR outperforms other registration algorithms and achieved superior registration accuracy on the inter-patient brain MRI and pre- and intra-operative uterus MRI registration tasks.

## 2. Related works

Some researchers have studied combining image registration with segmentation for the purpose of generating anatomical plausible results. Several studies (Lu et al., 2011; Pohl et al., 2006) proposed employing a Bayesian framework to model the association between these tasks such that both would restrict one other. Yezzi et al. (2003) proposed a geometric, variational framework for fusing segmenting and registering features from multiple images using active contours. Unal and Slabaugh (2005) proposed an approach for combined segmentation and registration in infinite dimensions by joining two partial differential equations (PDEs) for surface and deformation field evolutions to enable finding a locally optimum solution efficient. However, the framework proposed in Unal and Slabaugh (2005) is generalized in that shape priors are not essential, and the algorithm can execute without a training phase as long as a seed is placed someplace near the target region of interest. For cardiac perfusion images, research in Mahapatra et al. (2015) proposed a JRS framework to decompose temporal intensity images into sparse and low-rank components. They proposed segmenting using the sparse component and registering images with the low-rank component.

However, these conventional alternatives or joint optimization approaches have the apparent limitations of requiring extensive parameter tuning, high computational power and slow run-time for each image pair, making them insurmountable for some time-critical clinical applications. To overcome the above-mentioned limitations, deep learning-based joint optimization methods have been developed. Deep learning-based networks have recently achieved astounding success in a variety of fields, including medical image analysis (Cao et al., 2018; Fu et al., 2020; Fechter et al., 2017; Kiljunen et al., 2020; Liu et al., 2019), where deep learning models are able to compete with or even perform better than medical experts in some tasks (Hu et al., 2019; Maidens and Slamon, 2018; Tschandl et al., 2019; Ardila et al., 2019). For JRS, various deep learning-based techniques have been developed. There are two main types of joining frameworks in the literature: (1) joining through the loss function, and (2) joining via the architecture and also the loss function. Hu et al. (2018) proposed integrating segmentation and registration using a multi-resolution DSC loss function as an exemplar of the first strategy. Elmahdy et al. (2019a) proposed a methodology that is a combination of learning and iterative techniques, wherein the bladder is segmented by a CNN and transferred to an iterative-based registration approach. The authors included an

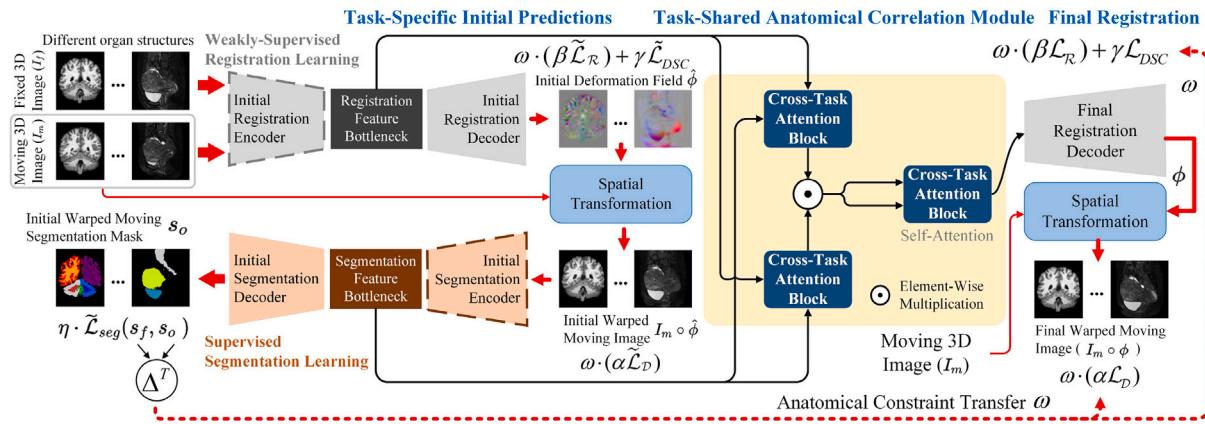
additional registration process to focus on the seminal vesicles and rectum, as well as domain-specific features such as air pocket inpainting and contrast clipping. Furthermore, research published in Elmahdy et al. (2019b) and Mahapatra et al. (2018) proposed a Generative Adversarial Network (GAN)-based strategy for joint registration and segmentation, in which a generative network predicts the correspondence between fixed and moving images and a discriminator network provides guidance on the quality of the deformed contours.

On the other hand, Xu and Niethammer (2019) provided a framework that concurrently trains a registration and a segmentation network as an exemplar of the second technique. The author suggested that these tasks be learned together throughout training, however the networks are deployed individually during testing. When the anatomical annotations are not accessible during the testing phase, the abovementioned training strategy provides a way that just the registration output to be predicted. Furthermore, Estienne et al. (2019) proposed a collaborative framework that incorporated affine and deformable registration as well as segmentation in a 3D end-to-end CNN network, in which the proposed technique can produce linear and elastic deformation models and efficiently register and segment medical volumes. The hybrid deep learning network for both tasks was relatively well-designed (Li et al., 2019), and it consistently improved as evaluated by longitudinal changes in the white matter tract. Liu et al. (2020) recently proposed Joint Synthesis, Segmentation, and Registration (JSSR) methodology, an end-to-end framework for registering and segmenting multi-modal images concurrently. The proposed JSSR framework consists of a generator, a registration and a segmentation network. First, JSSR converts source domain images to target domain images prior implementing intra-modal registration on the synthesized and target images. The segmentation module will then be used to provide adequate guidance based on semantic correspondences on the synthetic and target images, which is then being regularized by utilizing supervision from another fully-annotated dataset. The requirements for ground-truth segmentation have also been excluded from the segmentation part of the Bayesian inference with probabilistic atlas experiment (Qiu and Ren, 2021), and the concept of jointly learning segmentation and registration has been expanded to multi-modal image registration (Chen et al., 2021b).

However, the abovementioned approaches explored the idea of joining segmentation and registration, where to the best of our knowledge none have explored the idea of exploiting the correlation between the main registration task and segmentation task. To exploit this correlated anatomical information, with a view to enhance the accuracy of main registration task, we employ the JRS framework to explore the feasibility of utilizing supervised anatomical segmentation to better comprehend the spatial correspondence between the moving and fixed images. In comparison to prior works that combined unsupervised registration with supervised anatomical segmentation, we explicitly address the challenge of high anatomical variability among fixed and moving images, as well as investigate the method to best correlate both tasks, and how to optimize the anatomical knowledge transfer between these tasks.

## 3. Methodology

During clinical applications, radiologists typically incorporate related information from previous scans to compensate for patient misalignment and anatomical variations in organ position, which is important in improving the accuracy of the treatment. Meanwhile, to reduce treatment related side-effects, radiologists have to delineate organs that vary substantially in shape and volume between treatment fractions. As the coupling of image registration and segmentation is vital, this demonstrates how knowledge embedded in one task can be acquired by other tasks, demonstrating how it can be advantageous to learn related tasks together.



**Fig. 2.** The network architecture of our proposed Anatomical Correlation-Aware Deformable Medical Image Registration (AC-DMiR) approach, which combines the proposed task-shared anatomical correlation module with improved image realism via voxel-wise anatomical constraint transfer. The task-shared feature correlation module integrates the registration and segmentation features to explicitly fully comprehend the anatomical correlation between both tasks and provide additional information for the final registration process. Furthermore, as illustrated on the right-hand side of the figure, the anatomical plausibility of the warped moving images is refined during the training phase based on the anatomical constraint approximated by the segmentation prediction discrepancy.

Since it is extremely crucial to produce anatomically plausible results for ensuring high accuracy of clinical diagnosis and treatment, it is common for JRS approaches to combine image segmentation and registration in such a way that these tasks constrain one another during training. Existing work leverages the JRS framework by joining via the loss function (Hu et al., 2018), as well as the architecture and loss function (Xu and Niethammer, 2019; Elmahdy et al., 2019a) to enhance the effectiveness and realism of registration approaches. For the first strategy, local voxel-level loss functions (Ferrante et al., 2018, 2017) do not incorporate the comprehensive global context and thus do not necessarily correlate with greater anatomical plausibility. Our approach works in tandem with the latter strategy. We take advantage of supervised segmentation and task correlations to improve information flow between the two tasks and produce more realistic images after deformation.

In this work, we propose using a supervised segmentation network to learn anatomical prior-knowledge. As shown in Fig. 1, using supervised segmentation allows us to exploit the correlation between both tasks to improve the anatomical mapping between warped moving images and fixed images even more. First, as described in Section 3.2.3, we designed Cross-Task Attention Block to learn the inter-task feature correlation between registration and segmentation and transfer it to the target registration task. In our implementation, we utilized a joint registration and segmentation framework that is optimized with a joint loss during the learning process. Furthermore, we proposed an anatomical constraint strategy by calculating the discrepancy between the initial predictions of the segmentation network and the fixed label map. As registration and segmentation are naturally coupled, we make use of the anatomical constraint not only in the segmentation but also in the registration network as a loss function. This encourages the network parameters to converge to a point with high anatomical plausibility for the final warped moving image, which will be described in Section 3.2.5.

### 3.1. Anatomical correlation-aware architecture

We adapt our correlation-aware AC-DMiR framework to take advantage of the information about the spatial relationship between anatomical structures for registration and segmentation tasks leveraging recent advancements of the MTL framework (Xu et al., 2018; Vandenbende et al., 2020). The proposed approach is depicted in Fig. 2. The task-specific initial registration and segmentation are modeled using two hybrid Transformer-CNN backbone networks (Chen et al., 2022) for generating initial predictions. The next step is to use a

task-shared anatomical correlation module to explicitly learn the correlation between registration and segmentation, as well as to incorporate the complementary information from the other task to improve final registration predictions.

### 3.2. Task-specific initial predictions

#### 3.2.1. Unsupervised registration learning

Initial predictions are generated first in order to learn the correlation between registration and segmentation information subsequently. Let  $I_f, I_m$  be the fixed and moving images defined over an 3D spatial domain  $\Omega \subset \mathbb{R}^3$ . We assume that  $I_f$  and  $I_m$  were affinely aligned as part of the preprocessing, providing nonlinear misalignment as the only source of misalignment between the volumes. For initial registration, our registration network aims to compute an initial deformation field  $\hat{\phi} : \Omega \subset \mathbb{R}^3$  that aligns  $I_m$  to  $I_f$  on the field of view  $\Omega$  such that  $I_f(\mathbf{p})$  and  $[I_m \circ \hat{\phi}](\mathbf{p})$  are similar for each voxel  $\mathbf{p} \in \Omega$ . The network architectures with the loss and regularization functions are described in detail in the following sections.

*Image Similarity Measure.* Since the images are from the same imaging modality and have a similar intensity distribution, we utilized two similarity metrics for  $\tilde{\mathcal{L}}_D$ . The first is the Mean Squared Error (MSE) loss (i.e.,  $\tilde{\mathcal{L}}_{MSE}$ ), which measures the mean squared voxel-wise difference between  $I_f$  and  $I_m \circ \hat{\phi}$ :

$$\tilde{\mathcal{L}}_{MSE}(I_f, I_m \circ \hat{\phi}) = \frac{1}{|\Omega|} \sum_{\mathbf{p} \in \Omega} [I_f(\mathbf{p}) - [I_m \circ \hat{\phi}](\mathbf{p})]^2 \quad (1)$$

The second is the local normalized cross-correlation (LNCC) between  $I_f$  and  $I_m \circ \hat{\phi}$ . Let  $\bar{I}_f(\mathbf{p})$  and  $[\bar{I}_m \circ \hat{\phi}](\mathbf{p})$  be the mean voxel value within the  $n^3$  sized local window (we utilize  $n = 9$  in our work) centered on voxel  $\mathbf{p}$ , the LNCC loss ( $\tilde{\mathcal{L}}_{LNCC}$ ) of  $I_f$  and  $I_m \circ \hat{\phi}$  can be formulated as:

$$\begin{aligned} \tilde{\mathcal{L}}_{LNCC}(I_f, I_m \circ \hat{\phi}) = \\ \sum_{\mathbf{p} \in \Omega} \frac{\left( \sum_{\mathbf{p}_i} (I_f(\mathbf{p}_i) - \bar{I}_f(\mathbf{p})) ([I_m \circ \hat{\phi}](\mathbf{p}_i) - [\bar{I}_m \circ \hat{\phi}](\mathbf{p})) \right)^2}{\left( \sum_{\mathbf{p}_i} (I_f(\mathbf{p}_i) - \bar{I}_f(\mathbf{p}))^2 \right) \left( \sum_{\mathbf{p}_i} ([I_m \circ \hat{\phi}](\mathbf{p}_i) - [\bar{I}_m \circ \hat{\phi}](\mathbf{p}))^2 \right)} \end{aligned} \quad (2)$$

*Deformation Field Regularization.* Increasing the similarity score alone would encourage  $I_m \circ \hat{\phi}$  to be as similar to  $I_f$  as feasible visually. However, the resulting deformation field  $\hat{\phi}$  might not be smooth or realistic. To impose smoothness and continuity in the deformation fields, we introduce regularizer  $\tilde{\mathcal{L}}_R(\hat{\phi})$  to the loss function.  $\tilde{\mathcal{L}}_R(\hat{\phi})$  facilitates the displacement value at a location to be nearly equivalent to the values in its nearby locations. In this work, we experimented with

three regularizers for  $\tilde{\mathcal{L}}_R(\hat{\phi})$ . We first leveraged a diffusion regularizer (Balakrishnan et al., 2019) on the spatial gradients of displacement  $\mathbf{u}$ :

$$\tilde{\mathcal{L}}_{smooth}(\hat{\phi}) = \sum_{\mathbf{p} \in \Omega} \| \nabla \mathbf{u}(\mathbf{p}) \|^2 \quad (3)$$

where we approximate spatial gradients using differences between neighboring voxels. For  $\nabla \mathbf{u}(\mathbf{p}) = \left( \frac{\partial \mathbf{u}(\mathbf{p})}{\partial \{x,y,z\}} \right)$ , we approximate  $\frac{\partial \mathbf{u}(\mathbf{p})}{\partial \{x,y,z\}} \approx \mathbf{u}(p_{\{x,y,z\}} + 1) - \mathbf{u}(p_{\{x,y,z\}})$

The second regularizer, bending energy, was proposed in Rueckert et al. (1999) and it penalizes sharply curved deformations, making it helpful for abdominal organ deformable prediction (i.e., uterus deformation due to bladder-and-rectum filling). The formula for bending energy, which acts on the second derivative of the displacement field  $\mathbf{u}$ , is as follows:

$$\tilde{\mathcal{L}}_{bending}(\hat{\phi}) = \sum_{\mathbf{p} \in \Omega} \| \nabla^2 \mathbf{u}(\mathbf{p}) \|^2 \quad (4)$$

where the derivatives were computed using the same forward differences as before. Despite the fact that the smooth regularization from Eqs. (3) and (4) encourages smooth and continuous deformation, foldings may still occur as negative values of the Jacobian determinant, which are apparently not physically or anatomically possible. To prevent foldings, we aim to minimize the distance measure  $\tilde{\mathcal{L}}_D$  and the regularizer  $\tilde{\mathcal{L}}_R$  while maintaining a positive Jacobian determinant for each voxel  $\mathbf{p}$  in  $\Omega$ . To achieve this, we utilized an anti-fold loss  $\tilde{\mathcal{L}}_{anti-fold}$  proposed by Rühaak et al. (2017), which is defined as:

$$\tilde{\mathcal{L}}_{anti-fold}(\hat{\phi}) = \sum_{\mathbf{p} \in \Omega} \mathfrak{V}(\det(\nabla \mathbf{u}(\mathbf{p}))) \quad (5)$$

where:

$$\mathfrak{V}(v) = \begin{cases} \frac{(v-1)^2}{v}, & \text{if } v > 0 \\ +\infty, & \text{otherwise} \end{cases} \quad (6)$$

The input to  $\mathfrak{V}$  in Eq. (6) is substituted with  $v = \det(\nabla \mathbf{u}(\mathbf{p}))$  in Eq. (5), where  $\mathfrak{V}(v)$  refers to an index function that is utilized to penalize the gradient of the flow at regions with foldings and is minimized when  $v = 1$ . As a result, the  $\tilde{\mathcal{L}}_{anti-fold}$  provides two regularizing effects: (i) it inhibits folding by keeping the Jacobian determinants positive, and (ii) by biasing the optimization to preserve the same volume,  $\tilde{\mathcal{L}}_{anti-fold}$  prevents both shrinkage and expansion in  $\hat{\phi}$ , which is beneficial in registering medical images with significant spatial displacements (i.e., the changes of volume in bladder during uterine fibroid ablation).

*Auxiliary Data Loss Function.* A good deformation field  $\hat{\phi}$  can predict anatomical correspondences with high accuracy, which implies that the same corresponding anatomical structure in  $I_m \circ \hat{\phi}$  and  $I_f$  should overlap well. We use the Dice score (Dice, 1945) to calculate the volume overlap for segmentation masks and define  $\tilde{\mathcal{L}}_{DSC}$  as:

$$\tilde{\mathcal{L}}_{DSC}(s_f, s_{\hat{m}}) = 1 - \frac{1}{K} \sum_{k=1}^K \frac{2 \cdot \sum_{\mathbf{p} \in \Omega} s_f^k(\mathbf{p}) \cdot s_{\hat{m}}^k(\mathbf{p})}{\sum_{\mathbf{p} \in \Omega} (s_f^k(\mathbf{p}))^2 + \sum_{\mathbf{p} \in \Omega} (s_{\hat{m}}^k(\mathbf{p}))^2} \quad (7)$$

where  $s_f : \Omega \rightarrow [0, 1]^k$  and  $s_{\hat{m}} : \Omega \rightarrow [0, 1]^k$  denote the one-hot representation of the fixed and warped moving segmentation mask over all structures  $k \in [1, K]$ . An implementation of Eq. (7) directly may not be adequate to auto-differentiation networks since anatomical labels are categorical, making it incorrect to compute  $s_{\hat{m}}$  with a naive implementation of linear interpolation. We employed an approach similar to that proposed in Balakrishnan et al. (2019), in which  $s_f$  and  $s_{\hat{m}}$  were built as image volumes with  $K$  channels, each channel having a binary mask describing the segmentation of a single organ. The gradients of  $\tilde{\mathcal{L}}_{DSC}$  can then be backpropagated into the network using  $s_{\hat{m}}$ , which is computed by warping the  $K$ -channel  $s_m$  with  $\hat{\phi}$  using linear interpolation. We then multiply and add  $s_f$  and  $s_{\hat{m}}$  to determine the numerator and denominator of Eq. (7). The tilde in  $\tilde{\mathcal{L}}$  denotes that this is the initial prediction loss function.

### 3.2.2. Supervised segmentation learning

Given that target organ deformation is substantially correlated with surrounding organ properties, we consider leveraging anatomical information to constrain the final registration process. The input to the segmentation network is the output of the initial registration network, referred to as  $I_m \circ \hat{\phi}$ , where the segmentation network predicts the corresponding output label mask,  $s_o(\mathbf{p})$  (i.e.,  $s_o : \Omega \rightarrow [0, 1]^k$  denote the one-hot representation of the predicted segmentation mask of all structures  $k \in [1, K]$  from  $I_m \circ \hat{\phi}$ ). The DSC-based segmentation loss,  $\tilde{\mathcal{L}}_{seg}$ , which quantifies the overlap between the prediction  $s_o$  and the ground truth label  $s_f$  as follows, was used to train the segmentation network:

$$\tilde{\mathcal{L}}_{seg}(s_f, s_o) = 1 - \frac{1}{K} \sum_{k=1}^K \frac{2 \cdot \sum_{\mathbf{p} \in \Omega} s_f^k(\mathbf{p}) \cdot s_o^k(\mathbf{p})}{\sum_{\mathbf{p} \in \Omega} (s_f^k(\mathbf{p}))^2 + \sum_{\mathbf{p} \in \Omega} (s_o^k(\mathbf{p}))^2} \quad (8)$$

where the tilde in  $\tilde{\mathcal{L}}$  denotes the initial prediction loss function for the segmentation network. Different from the group-wise registration framework proposed in Li et al. (2021), our registration network is a learning-based pairwise medical image registration framework for subject-based deformable prediction. In our work, segmentation maps predicted by the initial segmentation decoder are compared with the ground-truth annotations of  $s_f$  for optimizing the anatomical plausibility of the final registered image, which is different from Li et al. (2021) that warp the obtained segmentation from the warped moving images back to the image native space to minimize the segmentation loss.

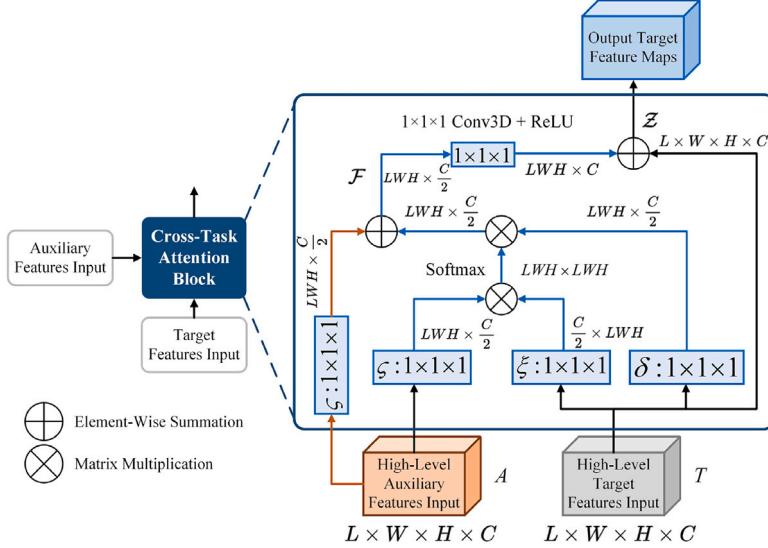
### 3.2.3. Task-shared anatomical correlation module

To learn the correlation between registration and segmentation, the registration and segmentation high-level features from the initial phase are fed into a task-shared anatomical correlation module. This is accomplished through the implementation of the proposed Cross-Task Attention Block (CTAB), which captures the mutual relationship and incorporates bottleneck features of high-level representations from both initial registration and segmentation networks. The design of the anatomical correlation module was inspired by research in the area of multi-task learning (Vandenhende et al., 2020), where comparable attention modules were employed to facilitate the joint learning of various tasks. Unlike the proposed non-local attention module in Wang et al. (2018), which computes self-attention on a single image, the proposed CTAB attempts to model anatomical correspondences between high-level features from two networks performing different tasks. The proposed CTAB structure is depicted in Fig. 3.

Let  $T \in \mathbb{R}^{L \times W \times H \times C}$  and  $A \in \mathbb{R}^{L \times W \times H \times C}$  represents the high-level target and auxiliary input feature maps of the block, respectively. The tensor shape of each 3D feature map after flattening is denoted as  $L$ ,  $W$  and  $H$ . The cross-task feature attention computed by CTAB can be formulated as:

$$f_i = \frac{\sum_{\forall j} \exp \left( \zeta(a_i)^T \xi(t_j) \right) \delta(t_j)}{\sum_{\forall j} \exp \left( \zeta(a_i)^T \xi(t_j) \right)} \quad (9)$$

where  $a_i$  and  $t_j$  are features from  $A$  and  $T$  at location  $i$  and  $j$  respectively,  $\zeta(\cdot)$ ,  $\xi(\cdot)$ , and  $\delta(\cdot)$  represent linear embedding layer. In Eq. (9), the function  $\exp(\cdot)$  computes the anatomical correlation between the high-level features of these two locations,  $a_i$  and  $t_j$ . The output  $f_i$  is a normalized comprehensive feature on all  $T$  locations, weighted by their correlations with the high-level auxiliary features on location  $i$ . Hence, the matrix  $\mathcal{F}$ , which is composed by  $f_i$ , contains non-local information from  $A$  to every position in  $T$ . Finally, to ensure effective back-propagation, the output of the proposed CTAB,  $\mathcal{Z}$ , is formulated as the element-wise summation of  $\mathcal{F}$  and  $T$ . As a result, the feature of a location  $k$  in  $\mathcal{Z}$  encapsulates anatomical correlation between the entire high-level target features and location  $k$  of the auxiliary feature map,



**Fig. 3.** The proposed Cross-Task Attention Block.

as well as information from the original high-level target feature map at  $k$ .

As shown in Fig. 2, given the registration and segmentation bottlenecks features,  $\mathcal{R}_B$  and  $S_B$  respectively, the output of task-shared anatomical correlation module  $\mathcal{W}_Q$  are calculated by:

$$\mathcal{W}_R = CTAB_R(\mathcal{R}_B, S_B) \quad (10)$$

$$\mathcal{W}_S = CTAB_S(S_B, \mathcal{R}_B) \quad (11)$$

$$\mathcal{W}_Q = CTAB_F(\mathcal{W}_R \odot \mathcal{W}_S, \mathcal{W}_R \odot \mathcal{W}_S) \quad (12)$$

where  $\odot$  represents the element-wise multiplication.  $CTAB(\tau, \Lambda)$  denotes the proposed Cross-Task Attention Block utilized for modeling the correlation between the target input features  $\tau$ , and auxiliary input features  $\Lambda$ . The anatomical correlation between  $\mathcal{R}_B$  and  $S_B$  is modeled by utilizing  $CTAB_R$ , and vice versa for  $CTAB_S$ . Meanwhile,  $CTAB_F$  deploys self-attention mechanisms to determine which parts of the input high-level features are essential based on contextual information.

With such design, we are able to facilitate high-level information flows between registration and segmentation tasks. On the one hand, the CTAB integrates anatomical priors from the segmentation task in the form of global constraints while excluding redundant information. As a result, we learn the correlation between the two tasks explicitly. On the other hand, we aim to acquire a more robust and generalizable anatomical correlation by proposing to share the attention from the segmentation network to the registration network.

#### 3.2.4. Registration-specific final decoder

We can then provide the final predictions for the whole network using the fused feature  $\mathcal{W}_Q$  from the task-shared anatomical correlation module. Similar to the initial predictions, we employ a registration-specific decoder to perform final predictions using  $\mathcal{W}_Q$  as input. As a result, the overall loss function for the entire network can be formulated as:

$$\begin{aligned} \mathcal{L}_{Total} = & \underbrace{\omega \cdot (\alpha \tilde{\mathcal{L}}_D + \beta \tilde{\mathcal{L}}_R) + \gamma \tilde{\mathcal{L}}_{DSC} + \eta \tilde{\mathcal{L}}_{seg}}_{\text{Initial Registration}} \\ & + \underbrace{\omega \cdot (\alpha \mathcal{L}_D + \beta \mathcal{L}_R) + \gamma \mathcal{L}_{DSC}}_{\text{Final Registration}} \end{aligned} \quad (13)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\eta$  are the hyperparameters that penalizes differences in appearance, local spatial variations in the deformation field, volume

overlap for anatomical structure, and volume overlap for predicted segmented structure, respectively.  $\omega$  are the hyperparameter for the voxel-wise anatomical constraint weight (described in Section 3.2.5). The final registration loss functions ( $\mathcal{L}$  without tilde) utilize identical formulations as their initial counterparts.

The architecture is summarized as follows: To explicitly learn the correlation between registration and segmentation, the proposed AC-DMiR includes initial task-specific networks and a task-shared anatomical correlation module. The final warped moving image  $I_m \circ \phi$  with deformation field  $\phi$  can then be generated by the final registration decoder.

#### 3.2.5. Deformation field refinement with anatomical constraint

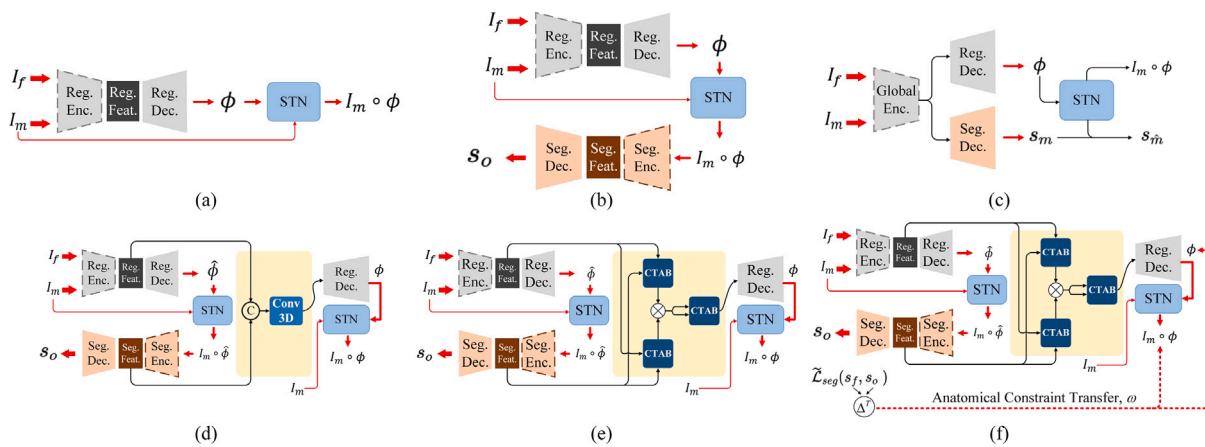
By computing the discrepancy between the predicted segmentation maps of the initial warped images and ground-truth annotations of  $s_f$ , we can approximate the anatomical constraint utilizing initial segmentation predictions. We assume the estimated voxel-wise anatomical constraint can be transferred from segmentation to registration since both of these tasks are complementary and correlated. We exploit this anatomical relation to obtain more realistic deformation results.

In particular, we leverage both high-level feature maps extracted from the initial segmentation network  $\varphi_S$  and the initial registration network  $\varphi_R$  to generate the final predictions of the warped image  $I_m \circ \phi$  given a pair of moving and fixed image input. First, the segmentation maps predicted by the initial segmentation decoder and the ground-truth fixed annotations are compared for voxel-wise prediction discrepancy. The discrepancy is then utilized as a voxel-wise anatomical constraint indicator. Since the two tasks are interrelated, we hypothesize that the anatomical priors can be transferred from segmentation to registration task. Voxels with a high segmentation map prediction discrepancy indicate a weaker anatomical correlation for this particular spot and, as a result, should be allocated a higher  $\omega$  weight in the target registration task. For the deformable prediction on the target registration task, the following equation is formulated to assign  $\omega$ :

$$\omega = \frac{1}{K} \sum_{k=1}^K \sum_{p \in \Omega} |MSE(s_f^k(p), s_o^k(p))| + 1 \quad (14)$$

where  $s_o = \varphi_S(I_m \circ \hat{\phi})$  and  $s_f$  represents the predicted segmentation maps by  $\varphi_S$  and the ground-truth annotations of  $I_f$ , respectively. As shown in Eq. (13), the voxel-wise weight  $\omega$  is applied on penalizing the (dis)similarity criterion  $\mathcal{L}_D$  and deformation smoothness criterion  $\mathcal{L}_R$  in both initial and final phase.

The voxel-wise weight  $\omega$  is designed to have a minimum value of 1. If the initial segmentation decoder gives identical predictions for a



**Fig. 4.** Network architectures used in the ablation experiment. (a) is the baseline Transmorph network architecture introduced in Chen et al. (2022); (b) SimpleSeg is the base Transmorph network architecture with auxiliary Transmorph-based segmentation network; (c) ReSNet is a coupled architecture with a global Swin transformer based-encoder and task-specific CNN decoder; (d) AC-DMiR (CM) is the architecture which enforces the correlation between both tasks using simple concatenation operations; (e) AC-DMiR (F) is the architecture which learns the anatomical correlation between the two tasks using our proposed task-shared anatomical correlation module; (f) is the fully proposed AC-DMiR network architecture. Details about the training loss used in each ablation network architecture are presented in Section 5.1.

voxel in the  $s_o$ , this indicates that the anatomical correlation in this region is very high, and the predicted warped moving image  $I_m \circ \phi$  is likely to be realistic and close to the actual  $I_f$ . Thus, the MSE between  $s_f$  and  $s_o$  becomes closer to 0, and the  $\omega$  will be closer to 1. If the difference of the predicted segmentation maps is large, then the anatomical correlation is weak, thus, it is hard to produce anatomically plausible results. In this case, the weight  $\omega$  becomes greater than 1, and the anatomical priors for this tissue region has a higher contribution to the registration in both initial and final training loss.

### 3.2.6. Summary

Combining the proposed anatomical correlation-aware architecture with the task-shared anatomical correlation transfer and the anatomical plausibility refinement with voxel-wise weight  $\omega$  leads to our Anatomical Correlation-Aware Deformable Medical Image Registration (AC-DMiR) framework. As shown in Fig. 2, we use a correlation-aware architecture, which incorporates a shared anatomical correlation module and initial task-specific networks. During the entire training process, the final predictions of the warped image  $I_m \circ \phi$  are refined using the voxel-wise anatomical correlation indicator introduced in our global segmentation-guided anatomical constraint in Section 3.2.5.

The proposed method can be readily integrated into any unsupervised medical image registration approach. To show that our proposed method is complementary to existing methodologies, we use TransMorph (Chen et al., 2022) as our base framework as it provides a simple yet strong baseline. TransMorph offers a hybrid Transformer-CNN model with an unlimited receptive field that allows for a more accurate understanding of the long-range spatial relationships between voxels of moving and fixed images.

## 4. Experiments

### 4.1. Datasets and preprocessing

We demonstrate the proposed AC-DMiR on the task of brain MRI and uterus MRI registration. The proposed method was thoroughly validated using two datasets comprising over 400 image pairs. The following sections describe the details of each dataset.

#### 4.1.1. Brain MRI registration

We utilized the OASIS (Marcus et al., 2007) dataset to construct a large, multistudy dataset comprising 414 T1-weighted brain MRI scans. Each dataset has its own set of acquisition details and health

conditions. All image volumes were resampled to 1 mm isotropic voxels with a  $256 \times 256 \times 256$  grid. We utilize FreeSurfer (Fischl, 2012) to perform conventional preprocessing procedures for each scan, such as skull stripping, resampling, and affine transformation, then crop the pre-processed images to  $160 \times 192 \times 224$ . For training, validation, and test sets, the dataset was randomly split into 289, 83, and 42 (7:2:1) volumes. Each image volume was randomly paired with two other volumes in each set to produce four  $I_f$  and  $I_m$  registration pairs, yielding 1156, 332, and 168 image pairs. We employed FreeSurfer to anatomically segment all MRIs, and we performed a visual inspection to identify any significant issues in segmentation results and affine alignment. For evaluating the registration accuracy of AC-DMiR, FreeSurfer was utilized to create label maps for 35 major anatomical regions.

#### 4.1.2. Uterus MRI registration

MRI has been widely used in High-Intensity Focused Ultrasound (HIFU) surgery to guide uterine fibroid ablation. During HIFU treatment, the patient rests still so that the physician can perform treatment precisely on the uterine fibroid without excessive damage to healthy tissues. However, during HIFU ablation therapy, the filling bladder, filling rectum, and emptying bladder (BRB) procedure for removing bowel loops in the sonication beam path contributes to unintended uterus deformations and deteriorate treatment outcomes. As a result, pre- to intra-operative MR registration has become a key method for accurate HIFU navigation.

This research used an MRI dataset from the National Engineering Research Center of Ultrasound Medicine (JC200/300 Haifu treatment system, Haifu, China) that includes 181 pre- and intra-operative uterus T1-weighted MR scans. In each patient's MR image, specific organs and structures were manually segmented. The structures segmented included the following: uterus, urinary bladder and vertebral column. Several medical students performed the manual segmentation, which was then corrected by an expert radiologist at the National Engineering Research Center of Ultrasound Medicine. The voxel sizes in the MRI volumes range from  $5.00 \times 0.65 \times 0.65$  mm to  $6.00 \times 0.90 \times 0.90$  mm. All scans were resampled to an identical voxel spacing of  $6.00 \times 0.78 \times 0.78$  mm using trilinear interpolation. The volumes were all cropped and zero-padded to have a size of  $32 \times 256 \times 256$  voxels. The pre- and intra-operative uterus MR images were rigidly registered using FreeSurfer. The dataset was split into 127, 36, and 18 (7:2:1) volumes for training, validation, and testing. We conducted three-fold cross-validation on the eighteen image volumes, resulting in 54 testing volumes in total.

#### 4.2. Baseline methods

We compared AC-DMiR to a variety of previous state-of-the-art registration methodologies. We begin by evaluating AC-DMiR with two non-deep-learning-based approaches, which are detailed below along with their hyperparameter settings:

- SyN (Avants et al., 2008): For inter-patient brain MR registration tasks, we employed the mean squared difference (MSQ) as the objective function, a Gaussian smoothing of 3 and three scales with 180, 100, and 40 iterations, respectively. For pre- to intra-operative uterus MR registration, we utilized the cross-correlation (CC) as the objective function, including a default Gaussian smoothing of 5 and three scales with 160, 80, and 40 iterations, respectively.
- NiftyReg (Modat et al., 2010): For all registration tasks, we utilized the sum of squared differences (SSD) as the objective function and bending energy as a regularizer. We employed three scales with 200 iterations each for inter-patient brain MR registration with a regularization weighting of 0.0004. For pre- to intra-operative uterus MR registration, we employed a regularization weighting of 0.0006 and three scales with 400 iterations each.

The proposed method was then compared to several current deep-learning-based approaches. For inter-patient brain MR registration, the loss function that comprises MSE loss from Eq. (1), diffusion regularization in Eq. (3), and Dice loss in Eq. (7) were employed unless otherwise stated for a fair comparison. We implemented the loss function that comprises LNCC loss in Eq. (2), bending energy in Eq. (4), and Dice loss in Eq. (7) for pre- to intra-operative uterus MR registration. The weights for smooth deformation field regularization and Dice loss, respectively, are denoted by hyperparameters  $\eta_{vm}$  and  $\mu_{vm}$ . For each approach, the following were the detailed parameter settings:

- VoxelMorph (Balakrishnan et al., 2019): The regularization hyperparameters  $\eta_{vm}$  and  $\mu_{vm}$  were set to 0.02 and 1, respectively, for inter-patient brain MR registration, where these values were described as the optimum values in Balakrishnan et al. (2019). For pre- to intra-operative uterus MR registration, we set  $\eta_{vm} = \mu_{vm} = 1$ .
- VoxelMorph-diff (Dalca et al., 2019): For inter-patient brain MR registration tasks, the diffeomorphic registration loss function proposed in Dalca et al. (2019) was used with  $\sigma$  set to 0.01 and  $\lambda$  set to 20. For pre- to intra-operative uterus MR registration, we used the surface-based registration loss function (Dalca et al., 2019) with  $\sigma_I = \sigma_S = 0.05$  and  $\lambda = 15$ . For both datasets, VoxelMorph-diff was trained with  $T = 7$  scaling and squaring steps.

We compared the registration quality of the proposed AC-DMiR to that of current Transformer-based networks that have acquired state-of-the-art registration accuracy in other applications (e.g., semantic segmentation, image classification, etc.). These models were modified to create three-dimensional deformation fields that warp the moving image. These methodologies and VoxelMorph (Balakrishnan et al., 2019) differ primarily in their network architecture; the loss function, and training details of the network remain unchanged. The Transformer-CNN architecture was employed for these models (i.e., ViT-V-Net (Chen et al., 2021a) and TransMorph (Chen et al., 2022)). The settings for the respective network hyperparameters were as follows:

- ViT-V-Net (Chen et al., 2021a): ViT was employed to establish this registration network. The default network hyperparameter settings provided in Chen et al. (2021a) were used in both datasets.

- TransMorph (Chen et al., 2022): The window sizes (i.e.,  $M_x \times M_y \times M_z$ ) utilized in Swin Transformer (Liu et al., 2021) were adjusted to 5, 6, 7 for MR brain registration and 4, 4, 4 for pre- to intra-operative uterus MR registration, respectively, due to the sizes of the image volumes. The same loss function parameters as VoxelMorph (Balakrishnan et al., 2019) were utilized for all tasks.

Besides, we compared the registration accuracy of the proposed AC-DMiR to that of the state-of-the-art deep learning framework that jointly learns networks for image registration and image segmentation (i.e., DeepAtlas (Xu and Niethammer, 2019) and Cross-stitch (Elmahdy et al., 2021)). The following settings were set for the abovementioned network hyperparameters:

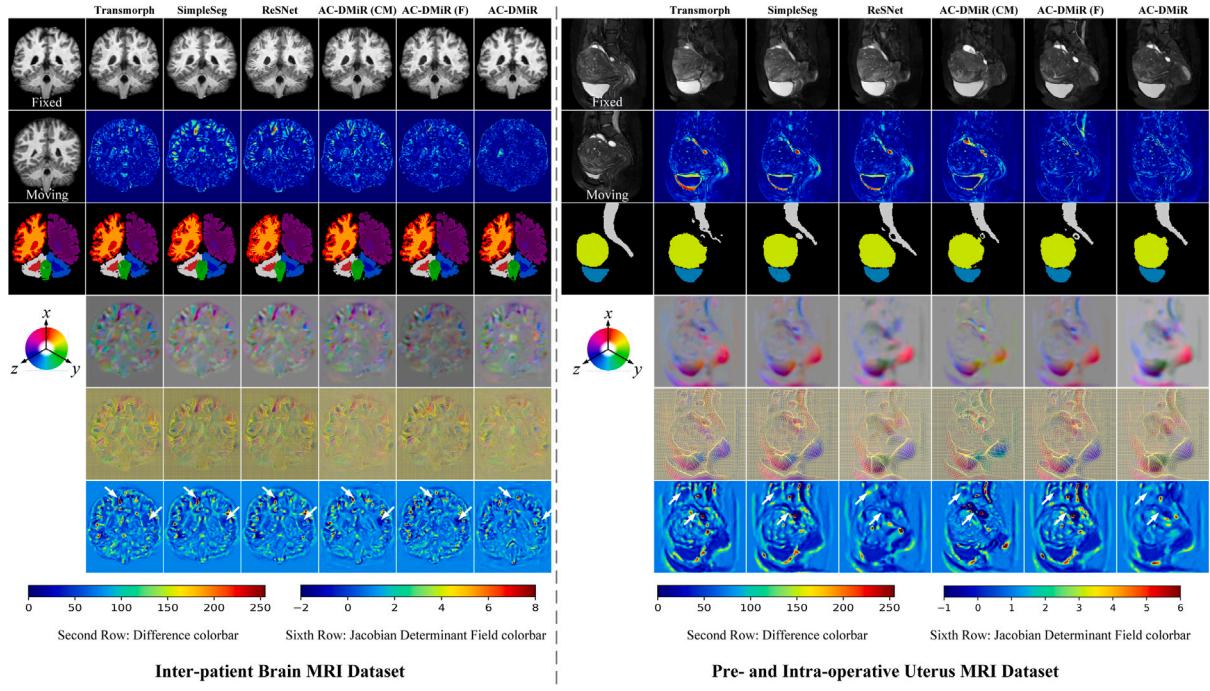
- DeepAtlas (Xu and Niethammer, 2019): For inter-patient brain MR registration, the loss weights (i.e.,  $\lambda_r$ ,  $\lambda_a$ ,  $\lambda_{sp}$ ) are set as 5000, 3, 3, respectively. We use the same loss weights as for the pre- to intra-operative uterus MRI experiment except for  $\lambda_r = 4500$ , since the registration for the uterus before and after the BRB procedure necessitates substantial deformations and hence less regularization.
- Cross-stitch (Elmahdy et al., 2021): The regularization hyperparameters (i.e.,  $\omega_0$ ,  $\omega_1$ ,  $\omega_2$ ,  $\omega_3$ ) utilized in the Cross-stitch network (Elmahdy et al., 2021) were adjusted to (5, 1, 1, 0.5) and (10, 0.5, 1.5, 1.0) for MR brain and pre- to intra-operative uterus MR registration, respectively.

#### 4.3. Implementation and training details

The proposed AC-DMiR was implemented using PyTorch with the support of two NVIDIA Quadro RTX 8000 GPUs. We utilize TransMorph (Chen et al., 2022) as the task-specific backbone network for both tasks in our anatomical correlation-aware architecture. The registration and segmentation feature bottlenecks are residual blocks with two  $3 \times 3 \times 3$  and three  $1 \times 1 \times 1$  3D convolution operations. A  $3 \times 3 \times 3$  3D convolution is utilized to map each 16-channeled feature map to the desired  $k \in [1, K]$  number of classes at the final layer for the segmentation task. Adam optimizer (Kingma and Ba, 2014) is utilized with a learning rate of  $1 \times 10^{-4}$  and follows a polynomial decay with an exponent of 0.9. AC-DMiR is trained in an end-to-end manner with a batch size of 1 for 500 epochs. During the training phase, the brain MR dataset was augmented by flipping in random directions, whereas the uterus MR dataset had no data augmentation. The hyperparameters for Swin Transformer in AC-DMiR are set according to the respective network hyperparameters of TransMorph (Chen et al., 2022). We set the weights for image similarity loss to  $\alpha = 2$  and  $\alpha = 4$  for the MR brain and uterus registration, respectively. Meanwhile, for both datasets, the weight for smooth deformation field regularization  $\beta$  and Dice loss  $\gamma$  were set according to the regularization hyperparameters  $\eta_{vm}$  and  $\mu_{vm}$  of the VoxelMorph baseline method, which are enhanced by our deformation field refinement with anatomical constraint regularization  $\omega$ . The hyperparameter  $\eta$  for regularize the segmentation network's accuracy was set to 1 for both datasets (details for choosing the value for  $\eta$  are described in Section 5.5). The entire training process of AC-DMiR is illustrated in Algorithm 1.

#### 4.4. Evaluation metrics

The registration accuracy of each model was evaluated for both MR brain and uterus registration tasks based on the volume overlap of anatomical segmentations, which was evaluated using the Dice score (Dice, 1945). For each patient, we computed the average of the Dice scores for all anatomical structures. Various registration methodologies were evaluated in terms of the mean and standard deviation of all patients' averaged scores.



**Fig. 5.** Qualitative analysis of the inter-patient brain MR registration task (left) and pre- to intra-operative uterus MR registration task (right) using different variants of network. The deformed moving images  $I_m \circ \phi$  are depicted in the first row, followed by the voxel-wise difference between  $I_m \circ \phi$  and the fixed image  $I_f$  in the second row, and warped moving label  $s_m$  in the third row. The fourth row depicts the deformation field  $\phi$  where spatial dimension  $x$ ,  $y$ , and  $z$  is mapped to each of the RGB color channels, respectively. The fifth and sixth row exhibits the deformation grid and the Jacobian determinant field.

Medical image registration also should produce realistic deformations and, as a result, should not generate deformations with foldings. Hence, we quantitatively examine the deformation fields' regularity. Particularly, the local properties of  $\phi$  around voxel  $p$  are captured by the Jacobian matrix  $J_\phi(p) = \nabla\phi(p) \in \mathbb{R}^{3 \times 3}$ . We count all non-background voxels that are less than zero (i.e.,  $J_\phi(p) \leq 0$ ) and do not exhibit diffeomorphic deformation (Ashburner, 2007). Furthermore, we evaluated the structural differences between the deformed moving and the target fixed images for both registration tasks utilizing the Structural Similarity Index (SSIM) score (Wang et al., 2004).

## 5. Results

### 5.1. Ablation study: Individual modules

The key novelty of our proposed AC-DMiR framework is to utilize supervised anatomical segmentation to effectively learn the shared correlation of the tissue/organ anatomy between tasks. We perform an ablation experiment on each of these components to validate our motivation. The architecture of each ablation network are illustrated in Fig. 4. TransMorph is used as a strong baseline since it can already capture deformable features relatively well without the incorporation of anatomical priors. Then, instead of implementing the task feature correlation module, we use supervised segmentation (denoted as Seg. Aux.) and add the segmentation auxiliary task to validate whether a simplistic method can improve the TransMorph baseline. We then examine whether a coupled architecture (denoted as Coup.) would provide prior knowledge for anatomies or tissue patterns by using a joint framework of image segmentation and registration that is made up of a global encoder and task-specific decoder. To evaluate whether a simple and direct concatenation method could be effective for feature fusing between two tasks (denoted as Feat. Fus.), we concatenate the high-level feature representations from both networks and forward it to a 3D convolutional operation instead of using the proposed Cross-Task Attention Block. Then, we evaluate our proposed task-shared anatomical correlation module (denoted as TACM) to check if explicitly

---

### Algorithm 1: Training pipeline for AC-DMiR

---

```

Input: Paired training dataset  $\{I_m, I_f\}$ , Number of training epochs  $N_{epoch}$ , Model hyperparameters  $\alpha, \beta, \gamma, \eta$ 
Output: Weight parameters of initial registration network, initial segmentation network and final registration decoder  $\theta_R, \theta_S, \theta_F$ 
Result: Deformation Field  $\phi$ , warped moving image  $I_m \circ \phi$ 
1 for  $n = 1, \dots, N_{epoch}$  do
2   if  $n = 1$  then
3     |  $\omega = 1$ ;
4   end
5    $\tilde{\mathcal{L}}_D \leftarrow \tilde{\mathcal{L}}_{MSE}(I_f, I_m \circ \hat{\phi}) + \tilde{\mathcal{L}}_{LNCC}(I_f, I_m \circ \hat{\phi})$ ;
6    $\tilde{\mathcal{L}}_R \leftarrow \tilde{\mathcal{L}}_{smooth}(\hat{\phi}) + \tilde{\mathcal{L}}_{bending}(\hat{\phi}) + \tilde{\mathcal{L}}_{anti-fold}(\hat{\phi})$ ;
7    $\tilde{\mathcal{L}}_{DSC} \leftarrow \tilde{\mathcal{L}}_{DSC}(s_f, s_{\hat{m}})$ ;
8    $\theta_R \leftarrow -\nabla_{\theta_R} \omega \cdot (\alpha \tilde{\mathcal{L}}_D + \beta \tilde{\mathcal{L}}_R) + \gamma \tilde{\mathcal{L}}_{DSC}$ ;
9    $\tilde{\mathcal{L}}_{seg} \leftarrow \tilde{\mathcal{L}}_{seg}(s_f, s_o)$ ;
10   $\theta_S \leftarrow -\nabla_{\theta_S} \eta \tilde{\mathcal{L}}_{seg}$ 
11   $\mathcal{L}_D \leftarrow \mathcal{L}_{MSE}(I_f, I_m \circ \phi) + \mathcal{L}_{LNCC}(I_f, I_m \circ \phi)$ ;
12   $\mathcal{L}_R \leftarrow \mathcal{L}_{smooth}(\phi) + \mathcal{L}_{bending}(\phi) + \mathcal{L}_{anti-fold}(\phi)$ ;
13   $\mathcal{L}_{DSC} \leftarrow \mathcal{L}_{DSC}(s_f, s_{\hat{m}})$ ;
14   $\theta_F \leftarrow -\nabla_{\theta_F} \omega \cdot (\alpha \mathcal{L}_D + \beta \mathcal{L}_R) + \gamma \mathcal{L}_{DSC}$ ;
15   $\omega \leftarrow \frac{1}{K} \sum_{k=1}^K \sum_{p \in \Omega} |MSE(s_f^k(p), s_o^k(p))| + 1$ ;
16 end

```

---

learning the anatomical correlation between the two tasks makes a significant difference. Finally, we implement our deformation field refinement, which is based on the anatomical constraint (denoted as AC) approximated by segmentation map prediction discrepancy. This final configuration corresponds to the proposed AC-DMiR framework.

To train the baseline model (Chen et al., 2022), conventional registration loss functions  $\mathcal{L}_D$ ,  $\mathcal{L}_R$  and  $\mathcal{L}_{DSC}$  were used for evaluating

**Table 1**

Ablation study of individual components in our proposed AC-DMiR framework on inter-patient MR brain and pre- to intra-operative uterus MR registration tasks. For each method, the Dice score for registration and segmentation, the Structural Similarity Index Metric (SSIM) score and the percentage of voxels with a non-positive Jacobian determinant are evaluated. The best values are denoted by bolded numbers.

Task	Method	Seg. Aux.	Coup.	Feat. Fus.	TACM	AC	Reg. Dice ↑	Seg. Dice ↑	SSIM ↑	$ J_\varphi  \leq 0 (\%) \downarrow$
Brain MRI registration	Baseline (Chen et al., 2022)						$0.735 \pm 0.115$	–	$0.890 \pm 0.106$	$0.527 \pm 0.324$
	SimpleSeg	✓					$0.738 \pm 0.128$	$0.619 \pm 0.106$	$0.897 \pm 0.121$	$0.424 \pm 0.354$
	ReSNet	✓	✓				$0.739 \pm 0.150$	$0.618 \pm 0.191$	$0.909 \pm 0.110$	$0.531 \pm 0.341$
	AC-DMiR (CM)	✓		✓			$0.741 \pm 0.120$	$0.620 \pm 0.164$	$0.916 \pm 0.116$	$0.955 \pm 0.244$
	AC-DMiR (F)	✓			✓		$0.746 \pm 0.185$	$0.620 \pm 0.162$	$0.918 \pm 0.218$	$0.845 \pm 0.135$
Uterus MRI registration	AC-DMiR	✓			✓	✓	$0.755 \pm 0.112$	$0.621 \pm 0.110$	$0.925 \pm 0.203$	$0.542 \pm 0.397$
	Baseline (Chen et al., 2022)						$0.718 \pm 0.148$	–	$0.909 \pm 0.217$	$1.875 \pm 0.384$
	SimpleSeg	✓					$0.719 \pm 0.366$	$0.788 \pm 0.296$	$0.912 \pm 0.223$	$1.145 \pm 0.647$
	ReSNet	✓	✓				$0.719 \pm 0.123$	$0.790 \pm 0.185$	$0.915 \pm 0.110$	$1.311 \pm 0.312$
	AC-DMiR (CM)	✓		✓			$0.725 \pm 0.208$	$0.787 \pm 0.177$	$0.918 \pm 0.118$	$1.213 \pm 0.642$
AC-DMiR	AC-DMiR (F)	✓			✓		$0.729 \pm 0.317$	$0.790 \pm 0.224$	$0.921 \pm 0.121$	$1.101 \pm 0.653$
	AC-DMiR	✓			✓	✓	$0.731 \pm 0.129$	$0.793 \pm 0.112$	$0.924 \pm 0.108$	$1.211 \pm 0.390$

**Table 2**

Quantitative evaluation results on inter-patient brain MR registration task. The Dice score for registration and segmentation, the SSIM score and the percentage of voxels with a non-positive Jacobian determinant are evaluated for various registration methods. Bolded numbers indicate the best values.

Method	Registration dice score ↑		Segmentation dice score ↑		SSIM ↑		$ J_\varphi  \leq 0 (\%) \downarrow$
	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	
Affine	$0.393 \pm 0.122$	0.382	–	–	$0.760 \pm 0.311$	0.749	–
SyN (Avants et al., 2008)	$0.645 \pm 0.215$	0.635	–	–	$0.849 \pm 0.212$	0.828	<0.0001
NiftyReg (Modat et al., 2010)	$0.646 \pm 0.115$	0.638	–	–	$0.850 \pm 0.112$	0.831	$0.045 \pm 0.067$
VoxelMorph (Balakrishnan et al., 2019)	$0.731 \pm 0.112$	0.721	–	–	$0.865 \pm 0.109$	0.834	$0.541 \pm 0.333$
VoxelMorph-diff (Dalca et al., 2019)	$0.724 \pm 0.132$	0.719	–	–	$0.854 \pm 0.112$	0.829	<0.0001
VIT-V-Net (Chen et al., 2021a)	$0.735 \pm 0.109$	0.724	–	–	$0.874 \pm 0.178$	0.870	$0.741 \pm 0.164$
TransMorph (Chen et al., 2022)	$0.735 \pm 0.115$	0.734	–	–	$0.890 \pm 0.106$	0.887	$0.527 \pm 0.324$
DeepAtlas (Xu and Niethammer, 2019)	$0.739 \pm 0.211$	0.737	$0.631 \pm 0.211$	0.621	$0.904 \pm 0.229$	0.891	$0.353 \pm 0.262$
Cross-stitch (Elmaghd et al., 2021)	$0.749 \pm 0.111$	0.750	$0.635 \pm 0.135$	$0.625$	$0.910 \pm 0.206$	0.904	$0.562 \pm 0.431$
AC-DMiR	$0.755 \pm 0.112$	0.754	$0.621 \pm 0.110$	0.619	$0.925 \pm 0.203$	$0.918$	$0.542 \pm 0.397$

loss with the same hyperparameter  $\alpha$ ,  $\beta$  and  $\gamma$  settings as mentioned in Section 4.3. For SimpleSeg and ReSNet variant networks, loss function which comprised of  $\mathcal{L}_D$ ,  $\mathcal{L}_R$ ,  $\mathcal{L}_{DSC}$  and  $\mathcal{L}_{seg}$  were used to train those models ( $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\eta$  are set according to the same settings mentioned in Section 4.3). Meanwhile, to train AC-DMiR (CM) and AC-DMiR (F), the total loss function in Eq. (13) is utilized with  $\omega = \eta = 1$ . Finally, complete loss function mentioned in Eq. (13) is utilized to calculate loss during the training of our proposed AC-DMiR. The parameters  $\alpha$ ,  $\beta$  and  $\gamma$  for AC-DMiR (CM), AC-DMiR (F), and AC-DMiR are the same as those described in Section 4.3.

Fig. 5 shows the qualitative results on a sample slice for brain and uterus MRI registration. Among the variants of network, AC-DMiR variants (i.e., AC-DMiR (CM) and AC-DMiR (F)) produced better qualitative results, with our proposed AC-DMiR producing the least error deformed moving images when compared with the fixed image. On the other hand, AC-DMiR showed better qualitative results (highlighted by the white arrows in the Jacobian field) with lower folded voxels (regions with dark blue color in the Jacobian field).

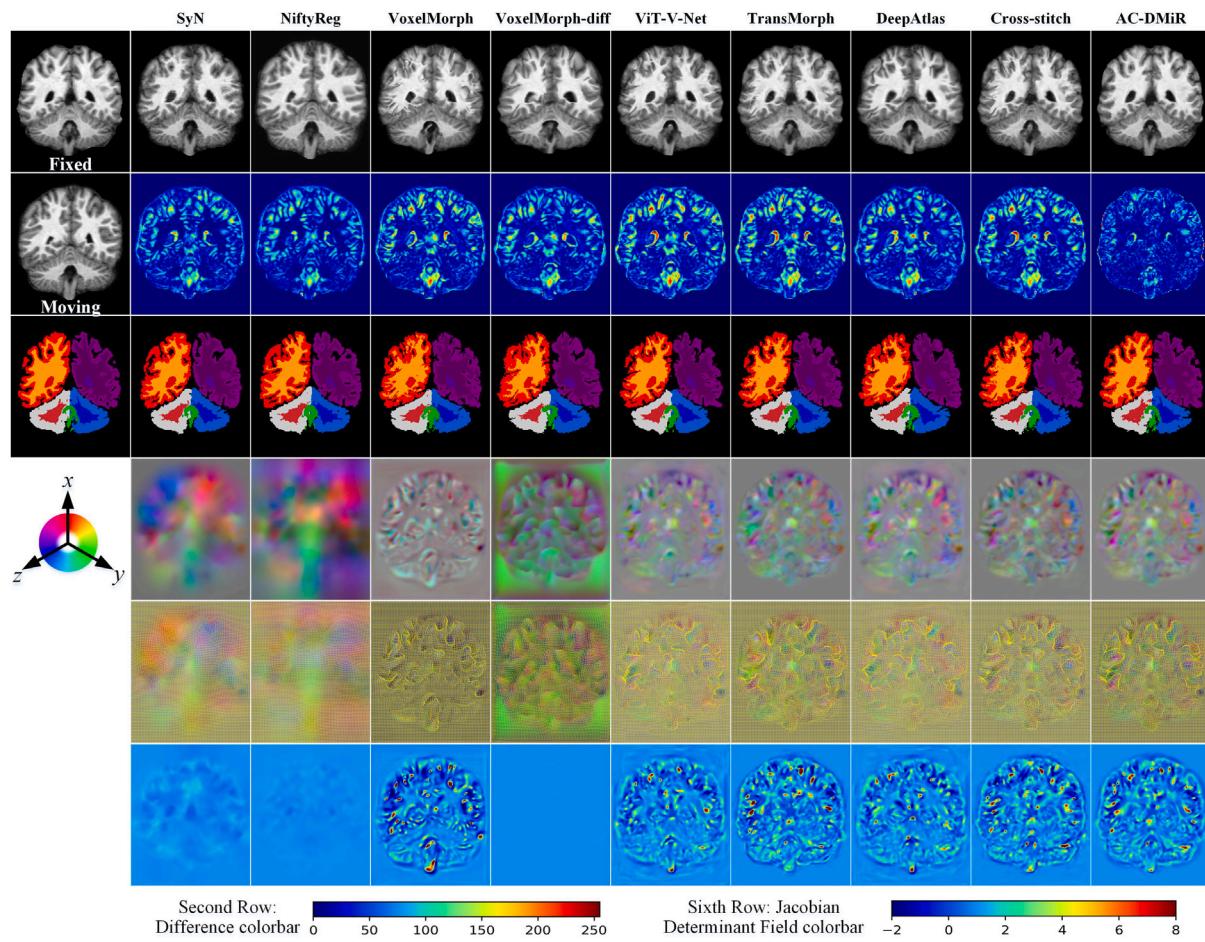
As shown in Table 1, utilizing segmentation features as an auxiliary task (denoted as SimpleSeg) without making major changes to the training method or network architecture can yield 0.738 and 0.719 Dice scores for MRI brain and uterus registration, respectively, compared to the TransMorph baseline. This supports the common perception that additional organ boundary information acquired during segmentation can contribute to the learning of deformable registration. The improvement, however, is not large, most likely since this simplistic method of learning two tasks simultaneously cannot guarantee good generalization registration accuracy in terms of the volume overlap between anatomical label map, which was evaluated using the Dice score (Kokkinos, 2017; Xu et al., 2018). By leveraging a multi-task network architecture, ReSNet yields slightly better registration accuracy compared to SimpleSeg, with Dice score of 0.739 for brain registration task. For uterus registration task, ReSNet yield the same registration accuracy compared to SimpleSeg, with 0.719 Dice score but less

distributed. On the other hand, we can utilize the high-level segmentation features more effectively and increase the realism of the warped images after final registration by explicitly modeling the correlation between registration and segmentation using the correlation-aware framework with anatomical feature correlation module AC-DMiR (F). This results in a 0.67% and 0.55% absolute improvement compared to AC-DMiR (CM), with Dice scores of 0.746 and 0.729 for MRI brain and uterus registration, respectively. When the correlation learning modules are removed but the high level registration and segmentation features are fused using simple concatenation method AC-DMiR (CM), the registration quality scores reduces to 0.741 and 0.725 Dice scores, respectively. This emphasizes the necessity of comprehending the correlation between the two tasks. In addition, we integrate our deformation field refinement with the anatomical constraint module into the network by comparing the prediction discrepancy between ground-truth annotations and predicted segmentation maps of initial warped images, which leads to our final proposed framework AC-DMiR. When compared to AC-DMiR (F), this leads to an incremental 1.2% and 0.3% improvement in absolute registration accuracy in the MRI brain and uterus registration tasks, respectively. As shown in Table 1, improved deformable registration accuracy can be achieved with our aforementioned parameter settings by using both the correlation-aware framework with task feature correlation and anatomical constraint. The findings indicate that each of the proposed components makes a substantial contribution.

## 5.2. Comparison to the state-of-the-art approaches

### 5.2.1. Brain MRI registration

Fig. 6 depicts the qualitative results on a sample slice for inter-patient brain MRI registration. Although AC-DMiR lacks an explicit constraint for diffeomorphic deformations, in this setting, the segmentation network's anatomical constraint and smoothness loss produce generally well-behaved and smooth results. Implementations like SyN,



**Fig. 6.** Qualitative analysis of the inter-patient brain MR registration task using different registration techniques. The deformed moving images  $I_m \circ \phi$  are depicted in the first row, followed by the voxel-wise difference between  $I_m \circ \phi$  and the fixed image  $I_f$  in the second row, and warped moving label  $s_m$  in the third row. The fourth row depicts the deformation field  $\phi$  where spatial dimension  $x$ ,  $y$ , and  $z$  is mapped to each of the RGB color channels, respectively. The fifth and sixth row exhibits the deformation grid and the Jacobian determinant field.

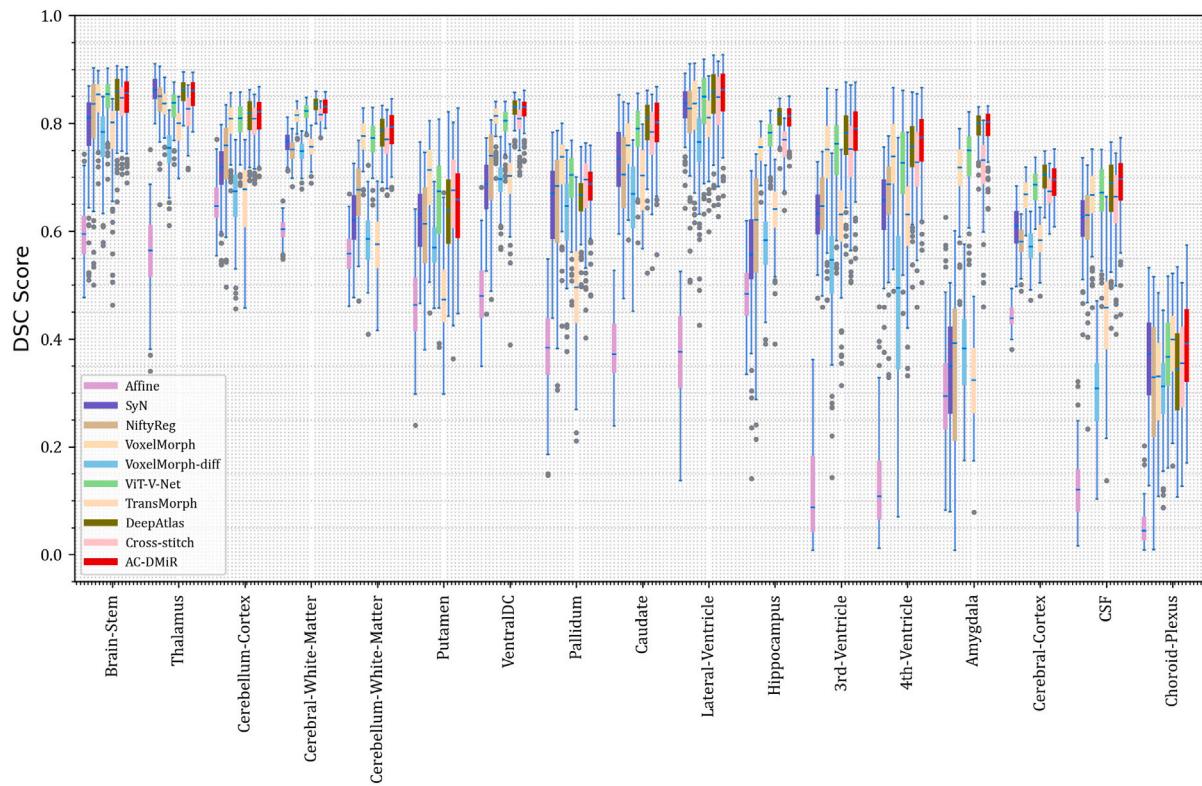
NiftyReg, and VoxelMorph-diff can strongly encourage or impose diffeomorphic deformations, but during our experiments, we found out that these had a negative impact on the results. When compared to other registration methods that do not incorporate implementations for diffeomorphic deformations, the proposed AC-DMiR in this work produced the best Dice scores, which also turned out to produce good deformation regularity.

The quantitative evaluations are shown in Table 2. Under our specific experimental settings mentioned in this work, our proposed AC-DMiR outperforms other methods, achieving the highest Dice score of 0.755, proving its effectiveness in handling deformable organ motion modeling problems. Although the Dice score of the proposed AC-DMiR is not the least distributed, it still outperformed the existing registration methods and generated anatomically plausible (i.e., 0.542% of  $|J_\phi| \leq 0$ ) results. In contrast, AC-DMiR increased the registration Dice score by 0.006 when compared to Cross-stitch and by 0.02 when compared to the baseline method TransMorph. The Wilcoxon signed-rank test revealed a  $p$ -value of  $p \ll 0.005$  for the best performing model (i.e., AC-DMiR) over the second best performing model (i.e., Cross-stitch) in terms of registration DSC scores, indicating that the proposed AC-DMiR outperforms other non-learnable and learnable registration methods.

Additional Dice score results are depicted as a boxplot in Fig. 7 for each anatomical structure. For the convenience of illustration, the Dice scores of the similar anatomical structures in the two hemispheres (i.e., the left and right cerebellum cortex) are averaged into one score. The AC-DMiR model achieves Dice scores comparable to learning-based registration approaches (i.e., DeepAtlas and Cross Stitch) for all

structures, with significantly higher registration quality scores on some structures, such as the cerebellum white matter, and slightly worse on others, such as the pallidum.

When compared to Cross-stitch, the segmentation network of our proposed AC-DMiR produces a slightly lower DSC score for segmentation accuracy. This can be interpreted as the segmentation network was designed to function as an auxiliary network to provide structural information to the primary registration network. While it is true that the segmentation network did not achieve the best segmentation performance (in terms of DSC score) when compared to other JRS framework with segmentation networks, its primary goal was not to achieve the highest segmentation accuracy. Instead, the segmentation network was utilized to provide structural information to the primary registration network through the cross-task attention mechanism. The cross-task attention mechanism fused high-level features from both the segmentation and registration tasks, allowing the primary registration network to make more informed decisions based on the structural information provided by the segmentation network. This ultimately led to the best registration accuracy in our experiments. Overall, the segmentation network played a crucial role in providing structural information to the primary registration network through the cross-task attention mechanism, leading to the best registration accuracy in our experiments.



**Fig. 7.** A quantitative evaluation of the numerous registration methodologies on the inter-patient brain MR registration task. Dice scores for several brain MR substructures utilizing the proposed AC-DMiR and existing image registration approaches are shown in boxplots.

### 5.2.2. Uterus MRI registration

The quantitative evaluation results are presented in [Table 3](#). It includes Dice scores for all abdominal organ structures, and the percentage of non-positive Jacobian determinants between the pre-operative MR scan and the deformed target intraoperative MR scan. The substantial difference between the pre- and intra-operative uterus MR scans can be revealed by examining at the Dice score of 0.292 without registration or affine transformation. Among the traditional registration methods, NiftyReg, which was initially designed for the neuroimaging-based registration approach, achieved the highest Dice score of 0.507. It is noteworthy that the network with diffeomorphic settings (i.e., VoxelMorph-diff) produced lower Dice and SSIM scores as a consequence of not having any folded voxels in the deformation. For the DSC and SSIM score of registration task, the *p*-value from the paired *t*-test between AC-DMiR (Transformer-based) and the best performing CNN-based JRS method (i.e., DeepAtlas) was  $\ll 0.005$ . When compared to other existing registration methodologies with the aforementioned parameter settings, the proposed AC-DMiR model achieved the highest Dice score of 0.731.

On the other hand, Cross-stitch demonstrated the best segmentation accuracy, similarly to our previous findings on brain MR registration task. Despite the performance of (in terms of DSC score) segmentation branch of AC-DMiR was not the state-of-the-art, the segmentation network was still able to contribute to the accuracy of the primary registration network through the use of the cross-task attention mechanism. This mechanism allows the primary registration network to fuse high-level features from both the registration and segmentation tasks, allowing for a more comprehensive understanding of the structural information present in the images. To further demonstrate the effectiveness of the cross-task attention mechanism, we have included results in [Section 5.3](#) from experiments visualizing the attention map of the primary registration network with and without the segmentation network. These results show that the primary registration network was able to achieve significantly better registration accuracy when using the

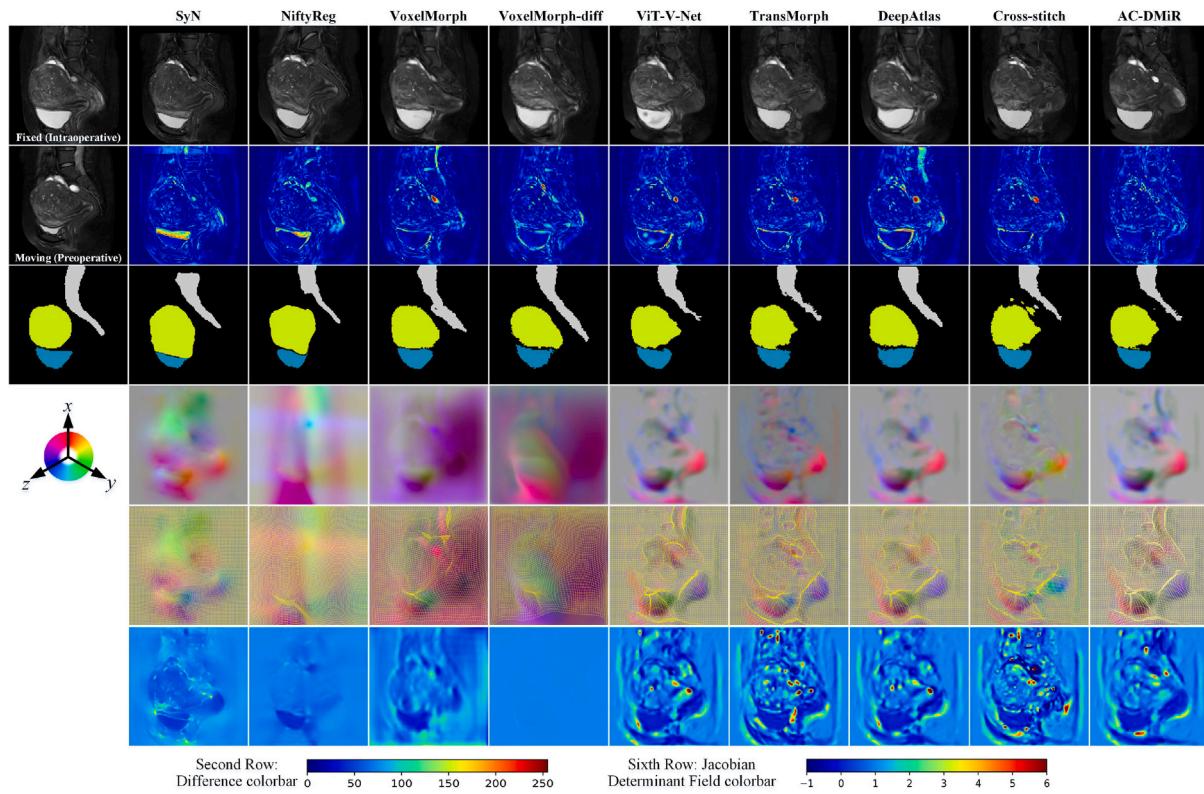
segmentation network as an auxiliary network, due to the additional structural information provided by the cross-task attention mechanism.

[Fig. 8](#) shows the qualitative results for a representative uterus MR slice. Similar to the results in the earlier sections, the proposed AC-DMiR generated smoother deformations (compared to methods without diffeomorphic deformations) and more realistic registration results (depicted by the difference map). Large displacements in pre-to intra-operative uterus MR registration may result in predictions of non-plausible deformable structures, in contrast to the comparatively minor displacements in brain MRI registration, but we are capable of improving the realism of the warped moving images by making use of the voxel-wise anatomical constraint inferred from segmentation. Additional boxplots of Dice scores on the various anatomical structures are shown in [Fig. 9](#).

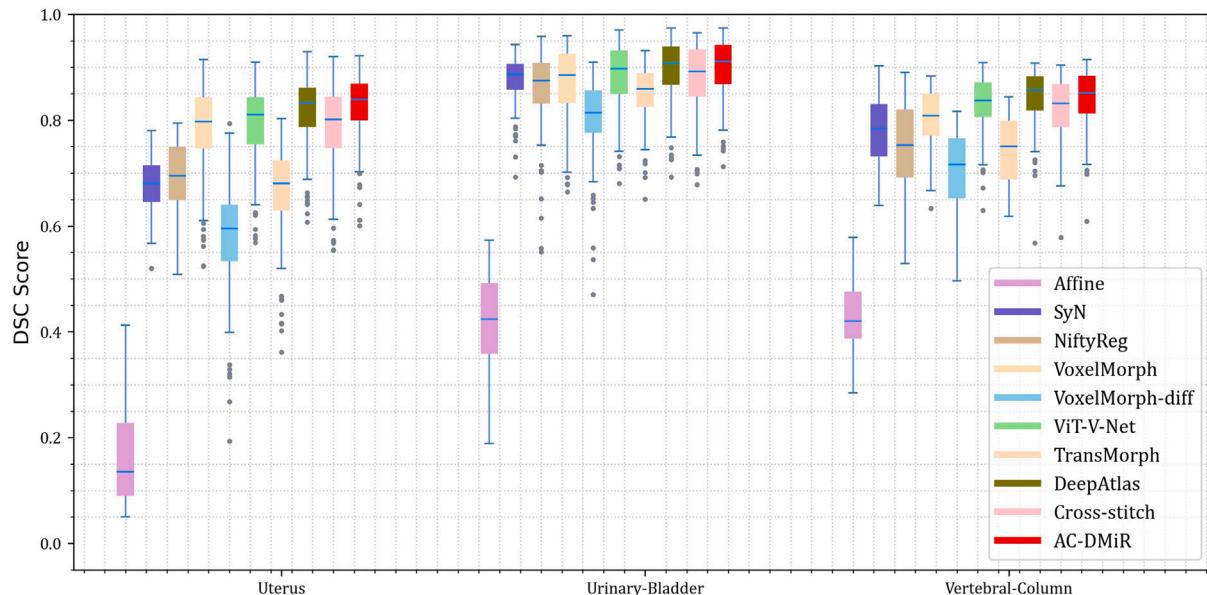
### 5.3. Cross task attention map visualization

In order to provide an in-depth explanation of the TACM, we depicted the focused regions of each ablation network architecture in this section. For each task, each channel obtained  $n$  attention maps ( $n = 384$  for registration of both the uterus and brain MRIs), which were then averaged over all of the input channel maps.

As shown in [Fig. 10](#), the attention maps that are produced by  $CTAB_R$  is able to emphasize the regions that contain large deformation while diminishing the less deformed regions. Attention maps from  $CTAB_S$  also focused on particular anatomical regions associated to various task. On the other hand, the attention maps at the  $CTAB_F$  are indeed increasingly focused to cover registration-specific anatomical regions by combining the input of  $CTAB_S$  and  $CTAB_R$ . The fusion of high-level anatomical features of labeled structures allowed in the preservation of essential deformed anatomical properties that could be further enhanced to distinguish between different deformation severities at different anatomy regions. Besides, it is noteworthy that



**Fig. 8.** Qualitative analysis of the pre- to intra-operative uterus MR registration task using different registration techniques. The deformed moving images  $I_m \circ \phi$  are depicted in the first row, followed by the voxel-wise difference between  $I_m \circ \phi$  and the fixed image  $I_f$  in the second row, and warped moving label  $s_m$  in the third row. The fourth row depicts the deformation field  $\phi$  where spatial dimension  $x$ ,  $y$ , and  $z$  is mapped to each of the RGB color channels, respectively. The fifth and sixth row exhibits the deformation grid and the Jacobian determinant field.

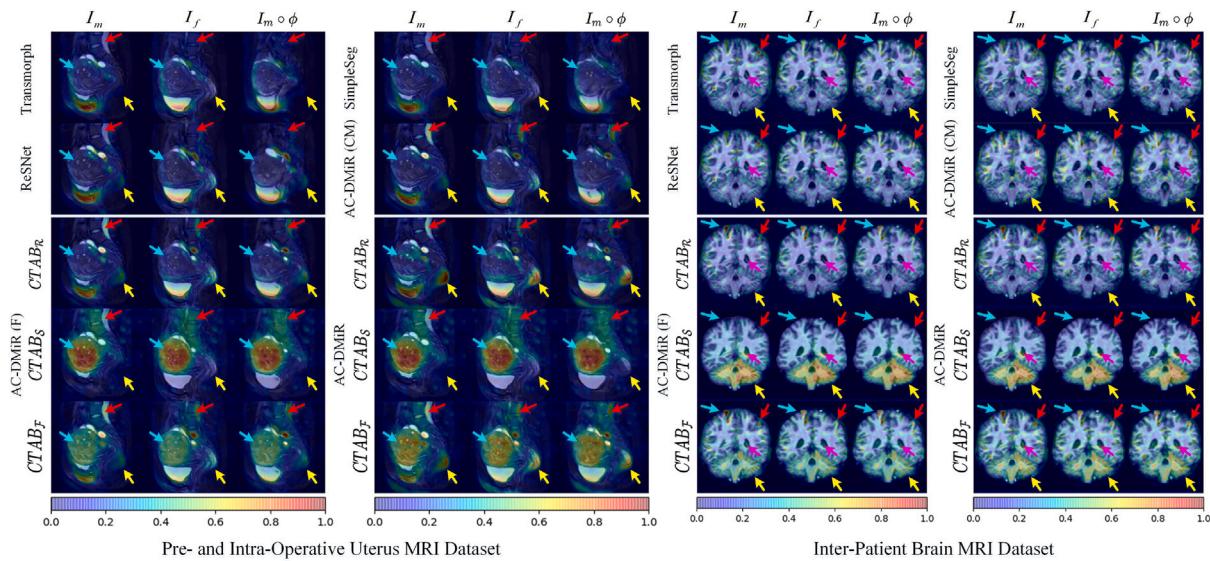


**Fig. 9.** A quantitative evaluation of the numerous registration methodologies on the pre- to intra-operative uterus MR registration task. Dice scores for several uterus MR substructures utilizing the proposed AC-DMiR and existing image registration approaches are shown in boxplots.

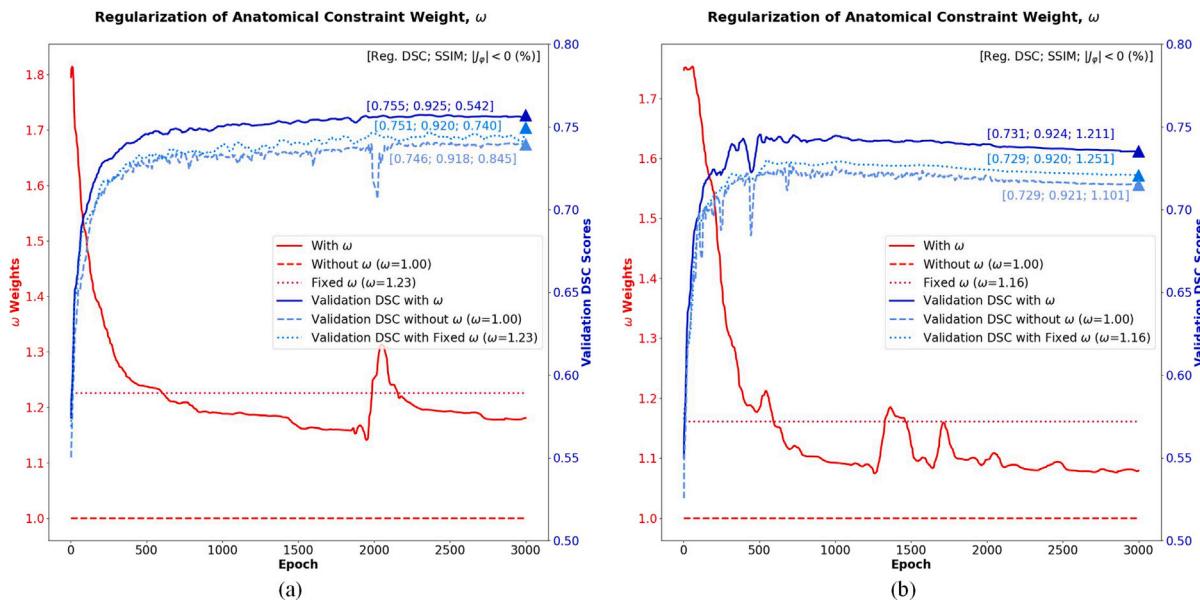
the attention masks' focused layouts varied and were impressively task-specific across different tasks.

For the uterus registration task,  $CTAB_S$  is able to capture more discriminative representation of anatomical regions (as indicated by cyan and red arrows).  $CTAB_F$  integrate features from  $CTAB_S$  and

$CTAB_R$  to produce the final cross-task attention features. Moreover,  $CTAB_S$  can improve the adaptive learning ability of  $CTAB_R$  on several deformed organs with the integration of our proposed anatomical constraint weight  $\omega$  (AC-DMiR), which is indicated by yellow arrows. For the brain registration task,  $CTAB_S$  have excellent coverage of



**Fig. 10.** Visualization of attention maps on both pre- and intra-operative uterus MRI (left) and inter-patient brain MRI (right) registration tasks. The attention feature maps are overlay on the moving image  $I_m$ , fixed image  $I_f$ , and warped moving image  $I_m \circ \phi$ , respectively, in each of the first, second, and third columns. Each of the attention maps was color-coded using a gradient to show whether the feature maps had enhanced (higher values with red) or diminished (lower values with blue).



**Fig. 11.** The evolution of the anatomical constraint units,  $\omega$  weights during training for inter-patient brain (a) and pre- to intra-operative uterus (b) MR registration tasks on the validation dataset. We compared our strategy with different settings of  $\omega$ . Blue line and red line indicate the validation DSC score and the changes of  $\omega$  weights throughout the entire training process, respectively. The average of metric scores (registration DSC score, SSIM score, and  $|J_\varphi| \leq 0\% \text{ (%)}$ ) obtained during testing phase were depicted at the final epoch of each strategy.

the deformed brain structure where  $CTAB_R$  pay less attention to (as indicated by yellow and purple arrows). Meanwhile, the blue and red arrows indicated that with  $\omega$ ,  $CTAB_F$  is able to cover various specific regions with small and large deformations.

The analysis demonstrates that the attention maps acquired from the high-level features of the segmentation and registration task give great coverage of anatomical regions to highlight important and severely deformed anatomical structures. The attention maps in  $CTAB_F$  become finer to cover various particular anatomical locations. Different tasks exhibited different high-level attentions, which reweighted attention to regions associated with the target task more intensely. Moreover, the attention maps were modified to accommodate the organ structures of various subjects. This shows that our AC-DMiR network architecture design can be a viable method for removing subject-specific individual

variability in clinical data that has not undergone spatial normalization and is instead raw and minimally pre-processed.

#### 5.4. Analysis of anatomical constraint weight $\omega$

In this experiment, we investigate the performance (in terms of DSC score) of the various anatomical constraint strategies introduced in Section 3.2.5 in order to select the best constraint method for the main registration tasks. Analysis of the behavior of the proposed anatomical constraint during training facilitates the understanding of how our global segmentation-guided anatomical constrain interacts with our registration network. Fig. 11 shows Dice scores for both the brain and uterus MRI registration task when sweeping  $\omega$  in Eq. (14). The  $\omega = 1$  corresponds to AC-DMiR (F) mentioned in Section 5.1, while  $\omega = \mu_\omega$

**Table 3**

Quantitative evaluation results on pre- to intra-operative uterus MR registration task. The Dice score for registration and segmentation, the SSIM score and the percentage of voxels with a non-positive Jacobian determinant are evaluated for various registration methods. Bolded numbers indicate the best values.

Method	Registration dice score ↑		Segmentation dice score ↑		SSIM ↑		$ J_\phi  \leq 0 (\%) ↓$
	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$
w/o registration	$0.292 \pm 0.115$	0.283	–	–	$0.710 \pm 0.217$	0.708	–
Affine	$0.350 \pm 0.188$	0.341	–	–	$0.756 \pm 0.220$	0.747	–
SyN (Avants et al., 2008)	$0.448 \pm 0.188$	0.436	–	–	$0.865 \pm 0.114$	0.865	$0.025 \pm 0.543$
NiftyReg (Modat et al., 2010)	$0.507 \pm 0.158$	0.510	–	–	$0.858 \pm 0.120$	0.859	$0.035 \pm 0.403$
VoxelMorph (Balakrishnan et al., 2019)	$0.716 \pm 0.132$	0.714	–	–	$0.893 \pm 0.121$	0.893	$1.248 \pm 1.029$
VoxelMorph-diff (Dalca et al., 2019)	$0.642 \pm 0.187$	0.621	–	–	$0.865 \pm 0.112$	0.860	<0.0001
ViT-V-Net (Chen et al., 2021a)	$0.722 \pm 0.144$	0.716	–	–	$0.905 \pm 0.114$	0.891	$2.099 \pm 0.557$
TransMorph (Chen et al., 2022)	$0.718 \pm 0.148$	0.715	–	–	$0.909 \pm 0.217$	0.900	$1.875 \pm 0.384$
DeepAtlas (Xu and Niethammer, 2019)	$0.727 \pm 0.140$	0.725	$0.799 \pm 0.116$	0.791	$0.916 \pm 0.214$	0.907	$1.603 \pm 0.758$
Cross-stitch (Elmahdy et al., 2021)	$0.721 \pm 0.135$	0.722	<b><math>0.815 \pm 0.115</math></b>	<b>0.795</b>	$0.919 \pm 0.112$	0.912	$2.049 \pm 1.028$
AC-DMiR	<b><math>0.731 \pm 0.129</math></b>	<b>0.731</b>	$0.793 \pm 0.112$	0.785	<b><math>0.924 \pm 0.108</math></b>	<b>0.923</b>	$1.211 \pm 0.390$

**Table 4**

Quantitative evaluation results on inter-patient brain and pre- to intra-operative uterus MR registration tasks when training with various  $\eta$  parameters. The Dice score for registration and segmentation, the SSIM score and the percentage of voxels with a non-positive Jacobian determinant are evaluated. Bolded numbers indicate the best values.

Task	$\eta$ settings	Registration dice score ↑		Segmentation dice score ↑		SSIM ↑		$ J_\phi  \leq 0 (\%) ↓$
		$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$	Median	$\mu \pm \sigma$
Brain MRI registration	0.5	$0.750 \pm 0.218$	0.751	$0.619 \pm 0.185$	0.615	$0.920 \pm 0.165$	0.916	$0.651 \pm 0.321$
	1	<b><math>0.755 \pm 0.112</math></b>	<b>0.754</b>	$0.621 \pm 0.110$	0.619	<b><math>0.925 \pm 0.203</math></b>	<b>0.918</b>	$0.542 \pm 0.397$
	5	$0.751 \pm 0.198$	0.749	$0.623 \pm 0.155$	0.621	$0.922 \pm 0.289$	0.915	$0.712 \pm 0.481$
	10	$0.748 \pm 0.201$	0.747	<b><math>0.624 \pm 0.167</math></b>	<b>0.622</b>	$0.918 \pm 0.297$	0.910	$0.771 \pm 0.215$
Uterus MRI registration	0.5	$0.728 \pm 0.181$	0.727	$0.790 \pm 0.185$	0.781	$0.922 \pm 0.148$	0.920	$1.569 \pm 0.418$
	1	<b><math>0.731 \pm 0.129</math></b>	<b>0.731</b>	$0.793 \pm 0.112$	0.785	<b><math>0.924 \pm 0.108</math></b>	<b>0.923</b>	$1.211 \pm 0.390$
	5	$0.730 \pm 0.166$	0.729	$0.797 \pm 0.163$	0.788	$0.921 \pm 0.214$	0.919	$1.617 \pm 0.520$
	10	$0.725 \pm 0.178$	0.724	<b><math>0.799 \pm 0.190</math></b>	<b>0.790</b>	$0.915 \pm 0.226$	0.915	$1.784 \pm 0.597$

(i.e.,  $\mu_\omega$  represents the mean value of  $\omega$  obtained during the training process) corresponds to AC-DMiR trained with only fixed  $\omega$ , without information flow improvement throughout the training process. In general, AC-DMiR with our proposed anatomical constraint strategy significantly outperforms AC-DMiR with  $\omega = 1$  and  $\omega = \mu_\omega$  settings in terms of DSC score (largest  $p$ -value  $< 10^{-9}$  when compared AC-DMiR with the other two  $\omega$  settings). Dice score during validation phase generally increases with obtained correlation between registration and segmentation tasks. Furthermore, the average of metric scores (registration DSC score, SSIM score, and  $|J_\phi| \leq 0 (\%)$ ) obtained during testing phase depicted at the final epoch of each settings shows that the proposed anatomical constraint method outperforms the AC-DMiR network with  $\omega = 1$  and  $\omega = \mu_\omega$  settings.

### 5.5. Analysis of weight $\eta$ for segmentation loss

**Table 4** tabulates the quantitative results for a range of  $\eta$  settings in the proposed AC-DMiR across both registration tasks. The registration accuracy starts to deteriorate when  $\eta$  is large. As  $\eta$  keep increasing, AC-DMiR has already been involved in over-fitting, in which our model continues to improve the DSC score for the segmentation network while harming the main registration task's accuracy. By utilizing segmentation networks as anatomical regularizers, we can observe that the optimum value of  $\eta = 1$  for the segmentation loss provides a good regularization of anatomical constraints for both brain and uterus MR registration tasks.

## 6. Discussion and conclusion

In this study, we formulate the registration and segmentation tasks as a multi-task learning setting. In order to achieve this, we provided a range of approaches using both the loss function and various architectural methodologies. We conducted experiments with various network architectures to determine which network setting maximizes the flow of prior information between these tasks. The proposed method is able to transfer high-level information between these two tasks by utilizing a feature-sharing mechanism via a task-shared anatomical correlation

module, which led to the best registration quality scores. Due to the precise inference of AC-DMiR, the proposed approach is highly promising for automatic registration of follow-up scans for medical image analysis.

A fundamental task in medical image analysis is the ability to consistently locate the same anatomical structure in various scans. For the auxiliary task in this study, we utilize semantic segmentation, but for this to be effective, we need to have a great number of labeled data for each anatomical structure of interest. In order to increase the potential of the proposed approach for multi-modal registration tasks, we will explore alternate contrastive learning strategies for learning anatomical embeddings that are consistent across body parts (i.e., without anatomical labels in training).

In conclusion, we proposed a novel multi-task learning framework for deformable medical image registration which effectively leverages the guidance from the supervision of auxiliary segmentation task to improve the realism of the warped moving image. The proposed AC-DMiR explicitly learns the correlation between registration and segmentation tasks to better transfer this anatomical prior knowledge to the target task. To achieve this, a task-shared anatomical correlation module is proposed. We further made use of the anatomical constraint, approximated by the prediction discrepancy from the task-specific initial segmentation network, to refine the realism of our registration predictions. By integrating our approach into an existing unsupervised registration framework, we achieved state-of-the-art registration accuracy on both inter-patient brain MRI and pre- to intra-operative MRI registration tasks. The findings confirmed our motivation and highlighted the significance of capturing the anatomical correlation between tasks to improve the registration method performance in terms of accuracy and realism.

### CRediT authorship contribution statement

**Hee Guan Khor:** Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Guochen Ning:** Data curation, Funding acquisition. **Yihua Sun:** Formal analysis, Writing – review

& editing. **Xu Lu**: Formal analysis, Writing – review & editing. **Xinran Zhang**: Formal analysis, Writing – review & editing. **Hongen Liao**: Conceptualization, Supervision, Methodology, Resources, Writing – review & editing, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data.

## Acknowledgments

The authors would like to acknowledge the supports of National Natural Science Foundation of China (82027807, 62201315, U22A2051), National Key Research and Development Program of China (2022YFC2405200), Beijing Municipal Natural Science Foundation, China (7212202) and China Postdoctoral Science Foundation (2021M701928).

## References

- Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., et al., 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* 25 (6), 954–961.
- Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. *NeuroImage* 38 (1), 95–113. <http://dx.doi.org/10.1016/j.neuroimage.2007.07.007>, URL: <https://www.sciencedirect.com/science/article/pii/S1053811907005848>.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12 (1), 26–41.
- Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* 38 (8), 1788–1800.
- Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., Shen, D., 2018. Deep learning based inter-modality image registration supervised by intra-modality similarity. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 55–63.
- Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y., 2022. Transmorph: Transformer for unsupervised medical image registration. *Med. Image Anal.* 82, 102615.
- Chen, J., He, Y., Frey, E.C., Li, Y., Du, Y., 2021a. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration. arXiv preprint [arXiv:2104.06468](https://arxiv.org/abs/2104.06468).
- Chen, Y., Xing, L., Yu, L., Liu, W., Pooya Fahimian, B., Niedermayr, T., Bagshaw, H.P., Buyounouski, M., Han, B., 2021b. MR to ultrasound image registration with segmentation-based learning for HDR prostate brachytherapy. *Med. Phys.* 48 (6), 3074–3083.
- Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2019. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Med. Image Anal.* 57, 226–236. <http://dx.doi.org/10.1016/j.media.2019.07.006>, URL: <https://www.sciencedirect.com/science/article/pii/S1361841519300635>.
- Dice, L.R., 1945. Measures of the amount of ecological association between species. *Ecology* 26 (3), 297–302.
- Elmahdy, M.S., Beljaards, L., Yousefi, S., Sokooti, H., Verbeek, F., Van Der Heide, U.A., Staring, M., 2021. Joint registration and segmentation via multi-task learning for adaptive radiotherapy of prostate cancer. *IEEE Access* 9, 95551–95568.
- Elmahdy, M.S., Jagt, T., Zinkstok, R.T., Qiao, Y., Shahzad, R., Sokooti, H., Yousefi, S., Incrocci, L., Marijnen, C., Hoogeman, M., et al., 2019a. Robust contour propagation using deep learning and image registration for online adaptive proton therapy of prostate cancer. *Med. Phys.* 46 (8), 3329–3343.
- Elmahdy, M.S., Wolterink, J.M., Sokooti, H., İsgüm, I., Staring, M., 2019b. Adversarial optimization for joint registration and segmentation in prostate CT radiotherapy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. In: Lecture Notes in Computer Science, vol. 11769, Springer, pp. 366–374.
- Estienne, T., Vakalopoulou, M., Christodoulidis, S., Battistela, E., Lerousseau, M., Carre, A., Klausner, G., Sun, R., Robert, C., Mougiakakou, S., et al., 2019. U-ResNet: Ultimate coupling of registration and segmentation with deep nets. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. In: Lecture Notes in Computer Science, vol. 11766, Springer, pp. 310–319.
- Fechter, T., Adebarh, S., Baltas, D., Ben Ayed, I., Desrosiers, C., Dolz, J., 2017. Esophagus segmentation in CT via 3D fully convolutional neural network and random walk. *Med. Phys.* 44 (12), 6341–6352.
- Ferrante, E., Dokania, P.K., Marini, R., Paragios, N., 2017. Deformable registration through learning of context-specific metric aggregation. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 256–265.
- Ferrante, E., Dokania, P.K., Silva, R.M., Paragios, N., 2018. Weakly supervised learning of metric aggregations for deformable image registration. *IEEE J. Biomed. Health Inf.* 23 (4), 1374–1384.
- Fischl, B., 2012. FreeSurfer. *NeuroImage* 62 (2), 774–781.
- Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X., 2020. Deep learning in medical image registration: a review. *Phys. Med. Biol.* 65 (20), 20TR01.
- Hu, L., Bell, D., Antani, S., Xue, Z., Yu, K., Horning, M.P., Gachoui, N., Wilson, B., Jaiswal, M.S., Befano, B., et al., 2019. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *JNCI: J. Natl. Cancer Inst.* 111 (9), 923–932.
- Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C.M., Emberton, M., Noble, J.A., Barratt, D.C., Vercauteren, T., 2018. Label-driven weakly-supervised learning for multimodal deformable image registration. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 1070–1074.
- Huo, Y., Xu, Z., Xiong, Y., Aboud, K., Parvatheneni, P., Bao, S., Bermudez, C., Resnick, S.M., Cutting, L.E., Landman, B.A., 2019. 3D whole brain segmentation using spatially localized atlas network tiles. *NeuroImage* 194, 105–119.
- Kiljunen, T., Akram, S., Niemelä, J., Löytyniemi, E., Seppälä, J., Heikkilä, J., Vuolukka, K., Kääriäinen, O.-S., Heikkilä, V.-P., Lehtiö, K., et al., 2020. A deep learning-based automated CT segmentation of prostate cancer anatomy for radiation therapy planning-a retrospective multicenter study. *Diagnostics* 10 (11), 959.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- Kokkinos, I., 2017. Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6129–6138.
- Kuckertz, S., Papenberg, N., Honegger, J., Morgas, T., Haas, B., Heldmann, S., 2020. Deep-learning-based CT-CBCT image registration for adaptive radio therapy. In: Medical Imaging 2020: Image Processing, Vol. 11313. SPIE, pp. 149–154.
- Li, B., Niessen, W.J., Klein, S., Groot, M.d., Ikram, M.A., Vernooij, M.W., Bron, E.E., 2019. A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. In: Lecture Notes in Computer Science, vol. 11766, Springer, pp. 645–653.
- Li, B., Niessen, W.J., Klein, S., Ikram, M.A., Vernooij, M.W., Bron, E.E., 2021. Learning unbiased group-wise registration (LUGR) and joint segmentation: evaluation on longitudinal diffusion MRI. In: Medical Imaging 2021: Image Processing, Vol. 11596. SPIE, pp. 136–144.
- Liu, F., Cai, J., Huo, Y., Cheng, C.-T., Raju, A., Jin, D., Xiao, J., Yuille, A., Lu, L., Liao, C., et al., 2020. Jssr: A joint synthesis, segmentation, and registration system for 3d multi-modal image alignment of large-scale pathological ct scans. In: European Conference on Computer Vision. Springer, pp. 257–274.
- Liu, C., Gardner, S.J., Wen, N., Elshaikh, M.A., Siddiqui, F., Movsas, B., Chetty, I.J., 2019. Automatic segmentation of the prostate on CT images using deep neural networks (DNN). *Int. J. Radiat. Oncol. Biol. Phys.* 104 (4), 924–932.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022.
- Lu, C., Chelikani, S., Papademetris, X., Knisely, J.P., Milosevic, M.F., Chen, Z., Jaffray, D.A., Staib, L.H., Duncan, J.S., 2011. An integrated approach to segmentation and nonrigid registration for application in image-guided pelvic radiotherapy. *Med. Image Anal.* 15 (5), 772–785.
- Mahapatra, D., Ge, Z., Sedai, S., Chakravorty, R., 2018. Joint registration and segmentation of xray images using generative adversarial networks. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 73–80.
- Mahapatra, D., Li, Z., Vos, F., Buhmann, J., 2015. Joint segmentation and groupwise registration of cardiac DCE MRI using sparse data representations. In: 2015 IEEE 12th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1312–1315.
- Maidens, J., Slamon, N.B., 2018. Artificial intelligence detects pediatric heart murmurs with cardiologist-level accuracy. *Circulation* 138 (Suppl.1), A12591.
- Mansilla, L., Milone, D.H., Ferrante, E., 2020. Learning deformable registration of medical images with anatomical constraints. *Neural Netw.* 124, 269–279.

- Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* 19 (9), 1498–1507.
- Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. *Comput. Methods Programs Biomed.* 98 (3), 278–284.
- Mok, T.C., Chung, A., 2020. Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4644–4653.
- Nielsen, R.K., Darkner, S., Feragen, A., 2019. Topaware: Topology-aware registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. In: Lecture Notes in Computer Science, vol. 11765, Springer, pp. 364–372.
- Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., De Marvao, A., Dawes, T., O'Regan, D.P., et al., 2017. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans. Med. Imaging* 37 (2), 384–395.
- Olut, S., Shen, Z., Xu, Z., Gerber, S., Niethammer, M., 2020. Adversarial data augmentation via deformation statistics. In: European Conference on Computer Vision. Springer, pp. 643–659.
- Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M., 2006. A Bayesian model for joint segmentation and registration. *NeuroImage* 31 (1), 228–239.
- Qiu, H., Qin, C., Schuh, A., Hammernik, K., Rueckert, D., 2021. Learning diffeomorphic and modality-invariant registration using B-splines. In: Heinrich, M., Dou, Q., de Bruijne, M., Lellmann, J., Schläfer, A., Ernst, F. (Eds.), Proceedings of the Fourth Conference on Medical Imaging with Deep Learning. In: Proceedings of Machine Learning Research, vol. 143, PMLR, pp. 645–664, URL: <https://proceedings.mlr.press/v143/qiu21a.html>.
- Qiu, L., Ren, H., 2021. U-RSNet: An unsupervised probabilistic model for joint registration and segmentation. *Neurocomputing* 450, 264–274.
- Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* 18 (8), 712–721.
- Rühaak, J., Polzin, T., Heldmann, S., Simpson, I.J.A., Handels, H., Modersitzki, J., Heinrich, M.P., 2017. Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration. *IEEE Trans. Med. Imaging* 36, 1746–1757.
- Shen, Z., Xu, Z., Olut, S., Niethammer, M., 2020. Anatomical data augmentation via fluid-based image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. In: Lecture Notes in Computer Science, vol. 12263, Springer, pp. 318–328.
- Tschandl, P., Codella, N., Akay, B.N., Argenziano, G., Braun, R.P., Cabo, H., Gutman, D., Halpern, A., Helba, B., Hofmann-Wellenhof, R., et al., 2019. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol.* 20 (7), 938–947.
- Unal, G., Slabaugh, G., 2005. Coupled PDEs for non-rigid registration and segmentation. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1. IEEE, pp. 168–175.
- Vandenbende, S., Georgoulis, S., Gool, L.V., 2020. Mti-net: Multi-scale task interaction networks for multi-task learning. In: European Conference on Computer Vision. Springer, pp. 527–543.
- Vandewincke, L., Willems, S., Robben, D., Van Der Veen, J., Crijns, W., Nuysts, S., Maes, F., 2020. Segmentation of head-and-neck organs-at-risk in longitudinal CT scans combining deformable registrations and convolutional neural networks. *Comput. Methods Biomed. Eng.: Imaging Vis.* 8 (5), 519–528.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13 (4), 600–612.
- Wang, H., Cao, Y., Syeda-Mahmood, T., 2014. Multi-atlas segmentation with learning-based label fusion. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 256–263.
- Wang, X., Girshick, R., Gupta, A., He, K., 2018. Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7794–7803.
- Xu, Z., Niethammer, M., 2019. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. In: Lecture Notes in Computer Science, vol. 11765, Springer, pp. 420–429.
- Xu, D., Ouyang, W., Wang, X., Sebe, N., 2018. Pad-net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 675–684.
- Yezzi, A., Zöllei, L., Kapur, T., 2003. A variational framework for integrating segmentation and registration through active contours. *Med. Image Anal.* 7 (2), 171–185.