



Published in final edited form as:

Med Image Anal. 2021 July ; 71: 102041. doi:10.1016/j.media.2021.102041.

Anatomy-guided Multimodal Registration by Learning Segmentation without Ground Truth: Application to Intra-procedural CBCT/MR Liver Segmentation and Registration

Bo Zhou^a, Zachary Augenfeld^a, Julius Chapiro^b, S. Kevin Zhou^d, Chi Liu^{a,b}, James S. Duncan^{a,b,c,*}

^aDepartment of Biomedical Engineering, Yale University, New Haven, CT, USA

^bDepartment of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA

^cDepartment of Electrical Engineering, Yale University, New Haven, CT, USA.

^dInstitute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

Abstract

Multimodal image registration has many applications in diagnostic medical imaging and image-guided interventions, such as Transcatheter Arterial Chemoembolization (TACE) of liver cancer guided by intra-procedural CBCT and pre-operative MR. The ability to register peri-procedurally acquired diagnostic images into the intra-procedural environment can potentially improve the intra-procedural tumor targeting, which will significantly improve therapeutic outcomes. However, the intra-procedural CBCT often suffers from suboptimal image quality due to lack of signal calibration for Hounsfield unit, limited FOV, and motion/metal artifacts. These non-ideal conditions make standard intensity-based multimodal registration methods infeasible to generate correct transformation across modalities. While registration based on anatomic structures, such as segmentation or landmarks, provides an efficient alternative, such anatomic structure information is not always available. One can train a deep learning-based anatomy extractor, but it requires large-scale manual annotations on specific modalities, which are often extremely time-consuming to obtain and require expert radiological readers. To tackle these issues, we leverage annotated datasets already existing in a source modality and propose an anatomy-preserving domain adaptation to segmentation network (APA2Seg-Net) for learning segmentation without target

*Corresponding author. bo.zhou@yale.edu (Bo Zhou), james.duncan@yale.edu (James S. Duncan).

Credit authorship contribution statement

Bo Zhou: Conceptualization, Methodology, Software, Visualization, Validation, Formal analysis, Writing original draft. **Zachary Augenfeld:** Methodology, Writing - review and editing. **Julius Chapiro:** Writing - review and editing. **S. Kevin Zhou:** Writing - review and editing. **Chi Liu:** Writing - review and editing, Supervision. **James S. Duncan:** Conceptualization, Methodology, Writing - review and editing, Supervision.

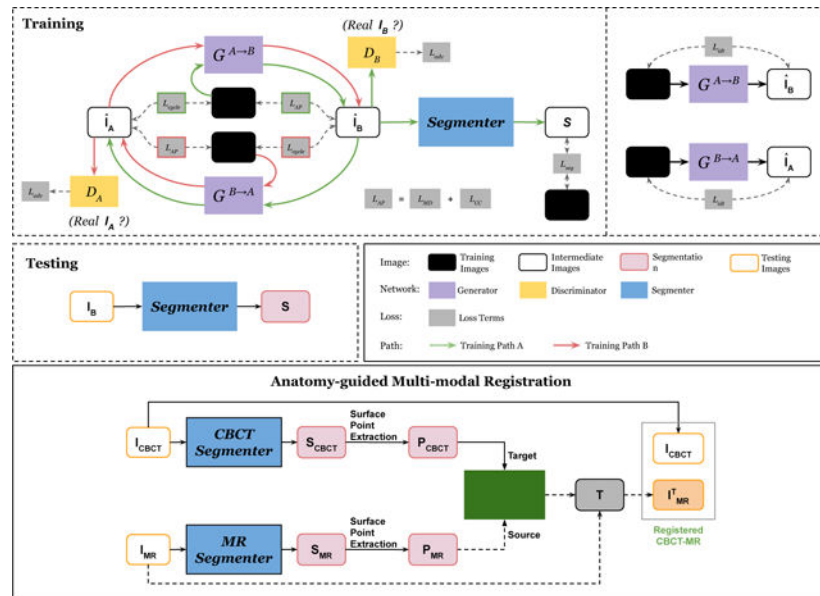
Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of interests

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.

modality ground truth. The segmenters are then integrated into our anatomy-guided multimodal registration based on the robust point matching machine. Our experimental results on in-house TACE patient data demonstrated that our APA2Seg-Net can generate robust CBCT and MR liver segmentation, and the anatomy-guided registration framework with these segmenters can provide high-quality multimodal registrations.

Graphical Abstract



Keywords

Multimodal registration; Unsupervised segmentation; Image-guided intervention; Cone-beam Computed Tomography

1. Introduction

Primary liver cancer is the fourth most common cancer and the third most common cause of cancer-related mortality worldwide with incidence rates rising across the globe and especially in the United States and Europe (Bray et al., 2018). Local image-guided therapies, such as Transcatheter Arterial Chemoembolization (TACE), are commonly used procedures that are performed in patients with intermediate to advanced stages as a palliative therapy option, capable of significantly prolonging patient survival (Pung et al., 2017). Most patients undergo multi-parametric multi-phasic contrast-enhanced MRI using gadolinium-enhanced T1 sequences both for diagnostic purposes as well as for the sake of therapy planning. This readily available multi-parametric information on tumor vascularity, size, location and even tissue properties is clinically underutilized for intra-procedural navigation primarily for technical reasons such as lack of practical image registration solutions. Intra-procedural navigation and targeting is instead achieved with serial, planar angiographic imaging as well as intra-procedural Cone-beam Computed Tomography (CBCT) imaging that provides a coarse cross-sectional dataset, which can then be used to map arterial supply of the tumor

and allow for accurate catheter guidance and intra-procedural feedback. While CBCT utilizes an x-ray source and the high-resolution 2D flat panel detector enables fast 3D organ visualization during procedures, CBCT suffers from low contrast-to-noise ratio (CNR), narrow abdominal tissue dynamic range, limited field-of-view (FOV), and motion/metal-induced artifacts, making it challenging to directly visualize and localize targeted tumors (Tacher et al., 2015; Pung et al., 2017). Therefore, multimodal image registration, i.e. mapping preoperative MR imaging and associated liver segmentations to intraprocedural CBCT is essential for accurate liver/tumor localization, targeting and subsequent drug delivery. Current workflows do not apply any quantitative measurements on the acquired CBCT images and the predominant technique is mere gestalt assessment of the images. Automatic multimodal registration is therefore highly desirable in image-guided interventional procedures.

Previous multimodal/monomodal image registration algorithms can be categorized into two classes: conventional iterative based approaches (Wyawahare et al., 2009; Maes et al., 1997; Avants et al., 2008; Rohr et al., 2001; Heinrich et al., 2012) and deep learning based approaches (Hu et al., 2018; Qin et al., 2019; Lee et al., 2019; Zhou et al., 2020a; Wang and Zhang, 2020; Arar et al., 2020; Mok and Chung, 2020). Conventional approaches utilize iterative maximization of intensity similarity metrics, such as mutual information (Maes et al., 1997), cross correlation (Avants et al., 2008) and difference in MIND (Heinrich et al., 2012), to find the optimal registration transformation between images. If paired key points between the images are available, landmark-based thin-plate splines registration can be applied to estimate the transformation between images. Previously, Al-Saleh et al. (2015, 2017) demonstrated the feasibility of temporomandibular joints MRI-CBCT registration via the above mentioned intensity-based and landmark-based registration methods. In the application of head and neck CBCT-CT registration, Zhen et al. (2012); Park et al. (2017) proposed to integrate the intensity matching between CT and CBCT into conventional iterative based approaches for more accurate CT-CBCT registrations. However, the intensity matching approaches are suitable for either monomodal or multimodal with similar imaging physics and cannot be adapted to CBCT-MR registration. More recently, Solbiati et al. (2018) proposed a two-stage registration for CBCT-CT liver registration, where manually annotated key points in the first stage are used for coarse alignment and conventional iterative registration based on mutual information is subsequently performed to refine the alignment.

With the recent advances in data driven learning (Zhou et al., 2020b), deep learning based methods have achieved comparable registration performance with a significantly higher inference speed. For monomodal registration, Balakrishnan et al. (2019) proposed the first deep learning based registration method using a deep convolutional network to predict the registration transformation between monomodality images, called VoxelMorph. Mok and Chung (2020) further improved its registration performance by adding the symmetric diffeomorphic properties into the network design. Moreover, Wang and Zhang (2020) developed a learning-based registration framework, called DeepFLASH, that utilizes low dimensional band-limited space for efficient transformation field computing. For multimodal registration, Hu et al. (2018) proposed to use the organ segmentations for weakly supervised training the transformation estimation network, where intensity-neutral supervision makes

the multimodal registration feasible. However, accurate manual organ segmentation is required for their approach and thus limits its applications. As an alternative to this approach, Qin et al. (2019) proposed to estimate the non-rigid transformation from disentangled representation of multimodal image contents. There are also recent studies of multimodal image registration for natural images (Arar et al., 2020), where source image appearance is first translated to fix image appearance, and then previously established monomodal registration methods (Balakrishnan et al., 2019) are applied. Although all the above methods achieve impressive results, they are limited to multimodal registration with no occluded FOV, sufficiently wide intensity range, or organ segmentations. Those conditions are hardly satisfied in many image-guided intervention procedures, such as TACE. More recently, Augenfeld et al. (2020) proposed to use manual CBCT liver annotation to train a CBCT segmenter and register based on the predicted CBCT segmentation and manually annotated MR. While demonstrating the feasibility of registration in TACE, segmenting liver on intraprocedural image is not clinical routine and training such segmenter from limited annotation data impede the segmentation and registration performance.

To tackle these issues, we present an anatomy-guided registration framework by learning segmentation without target modality ground truth. In previous works of learning segmentation without target modality, Zhang et al. (2018a) proposed a two-step strategy, where they first use CycleGAN (Zhu et al., 2017) to adapt the target domain image to the domain with a well-trained segmenter, and then predict the segmentation on the adapted image. However, the segmentation performance relies on the image adaptation performance, thus the two-step process may prone to error aggregation. To improve the CycleGAN performance in medical imaging, Zhang et al. (2018b) suggested adding two segmenters as additional discriminators for generating shape-consistent image adaption results. However, ground-truth segmentation is required for both source and target domains. Recently, Yang et al. (2020) proposed to add the Modality Independent Neighborhood Descriptor (MIND) loss (Heinrich et al., 2012) in the CycleGAN to constrain image structure during the adaptation. Similarly, Ge et al. (2019) proposed to incorporate correlation coefficient loss in the CycleGAN to constrain image structure. Both strategies demonstrated improvements in MR-to-CT translation. On the other hand, Huo et al. (2018) proposed Syn2Seg-Net that merges CycleGAN with a segmentation network on the target domain output, such that the segmentation network is trained on the target domain without target domain ground truth. However, the training image of the segmentation network relies on high-quality adapted images from the CycleGAN part of SynSeg-Net. Without anatomy-preserving constraint during the adaptation, the image could be adapted to a target domain image with incorrect anatomical contents, and negatively impact the subsequent segmentation network's training. Inspired by Huo et al. (2018) and with large-scale manual liver segmentation on conventional CT available from public dataset, such as LiTS (Bilic et al., 2019) and CHAOS (Kavur et al., 2020), we propose an anatomy-preserving domain adaptation to segmentation network (APA2Seg-Net) for learning segmentation without CBCT/MR ground truth. Specifically, we aim to use only conventional CT segmentation to train robust CBCT/MR segmenters in an anatomy-preserving unpaired fashion. The extracted anatomic information of CBCT/MR, i.e liver segmentations, guides our Robust Point Matching (RPM) to estimate

the multimodal registration transformation. Our experimental results on TACE patients demonstrate that our APA2Seg-Net based registration framework allows us to get robust target modality segmenters without ground truth, and enables accurate multimodal registration. Our code is available at <https://github.com/bbbbbbzhou/APA2Seg-Net>.

2. Methods

We propose a novel two-stage multimodal registration framework for mapping pre-operative MR to intraprocedural CBCT for liver image-guided interventions. In the first stage, our APA2Seg-Net is trained with 3 sources of images: paired conventional CT with liver segmentation and unpaired CBCT/MR, such that CBCT/MR segmenters can be extracted from our APA2Seg-Net for outputting the anatomic information. In the second stage, we extract the surface points of the outputted CBCT and MR segmentations, and input them into our RPM machine to predict MR to CBCT transformation. Finally, the transformation is applied to the pre-operative MR and the associated labels to register to the intraprocedural CBCT. The details are discussed in following sections.

2.1. Anatomy-preserving Adaptation to Segmentation Network

Our APA2Seg-Net consists of two parts - an anatomy-preserving domain adaptation network (APA-Net) and a segmentation network. The architecture and training/test stages are shown in Figure 1. The APA-Net is a cyclic adversarial network based on Zhu et al. (2017) with the addition of anatomy content consistency regularization. As illustrated in Figure 1, APA-Net adapts images between two domains: the conventional CT domain \mathbb{A} and the CBCT/MR domain \mathbb{B} . The anatomy consistency regularization ensures organ and tumor content information are not lost during the unpaired domain adaptation process, thus critical for training a robust segmenter in domain \mathbb{B} . Specifically, our APA2Seg-Net contains five networks, including two generators, two discriminators and one segmenter. The generator $G^{A \rightarrow B}$ adapts images from the conventional CT domain to the CBCT/MR domain, the generator $G^{B \rightarrow A}$ adapts the inverse way, the discriminator D_B identifies real CBCT or the adapted ones from $G^{A \rightarrow B}$, the discriminator D_A identifies real conventional CT or the adapted ones from $G^{B \rightarrow A}$, and the segmenter M_B predicts the segmentation \hat{S}_B on adapted image from generator $G^{A \rightarrow B}$. There are two training paths in our APA2Seg-Net. Path A first adapts conventional CT images I_A to \hat{I}_B in the CBCT/MR domain through $G^{A \rightarrow B}$. Then, \hat{I}_B is adapted back to the conventional CT domain as \hat{I}_A through $G^{B \rightarrow A}$. In parallel, \hat{I}_B is also feed into segmenter M_B to generate segmentation prediction \hat{S}_B . Similarly, path B first adapts CBCT/MR images I_B to \hat{I}_A in the conventional CT domain through $G^{B \rightarrow A}$. Then, \hat{I}_A is adapted back to the CBCT/MR domain as \hat{I}_B through $G^{A \rightarrow B}$.

Training supervision comes from five sources:

(a) adversarial loss \mathcal{L}_{adv} utilizes discriminators to classify if adapted image belong to specific domain. The adversarial objective aims to encourage G to generate adapted images that are indistinguishable to the discriminators. Two adversarial losses are introduced to train generators and discriminators:

$$\begin{aligned} \mathcal{L}_{adv}(G^{A \rightarrow B}, D_B, I_B, \hat{I}_B) &= \mathbb{E}_{I_B \sim \mathbb{B}}[\log D_B(I_B)] \\ &+ \mathbb{E}_{I_A \sim \mathbb{A}}[\log(1 - D_B(G^{A \rightarrow B}(I_A)))] \end{aligned} \quad (1)$$

$$\begin{aligned} \mathcal{L}_{adv}(G^{B \rightarrow A}, D_A, I_A, \hat{I}_A) &= \mathbb{E}_{I_A \sim \mathbb{A}}[\log D_A(I_A)] \\ &+ \mathbb{E}_{I_B \sim \mathbb{B}}[\log(1 - D_A(G^{B \rightarrow A}(I_B)))] \end{aligned} \quad (2)$$

(b) cycle-consistency loss \mathcal{L}_{cycle} constrains the image that returns to the original domain after passing through two generators to have minimal alternation to image content, such that a compound of two generators should be an identity mapping:

$$\begin{aligned} \mathcal{L}_{cycle} &= \mathbb{E}_{I_A \sim \mathbb{A}} \left[\left\| G^{B \rightarrow A}(G^{A \rightarrow B}(I_A)) - I_A \right\|_2^2 \right] \\ &+ \mathbb{E}_{I_B \sim \mathbb{B}} \left[\left\| G^{A \rightarrow B}(G^{B \rightarrow A}(I_B)) - I_B \right\|_2^2 \right] \end{aligned} \quad (3)$$

(c) segmentation loss \mathcal{L}_{seg} on the segmentation prediction from image \hat{I}_B . The segmentation prediction should be consistent with the ground truth label from the conventional CT domain \mathbb{A} :

$$\mathcal{L}_{seg} = \mathbb{E}_{I_A \sim \mathbb{A}} \left[1 - \frac{2|M_B(G^{A \rightarrow B}(I_A)) \cap S_A|}{|M_B(G^{A \rightarrow B}(I_A))| + |S_A|} \right] \quad (4)$$

(d) identity loss \mathcal{L}_{idt} regularizes the generators to be near an identity mapping when real samples of the target domain are provided. For example, if a given image looks like it is from the target domain, the generator should not map it into a different image. Therefore, the identity loss is formulated as:

$$\begin{aligned} \mathcal{L}_{idt} &= \mathbb{E}_{I_B \sim \mathbb{B}} \left[\left\| G^{A \rightarrow B}(I_B) - I_B \right\|_2^2 \right] \\ &+ \mathbb{E}_{I_A \sim \mathbb{A}} \left[\left\| G^{B \rightarrow A}(I_A) - I_A \right\|_2^2 \right] \end{aligned} \quad (5)$$

(e) anatomy-preserving loss \mathcal{L}_{AP} enforces the anatomical content is preserved before and after adaptation. Unlike conventional CycleGAN (Zhu et al., 2017) that does not use direct content constraint, we use both the MIND loss (Yang et al., 2020; Heinrich et al., 2012) and correlation coefficient loss (Ge et al., 2019) to preserve the anatomy in our unpaired domain adaptation process:

$$\mathcal{L}_{AP} = \lambda_{cc} \mathcal{L}_{cc} + \lambda_{md} \mathcal{L}_{md} \quad (6)$$

The first term \mathcal{L}_{cc} is the correlation coefficient loss, and is formulated as:

$$\begin{aligned} \mathcal{L}_{cc} = & \mathbb{E}_{I_A \sim \mathbb{A}} \left[\frac{\text{Cov}(G^{A \rightarrow B}(I_A), I_A)}{\sigma_{G^{A \rightarrow B}(I_A)} \sigma_{I_A}} \right] \\ & + \mathbb{E}_{I_B \sim \mathbb{B}} \left[\frac{\text{Cov}(G^{B \rightarrow A}(I_B), I_B)}{\sigma_{G^{B \rightarrow A}(I_B)} \sigma_{I_B}} \right] \end{aligned} \quad (7)$$

where Cov is the variance operator and σ is the standard deviation operator. The second term \mathcal{L}_{md} is the MIND loss (Yang et al., 2020), and is formulated as:

$$\begin{aligned} \mathcal{L}_{md} = & \mathbb{E}_{I_B \sim \mathbb{B}} \left[\|F(G^{A \rightarrow B}(I_A)) - F(I_A)\|_1 \right] \\ & + \mathbb{E}_{I_A \sim \mathbb{A}} \left[\|F(G^{B \rightarrow A}(I_B)) - F(I_B)\|_1 \right] \end{aligned} \quad (8)$$

where F is a modal-independent feature extractor defined by:

$$F_x(I) = \frac{1}{Z} \exp\left(-\frac{K_x(I)}{V_x(I)}\right) \quad (9)$$

where $K_x(I)$ is a distance vector of image patches around voxel x with all the neighborhood patches within a non-local region in image I . $V_x(I)$ is the local variance at voxel x in image I . Here, dividing $K_x(I)$ with $V_x(I)$ aims to reduce the influence of image modality and intensity range, and Z is a normalization constant to ensure that the maximum element of F_x equals to 1. In our anatomy-preserving loss, the weight parameters are set to $\lambda_{cc} = 1$ and $\lambda_{md} = 1$ to achieve balanced training. The anatomy-preserving loss ensures the adaptation only alter the appearance of image while maintaining the anatomical content, such that segmentation network M_B can be trained correctly to recognize the anatomical content in the adapted image.

Finally, the overall objective is a weighted combination of all loss listed above:

$$\mathcal{L}_{all} = \lambda_1 \mathcal{L}_{cycle} + \lambda_2 \mathcal{L}_{adv} + \lambda_3 \mathcal{L}_{seg} + \lambda_4 \mathcal{L}_{idt} + \lambda_5 \mathcal{L}_{AP} \quad (10)$$

where weight parameters are set to $\lambda_1 = 10$ and $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1$ to achieve a balanced training and near-optimal performance according to our hyper-parameter search. The sub-networks' details are shown in Figure 2. Specifically, we use a decoder-encoder network with 9 residual bottleneck for our generators, a 3-layer CNN for our discriminators. Our segmenter is a 5-level UNet with concurrent SE module Roy et al. (2018) concatenated to each level's output.

2.2. Anatomy-guided Multimodal Registration

Using our APA2Seg-Net trained by conventional CT with liver segmentation, we can obtain CBCT and MR segmenters. Then, the segmenters are deployed in our anatomy-guided multimodal registration to guide the Robust Point Matching (RPM) machine to predict the transformation between MR and CBCT images. The registration pipeline is shown in Figure 3. The CBCT and MR segmenters from APA2Seg-Net predict the CBCT and MR

segmentations. Then, we extract the surface points from the CBCT and MR segmentations and input them into the RPM.

RPM is a point-based registration framework based on deterministic annealing and soft assignment of correspondences between point sets (Gold et al., 1998), which is robust to point outliers. Specifically, given two point sets \mathcal{M} and \mathcal{S} , RPM aims to find the affine transformation T that best relates the two point sets. Reformulating the transform T into transformation matrix A and translation vector t form, we have $T(m) = Am + t$ where A is composed of scale, rotation, horizontal shear, and vertical shear parameters, denoted as a , θ , b , and c , respectively. The registration cost function can be written as:

$$C = \sum_{j=1}^N \sum_{i=1}^M \mu_{ij} \|s_j - (Am_i + t)\|^2 + \alpha(a^2 + b^2 + c^2) - \beta \sum_{j=1}^N \sum_{i=1}^M \mu_{ij} \quad (11)$$

where μ is the point match matrix with $u_{ij} = 1$ if point m_i corresponds to point s_j and $u_{ij} = 0$ otherwise. The first term minimizes the distance between point sets, and the second term constrains the transformation to avoid large numbers or dramatic transformations. The third term biases the cost toward stronger point correlation by decreasing the cost function. Then, the cost function can be iteratively solved by the soft assignment algorithm. The soft assignment between point sets allows point registration with exclusion of the outlying points and avoids local minima, which fits the problem of CBCT-MR registration well, since the liver is often partially occluded in CBCT due to limited FOV. In addition, for the purpose of intraprocedural registration, RPM provides high-speed registration as it is based on points. The generated transformation T is then applied to original MR image to create registered CBCT-MR pair, which provides better visualization of tumor during image-guided intervention.

In our implementation, we extract the liver surface points from CBCT and MR segmentation for RPM. Other types of point features from segmentation can also be used in RPM, such as landmarks and skeletons.

3. Experimental Results

3.1. Data and Setup

In the conventional CT domain, we collected 131 and 20 CT volumes with liver segmentation from LiTS (Bilic et al., 2019) and CHAOS (Kavur et al., 2020), respectively. In the CBCT/MR domain, we collected 16 in-house TACE patients with both intraprocedural CBCT and pre-operative MR for our segmentation and registration evaluations. All the CBCT data were acquired using a Philips C-arm system with a reconstructed image size of $384 \times 384 \times 297$ and voxel size of $0.65 \times 0.65 \times 0.65 \text{ mm}^3$. The MR data were acquired using different scanners with different spatial resolutions. Thus, we resampled all the CBCT, MR and conventional CT to an isotropic spatial resolution of $1 \times 1 \times 1 \text{ mm}^3$.

In the CBCT APA2Seg-Net setup, the conventional CT inputs were first randomly cropped in the axial view using a circular spherical mask on the liver region to simulate the limited

FOV in CBCT, as demonstrated in Figure 4. The circular mask maintains the same cropping geometry observed in CBCT with a radius of 125mm . As a result, we obtained 13,241 2D conventional CT images with liver segmentation, and 3,792 2D CBCT images. In the MR APA2Seg-Net setup, both the conventional CT and MR input were zero-padded or cropped to keep an axial FOV of $410 \times 410\text{mm}^2$. As a result, we obtained 13,241 2D conventional CT images with liver segmentation, and 1,128 2D MR images. All the 2D images were resized to 256×256 for APA2Seg-Net inputs. With 16 TACE patients in our dataset, we performed four-fold cross-validation with 12 TACE used as training and 4 patients used as testing in each validation.

3.2. Segmentation Results

After training, we extracted the segmenters from the APA2Seg-Net for prediction of liver segmentations on both CBCT and MR. For qualitative study, we compared our segmentation performance with: **i.** CBCT/MR-to-CT CycleGAN concatenated with conventional CT segmenter (CycleGAN+Seg_{CT}), where the CT segmenter is trained on conventional CT images with liver annotations (Seg_{CT}); **ii.** APA2Seg-Net without MIND loss and CC loss for anatomy preserving constraint during the training (Ours-MD-CC); **iii.** APA2Seg-Net with MIND loss only for anatomy preserving constraint during the training (Ours+MD-CC); **iv.** APA2Seg-Net with CC loss only for anatomy preserving constraint during the training (Ours-MD+CC); **v.** APA2Seg-Net with both MIND loss and CC loss for anatomy preserving constraint during the training (Ours+MD+CC); and **vi.** the segmenter trained on target domain images with limited liver annotations (Supervised Seg_{CBCT} / Seg_{MRI}).

Qualitative comparison of CBCT segmentation results are shown in Figure 5. As we can see, CBCT in TACE suffers from limited FOV, metal artifacts, and low CNR. CycleGAN+Seg_{CT} is non-ideal because it requires adapting the input CBCT to conventional CT first, and the segmentation relies on the translated image quality. However, the unpaired and unconstrained adaption from CBCT to CT is difficult as it consists of metal artifact removal and liver boundary enhancement. The multi-stage inference in CycleGAN+Seg_{CT} aggregates the prediction error into the final segmentation. On the other hand, our APA2Seg-Net with anatomy-preserving constraint and one-stage inference mechanism achieved significantly better CBCT liver segmentation results. We found combining CC loss and MIND loss for our anatomy-preserving constraint in APA2Seg-Net yields the best results. We also found our identity loss that helps maintain the target domain feature during the adaptation process provides us better segmentation performance. Furthermore, compared to the segmenters trained on target domains using relatively limited annotation data (2844 2D images), our APA2Seg-Net trained from large-scale conventional CT data (13,241 2D images) can provide slightly better segmenters. Qualitative comparison of MR segmentation results are illustrated in Figure 6. Similar observations can be found in the MR segmentation results.

Dice Similarity Coefficient (DSC) and Average Symmetric Surface Distance (ASD) were used to evaluate the quantitative segmentation performance. Table 1 summarizes the quantitative comparison of CBCT and MR segmentation results. As we can see, our APA2Seg-Net achieved the best CBCT and MR segmentation in terms of DSC and ASD, indicating the best overall liver segmentation.

3.3. Registration Results

With the CBCT and MR segmenters extracted from APA2Seg-Nets, we integrate the segmenters into our anatomy-guided multimodal registration pipeline for registering MR to CBCT. The CBCT and MR liver segmentation from segmenters are inputted into RPM to generate the transformation parameters. For qualitative studies, we first compared our registration results with classical previous works of intensity-based affine registration and intensity-based B-spline registration (Wyawahare et al., 2009; Maes et al., 1997). We also compared our registration results with intensity-based affine/B-spline registration based on CT images translated from CBCT and MR using APA-Net - similar to the idea in Arar et al. (2020). Two examples are illustrated in Figure 7. The ground truth (GT) CBCT liver mask (green) and the transformed GT MR liver mask (blue) are overlaid on the CBCT image to qualitatively evaluate the registration performance. As we can observe, neither intensity-based registration methods can correctly estimate the MR transformation, while our anatomy-guided registration, as demonstrated in the last column of Figure 7, can more accurately map the MR to CBCT images. Compared to the RPM registration based on ground truth liver segmentations, our anatomy-guided registration based on APA2Seg-Net's segmenter provides similar registration performance. Additional registration results using our method are shown in Figure 8.

For quantitative registration evaluation, we first evaluated the averaged error of transformation parameters where the transformation from human annotation based RPM registration is used as ground truth. Based on 3D affine transformation equation:

$$\begin{aligned}
 T &= \begin{bmatrix} 1 & 0 & 0 & \Delta x \\ 0 & 1 & 0 & \Delta y \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & h_{xy} & h_{xz} & 0 \\ h_{yx} & 1 & h_{yz} & 0 \\ h_{zx} & h_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} s_x & s_x h_{xy} & s_x h_{xz} & \Delta x \\ s_y h_{yx} & s_y & s_y h_{yz} & \Delta y \\ s_z h_{zx} & s_z h_{zy} & s_z & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{12}
 \end{aligned}$$

twelve 3D affine transformation parameters were evaluated: s_x, s_y, s_z are the scaling factors on the x,y,z directions, $h_{xy}, h_{xz}, h_{yx}, h_{yz}, h_{zx}, h_{zy}$ are parameters that control the shear transformation, and $x, y,$ and z are the translation on the x,y,z directions. The average errors of the parameters are reported in Table 2. As we can observe, our registration method achieves the least errors in estimating transformation parameters.

To further validate our registration, we computed the DSC and ASD metrics between the human annotated CBCT liver segmentation and the transformed human annotated MR liver segmentation using transformation generated from different methods. The results are summarized in Table 3. Please note that due to the limited FOV of CBCT, the liver mask in CBCT is often truncated while the liver mask in MR is intact. Therefore, the upper limit/gold standard of the metrics are not DSC=1 and ASD=0, but assumed to be the registration results based human annotated liver segmentation. As we can observe from the table, our

segmenter-based RPM registration achieved a mean DSC of 0.847 that is comparable to the human annotation based RPM registration with a mean DSC of 0.853. Our anatomy-guided method is also significantly better than the intensity-based methods. In Figure 9, we visualize the case-by-case DSC/ASD differences between our method and the human annotation based registration. Our method can achieve similar registration performance across all 16 cases as compared to the human annotation based registration. A maximal DSC difference less than 0.07 and a maximal ASD difference less than 3mm can be observed. Furthermore, we compared our APA2Seg-Net segmenter-based RPM registration to the target domain supervised segmenter-based RPM registration in Table 3. We found that our method also outperforms the supervised method that requires annotations on the target domains with the difference significant at $p < 0.05$ for both DSC and ASD.

4. Conclusion and Discussion

In this work, we proposed an anatomy-guided registration framework by learning segmentation without target modality ground truth. Specifically, we developed an APA2Seg-Net to learn CBCT and MR segmenters without ground truth, which are then plugged into the anatomy-guided registration pipeline for mapping MR to CBCT. We overcame three major difficulties in multimodal image registration. First, we proposed an anatomy-based registration framework that utilizes point clouds of the segmented anatomy, instead of relying on multimodal image intensity which may have significant distribution differences. To obtain robust segmenters of target modality without ground truth, we proposed a segmentation network training scheme without using target modality ground truth, which mitigates the manual annotation requirement on the target modality. Then, we also proposed to use RPM-based point registration that is robust to partially occluded view (point outliers), a scenario commonly observed in TACE and other image-guided intervention procedures.

We demonstrated the successful application on TACE, in which pre-operative diagnostic MR are registered to intraprocedural CBCT for guiding TACE procedures. Firstly, our method achieved the superior segmentation performance even when compared to the fully supervised methods that requires annotations on the target domains. As annotating new domain data, i.e. intraprocedural CBCT, is not a clinical routine and is time-consuming, one may only obtain limited amount of labeled data for supervised training on the target domain. Thus, it cannot provide sufficient data variability for generating a robust model. On the other hand, our APA2Seg-Net utilizing large-scale conventional CT dataset offers much larger data variability, thus achieved superior segmentation performance even without using ground truth annotations from the target domain. Then, given the more robust segmenters from APA2Seg-Net, our registration pipeline based on these segmenters and RPM can also offer superior registration performance. In Table 3, our method is able to reduce the ASD between MR and CBCT liver segmentation from 4 cm based on previous intensity-based affine registration to 0.5cm, and reduce the translation difference from 9.6cm based on previous intensity-based affine registration to 2cm, as demonstrated in Table 2. With our method, the registration errors now fall within a more acceptable range. This allows MR to be more accurately registered to CBCT, reinforcing the utility of MR-derived information within the clinical TACE image guidance environment.

The presented work also has potential limitations. First of all, the CBCT segmentation performance is far from perfect with a mean DSC of 0.893. In our current APA2Seg-Net implementation, only 2D networks were considered in this study since the amount of training data is not large enough to train a robust 3D network. However, the proposed APA2Seg-Net can be extended to 3D with the expense of higher GPU memory consumption and longer computation time, which would potentially provide better segmentation results if a large amount of 3D training scans is available. As a matter of fact, Zhang et al. (2018b); Cai et al. (2019) had demonstrated the promising results from 3D synthesis and segmentation. On the other hand, our APA2Seg-Net is an open framework with flexibility in network components. While we used 2D Res-Net/Patch GAN/concurrent-SE-UNet as our generator/discriminator/segmenter, we do not claim optimality of the combination for segmentation. Other image segmentation networks, such as attention UNet (Oktay et al., 2018), multi-scale guided attention network (Sinha and Dolz, 2020), and adversarial image-to-image network (Yang et al., 2017), could be deployed in our framework and might yield better segmentation performance on different applications. Secondly, as the segmentation is imperfect from our APA2Seg-Net, it leads to difference between registration based on our segmentation and human segmentation. However, the impact of imperfect CBCT/MR segmentation is mitigated through our RPM based registration. As we can observe from Table 3 and Figure 7, the human segmentation based registration is very close to the APA2Seg-Net segmentation-based registration in terms of qualitative visualization and quantitative comparison (< 0.01 in terms of Dice). Thirdly, we considered affine transformation in our CBCT-MR liver registration, and incorporating non-rigid registration could potentially provide more accurate internal structure mappings. However, liver in CBCT is often truncated due to limited FOV. Therefore, using non-rigid registration would lead to incorrect matching on the truncated boundary. Future works on incorporating non-rigid registration while rejecting the point outliers outside of FOV is needed. Lastly, we evaluated the registration performance on the entire liver, while registration performance on other important landmarks, such as tumor location, is not included here. We will evaluate other landmark's alignment in our future clinical feasibility studies.

The design of our anatomy-guided registration framework by learning segmentation without ground truth also suggests several interesting topics for future studies. First of all, our method could be adapted to other multimodal registration tasks that conventional registration techniques are not applicable, such as MR-Ultrasound (US) registration for neurosurgery (Rivaz et al., 2014) and prostate interventions (Hu et al., 2012), where occluded FOV and intensity inhomogeneity is often observed in US. There are several public datasets containing MR brain tumor and MR prostate with ground truth segmentation Simpson et al. (2019), which make it possible to adapt our method to these applications. More specifically, we could use APA2Seg-Net to obtain the US and MR segmenters, which are then embedded into our anatomy-guided registration pipeline for real-time MR-US alignments. Similar to our idea in registration, Sultana et al. (2019) recently proposed a prostate US-PET/CT registration algorithm based on segmentation for dose planning Dréan et al. (2016), in which our APA2Seg-Net could potentially provide the US and PET/CT prostate segmenter as well. Secondly, our method could also be adapted to landmark-based registration tasks. While anatomy-guided registration framework based on segmentation is demonstrated in this work,

the segmenter in APA2Seg-Net could be replaced with a detector for learning keypoint detection without ground truth on target domain. Then, the keypoint detector could also be embedded into our anatomy-guided registration pipeline for keypoint based alignments.

In summary, we proposed an anatomy-guided registration framework by learning segmentation without target modality ground truth based on APA2Seg-Net. We demonstrated the successful application on intraprocedural CBCT-MR liver registration. In the future, we will assess the tumor-of-interest's registration accuracy and evaluate the clinical impact of real-time intraprocedural MR-CBCT liver registration.

Acknowledgments

This work was supported by funding from the National Institutes of Health/ National Cancer Institute (NIH/NCI) under grant number R01CA206180. BZ was supported by the Biomedical Engineering Ph.D. fellowship from Yale University.

References

- Al-Saleh M, Jaremko J, Alsufyani N, Jibri Z, Lai H, Major P, 2015. Assessing the reliability of mri-cbct image registration to visualize temporomandibular joints. *Dentomaxillofacial Radiology* 44, 20140244. [PubMed: 25734241]
- Al-Saleh MA, Alsufyani N, Lai H, Lagravere M, Jaremko JL, Major PW, 2017. Usefulness of mri-cbct image registration in the evaluation of temporomandibular joint internal derangement by novice examiners. *Oral surgery, oral medicine, oral pathology and oral radiology* 123, 249–256.
- Arar M, Ginger Y, Danon D, Bermanno AH, Cohen-Or D, 2020. Unsupervised multi-modal image registration via geometry preserving image-to-image translation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13410–13419.
- Augenfeld Z, Lin M, Chapiro J, Duncan J, 2020. Automatic multimodal registration via intraprocedural cone-beam ct segmentation using mri distance maps, in: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020)*, IEEE. pp. 1217–1220.
- Avants BB, Epstein CL, Grossman M, Gee JC, 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* 12, 26–41. [PubMed: 17659998]
- Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV, 2019. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging* 38, 1788–1800.
- Bilic P, Christ PF, Vorontsov E, Chlebus G, Chen H, Dou Q, Fu CW, Han X, Heng PA, Hesser J, et al., 2019. The liver tumor segmentation benchmark (lits). *arXiv preprint arXiv:1901.04056*.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, 2018. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 68, 394–424. [PubMed: 30207593]
- Cai J, Zhang Z, Cui L, Zheng Y, Yang L, 2019. Towards cross-modal organ translation and segmentation: A cycle-and shape-consistent generative adversarial network. *Medical image analysis* 52, 174–184. [PubMed: 30594770]
- Dréan G, Acosta O, Lafond C, Simon A, de Crevoisier R, Haigron P, 2016. Interindividual registration and dose mapping for voxelwise population analysis of rectal toxicity in prostate cancer radiotherapy. *Medical Physics* 43, 2721–2730. [PubMed: 27277019]
- Ge Y, Xue Z, Cao T, Liao S, 2019. Unpaired whole-body mr to ct synthesis with correlation coefficient constrained adversarial learning, in: *Medical Imaging 2019: Image Processing*, International Society for Optics and Photonics. p. 1094905.
- Gold S, Rangarajan A, Lu CP, Pappu S, Mjolsness E, 1998. New algorithms for 2d and 3d point matching: Pose estimation and correspondence. *Pattern recognition* 31, 1019–1031.

- Heinrich MP, Jenkinson M, Bhushan M, Matin T, Gleeson FV, Brady M, Schnabel JA, 2012. Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical image analysis* 16, 1423–1435. [PubMed: 22722056]
- Hu Y, Ahmed HU, Taylor Z, Allen C, Emberton M, Hawkes D, Barratt D, 2012. Mr to ultrasound registration for image-guided prostate interventions. *Medical image analysis* 16, 687–703. [PubMed: 21216180]
- Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, Wang G, Bandula S, Moore CM, Emberton M, et al., 2018. Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis* 49, 1–13. [PubMed: 30007253]
- Huo Y, Xu Z, Bao S, Assad A, Abramson RG, Landman BA, 2018. Adversarial synthesis learning enables segmentation without target modality ground truth, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 1217–1220.
- Kavur AE, Gezer NS, Bari M, Conze PH, Groza V, Pham DD, Chat-terjee S, Ernst P, Özkan S, Baydar B, et al., 2020. Chaos challenge– combined (ct-mr) healthy abdominal organ segmentation. arXiv preprint arXiv:2001.06535.
- Lee MC, Oktay O, Schuh A, Schaap M, Glocker B, 2019. Image-and-spatial transformer networks for structure-guided image registration, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 337–345.
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P, 1997. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging* 16, 187–198. [PubMed: 9101328]
- Mok TC, 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.
- Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, et al., 2018. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
- Park S, Plishker W, Quon H, Wong J, Shekhar R, Lee J, 2017. Deformable registration of ct and cone-beam ct with local intensity matching. *Physics in Medicine & Biology* 62, 927. [PubMed: 28074785]
- Pung L, Ahmad M, Mueller K, Rosenberg J, Stave C, Hwang GL, Shah R, Kothary N, 2017. The role of cone-beam ct in transcatheter arterial chemoembolization for hepatocellular carcinoma: a systematic review and meta-analysis. *Journal of Vascular and Interventional Radiology* 28, 334–341. [PubMed: 28109724]
- Qin C, Shi B, Liao R, Mansi T, Rueckert D, Kamen A, 2019. Unsupervised deformable registration for multi-modal images via disentangled representations, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 249–261.
- Rivaz H, Chen SJS, Collins DL, 2014. Automatic deformable mr-ultrasound registration for image-guided neurosurgery. *IEEE transactions on medical imaging* 34, 366–380. [PubMed: 25248177]
- Rohr K, Stiehl HS, Sprengel R, Buzug TM, Weese J, Kuhn M, 2001. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on medical imaging* 20, 526–534. [PubMed: 11437112]
- Roy AG, Navab N, Wachinger C, 2018. Concurrent spatial and channel squeeze & excitation in fully convolutional networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 421–429.
- Simpson AL, Antonelli M, Bakas S, Bilello M, Farahani K, Van Ginneken B, Kopp-Schneider A, Landman BA, Litjens G, Menze B, et al., 2019. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063.
- Sinha A, Dolz J, 2020. Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*.
- Solbiati M, Passera KM, Goldberg SN, Rotilio A, Ierace T, Pedicini V, Poretti D, Solbiati L, 2018. A novel ct to cone-beam ct registration method enables immediate real-time intraprocedural three-dimensional assessment of ablative treatments of liver malignancies. *Cardiovascular and interventional radiology* 41, 1049–1057. [PubMed: 29492634]

- Sultana S, Song DY, Lee J, 2019. Deformable registration of pet/ct and ultrasound for disease-targeted focal prostate brachytherapy. *Journal of Medical Imaging* 6, 035003. [PubMed: 31528661]
- Tacher V, Radaelli A, Lin M, Geschwind JF, 2015. How i do it: cone-beam ct during transarterial chemoembolization for liver cancer. *Radiology* 274, 320–334. [PubMed: 25625741]
- Wang J, Zhang M, 2020. Deepflash: An efficient network for learning-based medical image registration, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4444–4452.
- Wyawahare MV, Patil PM, Abhyankar HK, et al., 2009. Image registration techniques: an overview. *International Journal of Signal Processing, Image Processing and Pattern Recognition* 2, 11–28.
- Yang D, Xu D, Zhou SK, Georgescu B, Chen M, Grbic S, Metaxas D, Comaniciu D, 2017. Automatic liver segmentation using an adversarial image-to-image network, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 507–515.
- Yang H, Sun J, Carass A, Zhao C, Lee J, Prince JL, Xu Z, 2020. Unsupervised mr-to-ct synthesis using structure-constrained cyclegan. *IEEE Transactions on Medical Imaging*.
- Zhang Y, Miao S, Mansi T, Liao R, 2018a. Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 599–607.
- Zhang Z, Yang L, Zheng Y, 2018b. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9242–9251.
- Zhen X, Gu X, Yan H, Zhou L, Jia X, Jiang SB, 2012. Ct to cone-beam ct deformable registration with simultaneous intensity correction. *Physics in Medicine & Biology* 57, 6807. [PubMed: 23032638]
- Zhou B, Tsai YJ, Liu C, 2020a. Simultaneous denoising and motion estimation for low-dose gated pet using a siamese adversarial network with gate-to-gate consistency learning, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 743–752.
- Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, Prince JL, Rueckert D, Summers RM, 2020b. A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises. *ArXiv preprint arXiv:2008.09104*.
- Zhu JY, Park T, Isola P, Efros AA, 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232.

Highlights

- We proposed an anatomy-guided multimodal registration framework and demonstrated the successful application to CBCT-MR liver registration.
- For obtaining the anatomic structure information from CBCT/MR, we proposed an anatomy-preserving domain adaptation to segmentation network (APA2Seg-Net) that allows us to learn segmentation without ground truth.
- With the CBCT/MR segmenters obtained from APA2Seg-Net, we proposed a registration pipeline that allows MR and MR-derived information to be accurately registered to CBCT liver to facilitate the TACE procedure.
- We believe our design can be extended to similar image-guided intervention procedures, when standard registration techniques are limited due to the non-ideal image quality conditions.

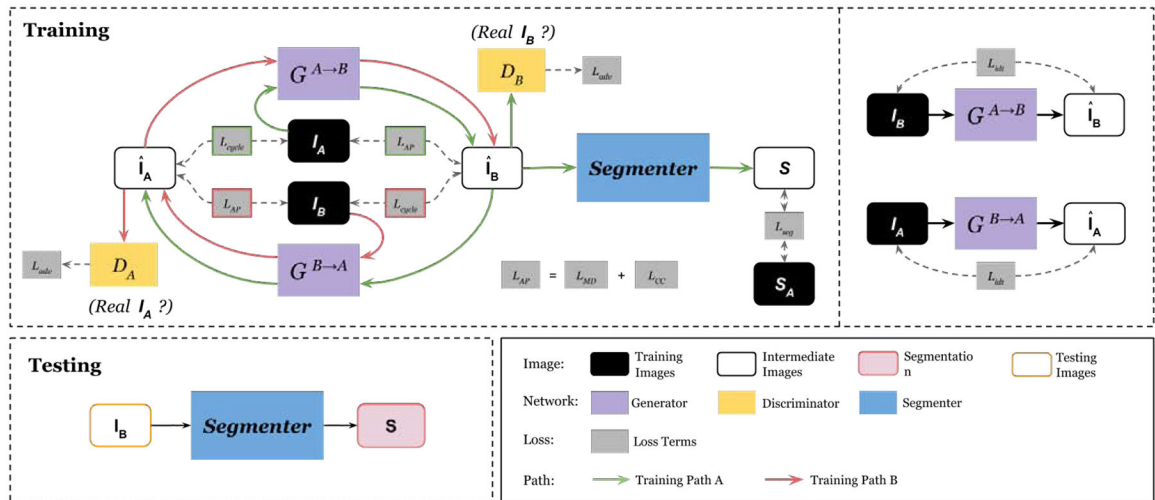


Fig. 1. Illustration of our Anatomy-Preserving domain Adaptation to Segmentation Network (APA2Seg-Net). It consists of an anatomy-preserving domain adaptation network (left portion), and a segmenter for target domain. During test phase, the segmenter is extracted from APA2Seg-Net for predicting structural information, i.e., segmentation.

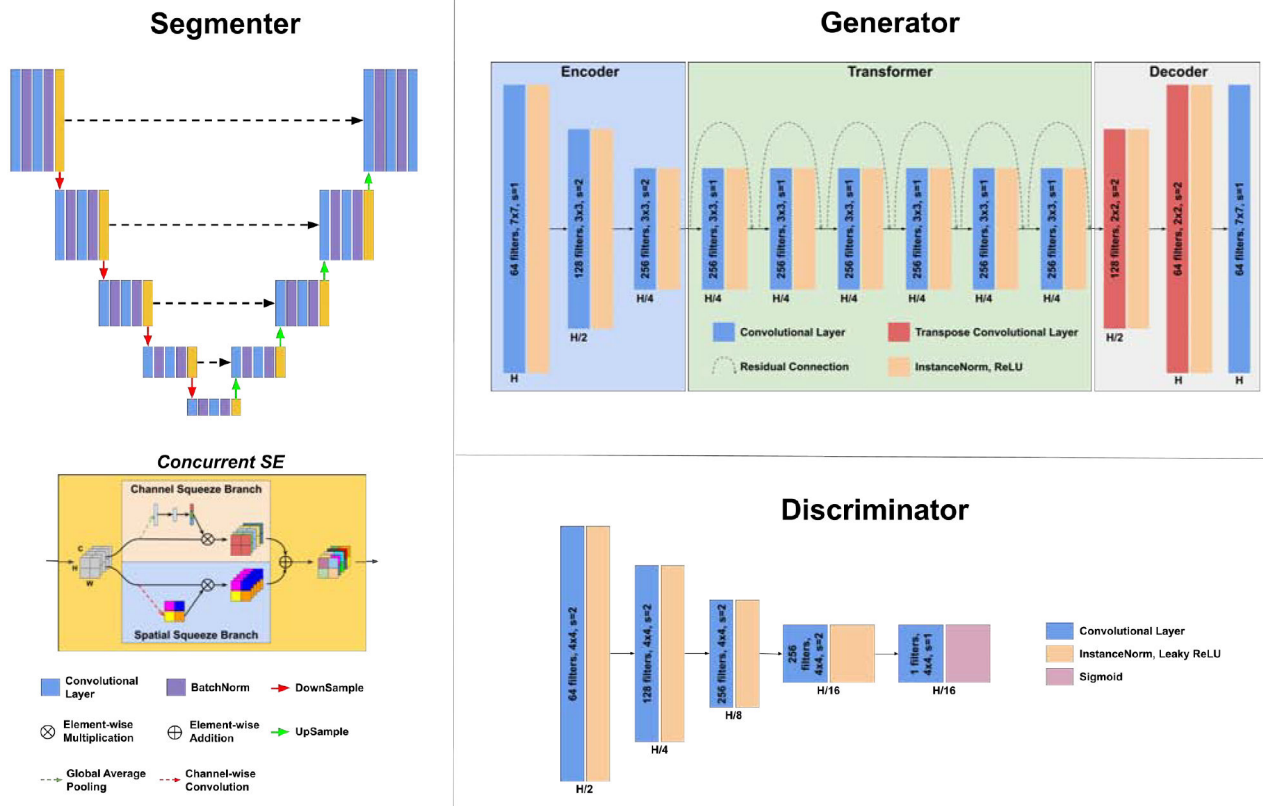


Fig. 2. Illustration of our segementer, generator, and discriminator network structures in the APA2Seg-Net. 5-level U-Net with the concurrent squeeze and excitation module is used for our segementer. An autoencoder with multiple residual bottleneck is used for our generator. The feature size shrinks in the encoder phase, stays constant in the transformer phase, and expands again in the decoder phase. The feature size of the layer outputs is listed below it, in terms of the input image size, H. On each layer is listed the number of filters, the size of those filters, and the stride.

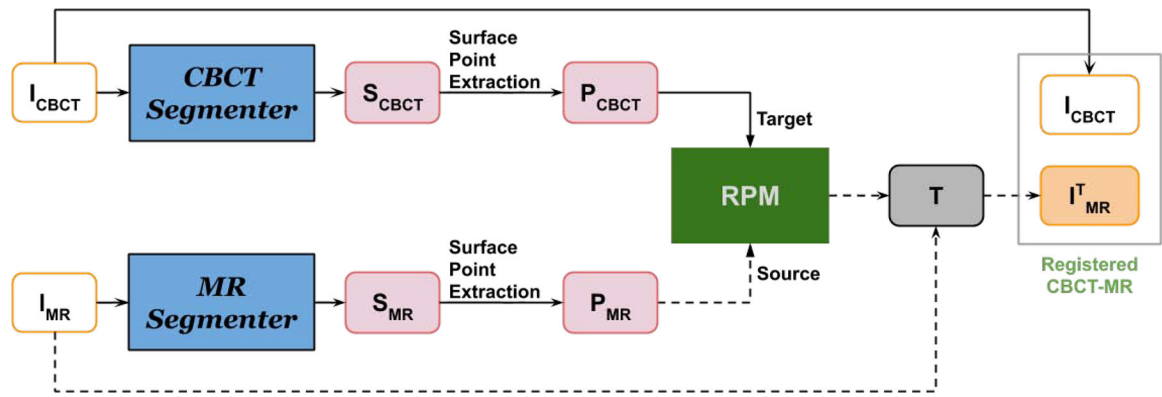


Fig. 3. Our anatomy-guided multimodal registration pipeline. The segmenters are obtained from APA2Seg-Net in Figure 1.

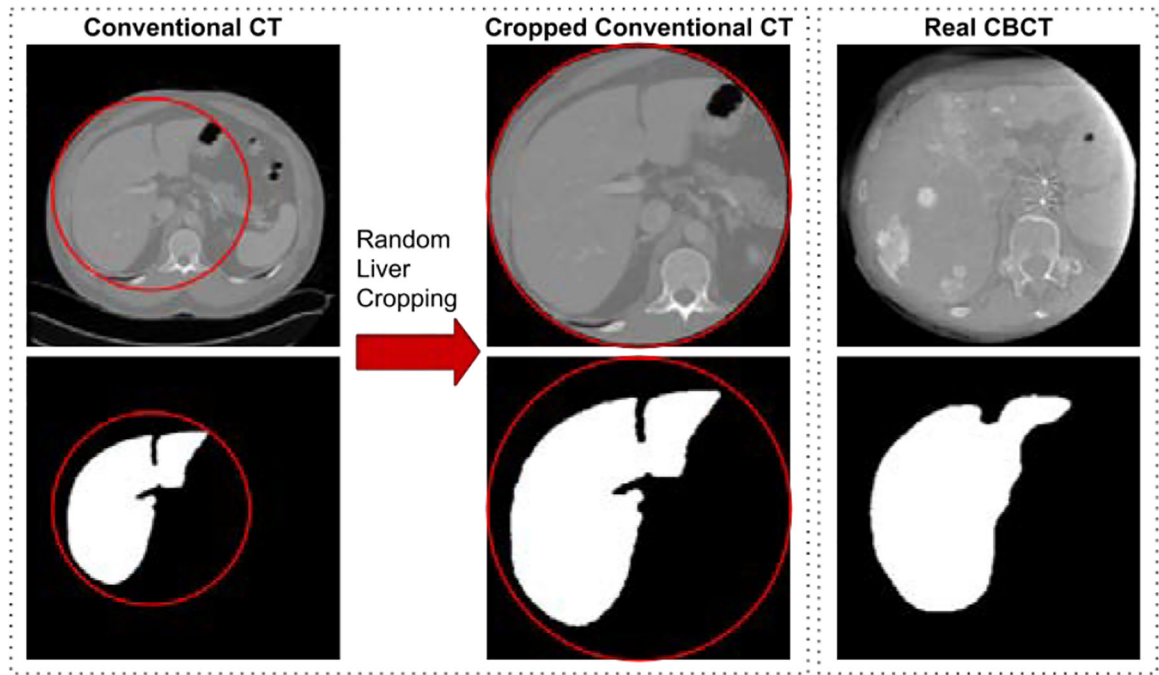


Fig. 4. Generation of limited FOV CT from conventional CT as input for our APA2Seg-Net.

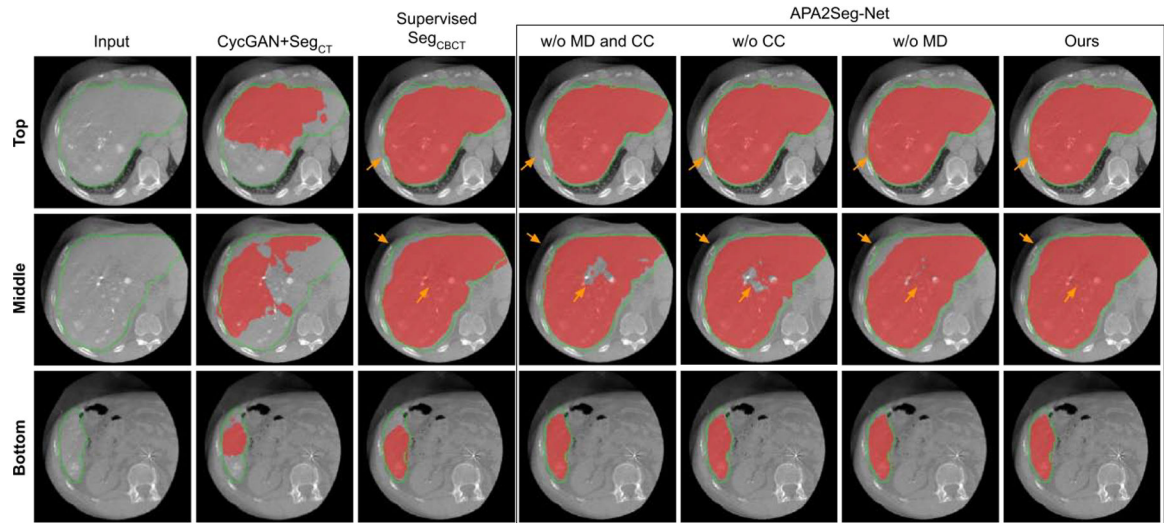


Fig. 5. Comparison of CBCT segmentation at different liver latitudes. Red mask: liver segmentation prediction. Green contour: liver segmentation ground truth. Results on APA2Seg-Net with or without CC loss and MIND loss are shown in the box.

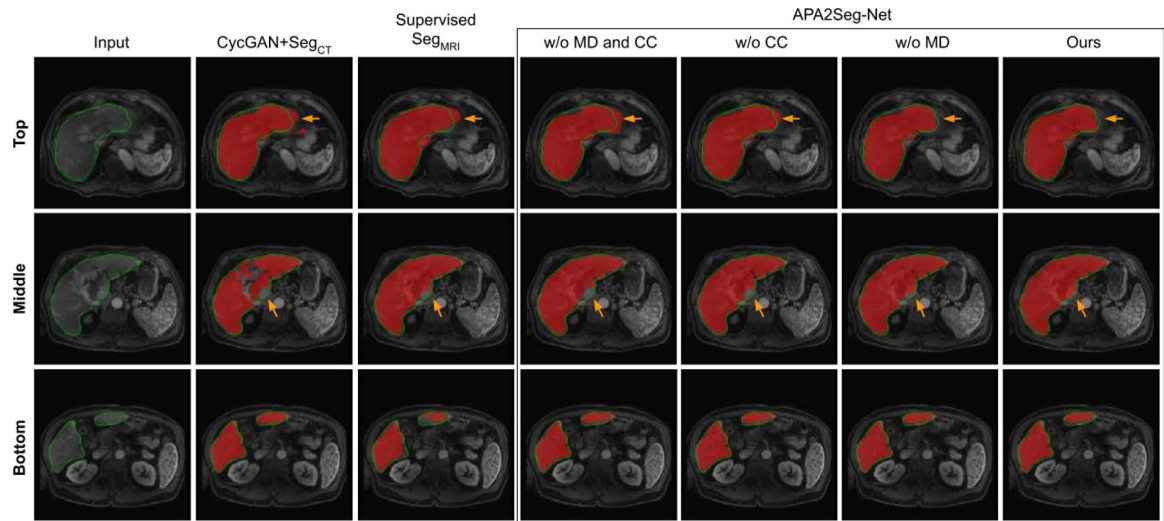


Fig. 6. Comparison of MRI segmentation at different liver latitudes. Red mask: liver segmentation prediction. Green contour: liver segmentation ground truth. Results on APA2Seg-Net with or without CC loss and MIND loss are shown in the box.

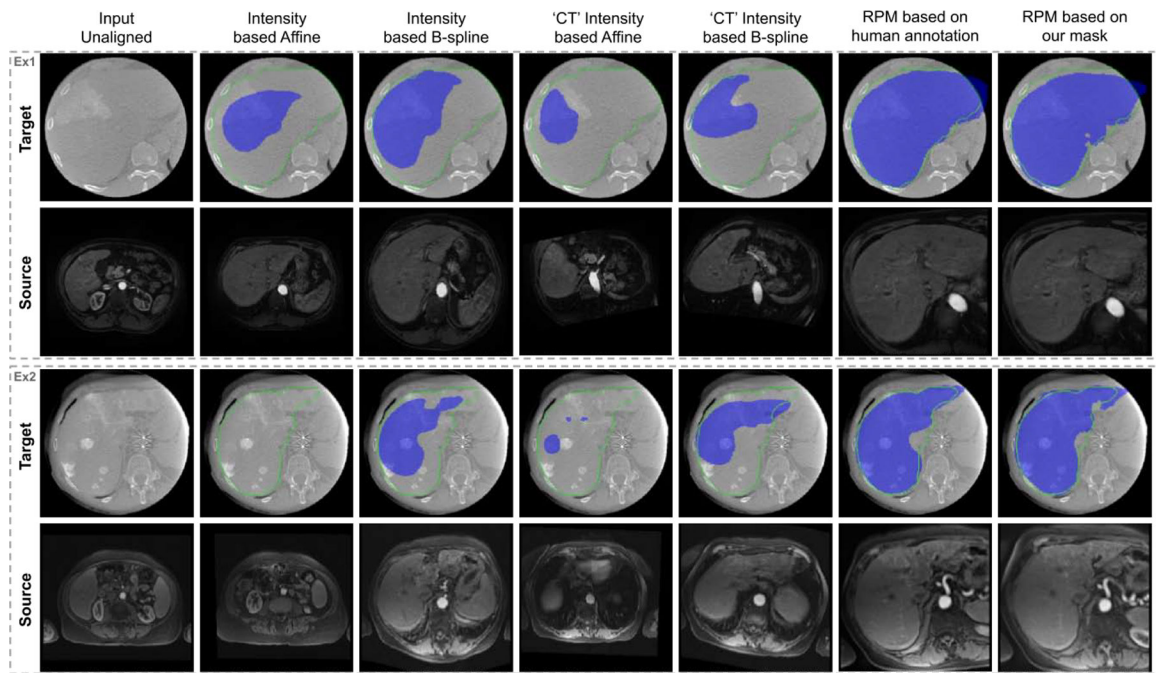


Fig. 7. CBCT (target) and MRI (source) registration results. Deformation fields are applied on ground truth MRI liver mask and overlaid on CBCT images (blue). Ground truth CBCT liver mask (green contour) is overlaid on CBCT images as well. ‘CT’ Intensity based Affine means intensity-based affine registration based on CT images translated from CBCT and MR using APA-Net. ‘CT’ Intensity based BSpline means intensity-based BSpline registration based on CT images translated from CBCT and MR using APA-Net.

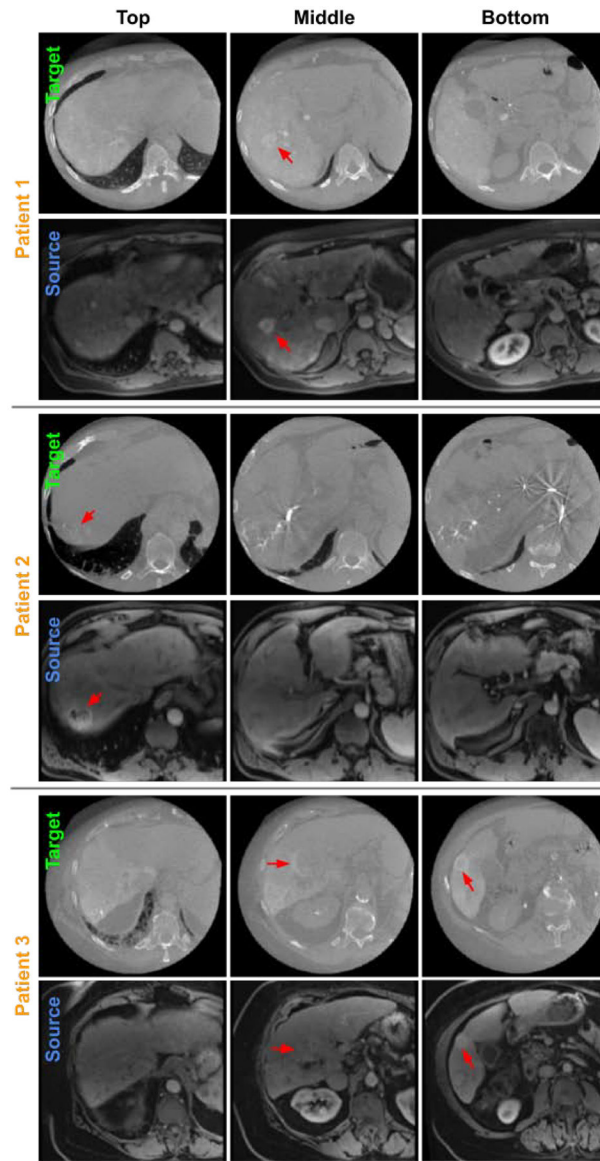


Fig. 8. Three examples of CBCT (target) and MRI (source) registration results visualized at 3 liver latitudes. RPM registration is performed based on APA2Seg-Net segmentation. Liver tumors are located by red arrows in CBCT and registered MRI.

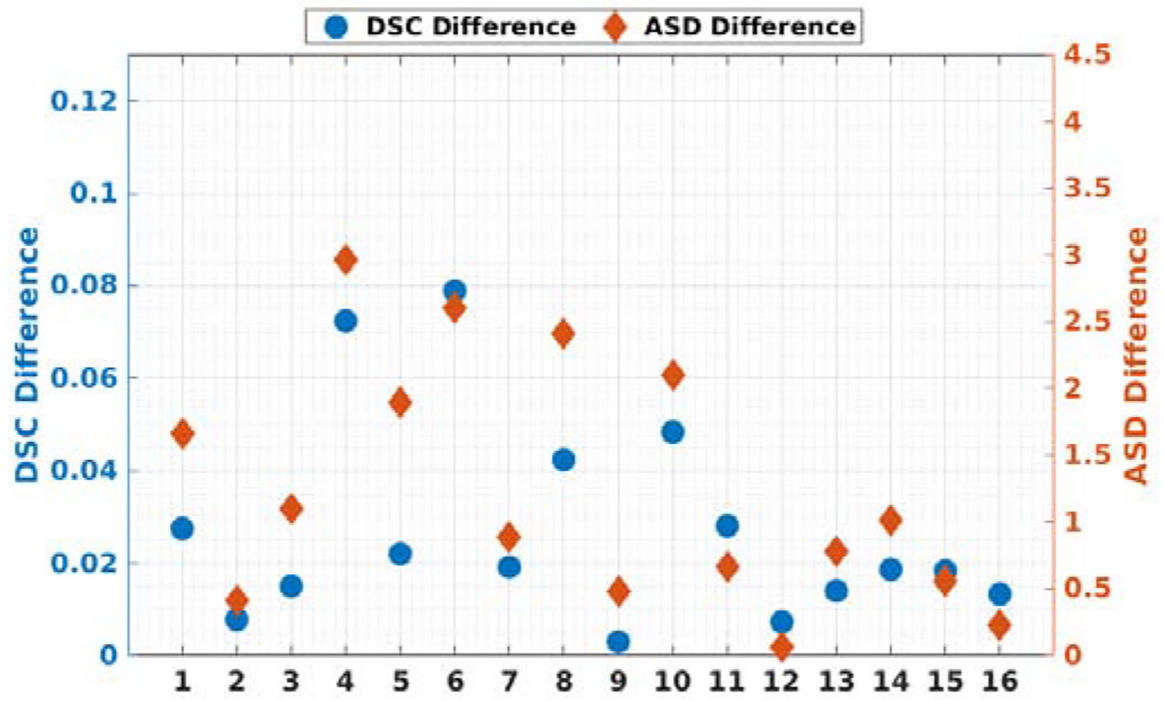


Fig. 9. Plots of all 16 patients' DSC and ASD differences between RPM registration based on human annotation and RPM registration based on our segmentation.

Quantitative comparison of CBCT and MRI segmentation results using DSC and ASD(mm). Best results are marked in red. Underline means supervised training with ground truth segmentation on the target domain, i.e. CBCT or MRI segmentation. The negative sign '-' means without the corresponding loss component in our APA2Seg-Net.

Table 1.

CBCT	CycGAN+SegCT	SegCBCT	Ours-Idt	Ours-MD-CC	Ours-CC	Ours-MD	Ours
Median DSC	0.685	0.882	0.877	0.870	0.878	0.887	<u>0.903</u>
Mean±Std DSC	0.695 ± 0.092	0.874 ± 0.035	0.873 ± 0.056	0.862 ± 0.051	0.871 ± 0.051	0.882 ± 0.038	<u>0.893 ± 0.034</u>
Median ASD	10.144	9.190	6.863	7.289	5.918	6.476	<u>5.882</u>
Mean±Std ASD	10.697 ± 2.079	10.742 ± 4.998	6.948 ± 2.138	7.459 ± 2.769	5.971 ± 1.823	6.086 ± 1.415	<u>5.886 ± 1.517</u>
MRI	CycGAN+SegCT	SegMRI	Ours-Idt	Ours-MD-CC	Ours-CC	Ours-MD	Ours
Median DSC	0.907	0.907	0.913	0.915	0.917	0.916	<u>0.918</u>
Mean±Std DSC	0.900 ± 0.044	0.859 ± 0.102	0.914 ± 0.028	0.912 ± 0.029	0.917 ± 0.026	0.916 ± 0.025	<u>0.921 ± 0.022</u>
Median ASD	1.632	2.838	1.619	1.681	1.498	1.532	<u>1.491</u>
Mean±Std ASD	2.328 ± 2.070	3.660 ± 2.522	1.932 ± 1.303	1.916 ± 1.142	1.976 ± 1.311	2.047 ± 1.382	<u>1.860 ± 1.099</u>

Table 2.

Quantitative comparison of average transformation parameter errors. Underline means supervised trained model using ground truth segmentation on the target domain.

Names	Intst-Affine	'CT'-Intst-Affine	RPM(Ours)	RPM(Seg)
s_x	3.23	3.33	0.23	0.33
s_y	0.51	0.62	0.07	0.15
s_z	0.62	0.67	0.06	0.26
h_{xy}	0.19	0.67	0.04	0.09
h_{xz}	0.32	0.65	0.04	0.08
h_{yx}	0.22	1.39	0.12	0.21
h_{yz}	0.29	0.43	0.07	0.15
h_{zx}	0.68	0.71	0.25	0.72
h_{zy}	0.28	0.33	0.09	0.39
x	79.85 mm	73.76 mm	25.13 mm	65.71 mm
y	64.48 mm	38.44 mm	15.05 mm	27.03 mm
z	96.12 mm	161.32 mm	22.74 mm	52.54 mm

Table 3.

Quantitative comparison of CBCT-MRI registration results using DSC and ASD(mm). Best and second best results are marked in red and blue, respectively. Underline means supervised trained model using ground truth segmentation on the target domain. IMG-BSpline and IMG-Affine mean intensity-based BSpline registration and affine registration, respectively. 'CT'-IMG-BSpline means intensity-based BSpline registration based on CT images translated from CBCT and MR using APA-Net.

	IMG-BSpline	IMG-Affine	'CT'-IMG-BSpline	RPM(Human Seg)	RPM(Our's)	RPM(Seg)
Median DSC	0.366	0.156	0.332	0.852	0.848	0.769
Mean±Std DSC	0.401 ± 0.168	0.168 ± 0.049	0.304 ± 0.182	0.853 ± 0.054	0.844 ± 0.029	0.755 ± 0.101
Median ASD	25.225	40.365	28.424	4.921	5.853	9.347
Mean±Std ASD	25.016 ± 8.809	40.682 ± 7.187	33.645 ± 13.511	5.095 ± 1.934	5.629 ± 0.909	8.649 ± 2.818