Generation of high-resolution target objects from multi-view, unaligned low-resolution MRIs

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

3D medical image segmentation often requires high-resolution images that are isotropic in all directions to obtain accurate, realistic segmentation results. However, high-resolution MRI images are not easy to obtain due to various limitations, such as techniques and costs. Especially for body regions other than the brain, the resolution of MRI is often very low in the out-of-plane direction, which makes 3D segmentation based on these images unsatisfactory. Previously, generating high-resolution images from low-resolution MRIs often assumed that multiple views were orthogonal to each other and perfectly registered, which is infeasible in real-world practice.

We innovatively design a two-stage training based on intertwining segmentation and unsupervised registration networks by introducing two new regularizations in the second stage to make registration and segmentation mutually reinforcing; finally, we introduce a MRI-compliant multi-view fusion method to generate high-resolution target object masks. With the validation of multiple resolutions and multiple MRI datasets, our method achieves optimal results. The design process of our method does not require the involvement of any HD images, which provides the breadth of our method and can be simply extended to any body parts where HD images are not available.

Keywords: List of keywords, comma separated.

1. Introduction

As a non-invasive, low-radiation, orientation free 3D imaging modality, MRI plays an important role in the field of clinical examination and disease diagnosis. The resolution of MR images is often very low in out-of-plane directions due to hardware limitations, acquisition time constraints, human motion, cost controls, etc (Somasundaram and Kalavathi, 2012). On the other hand, with the development of machine learning and deep learning technologies, 3D medical image segmentation can quickly segment target regions, such as tumors and organs, which can better assist in diagnosis and the development of surgical plans. However, due to the anisotropy of MRI, the low resolution in the out-of-plane direction makes the 3D segmentation acting directly on the original image very coarse, losing many important details and being far from the real morphology and structure of the target subjects, which hinders the subsequent diagnosis.

Machine learning based super-resolution (SR) methods have been widely used for reconstruction of a high resolution magnetic resonance (MR) volume from low resolution scans (Van Reeth et al., 2012; Plenge et al., 2012). Most super-resolution methods use a cost function that includes a fidelity term related to an slice acquisition model and a regularization term to constrain the space of solutions. Based on the slice acquisition model, various

© 2023 Anonymous.

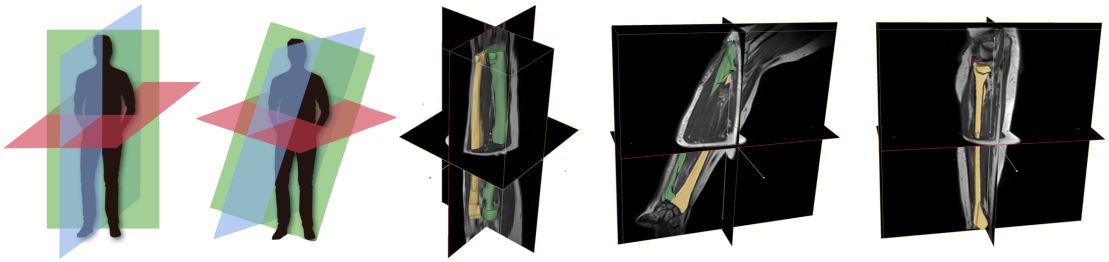

Figure 1: Illustration of input MRIs and the orientation of three different volumes. From left to right: (ideal case) anatomical coordinate system aligned with image coordinate system; (real case) anatomical coordinate system not aligned with image coordinate system; axial volume; coronal volume; sagittal volume.

methods were developed to find a feasible solution. Gholipour et al. (2010) propose a robust M-estimation solution that minimizes the error between the model-generated slices and the acquired slices. Jia et al. (2017) up-sample and fuse three orthogonal anisotropic input scans to create high-resolution isotropic 3D MR image based on sparse representation. Sui et al. (2019) introduce a multi-scale gradient field prior that improves both spatial smoothness and edge preservation. While in recent years, several studies have further investigated convolutional neural networks (CNNs) based architectures for high resolution MRI reconstruction (Pham et al., 2019; Jurek et al., 2020). Ebner et al. (2020) present a fully automatic framework for fetal brain reconstruction that includes coarse fetal brain localization, fine segmentation using CNN and outlier-robust super-resolution reconstruction. Chai et al. (2020) use a generative adversarial network (GAN) to restore the through-plane slices by modeling the restoration process as an image imputation task.

However, current MRI high resolution reconstruction methods from multiple low-resolution views often require **well-aligned** and perfectly orthogonal images (), which can be difficult to obtain in practice. Figure 1 illustrates how different views are oriented in a real case. Because of the patient's movement and the fact that different MRI volumes are not taken simultaneously, the anatomical coordinate system is not necessarily aligned with the image coordinate system. In addition, the resolution in the out-of-plane direction is usually much lower than that in the in-plane direction, which makes it impossible to extract precise, high-resolution masks from a single MRI volume. In order to recreate a high-resolution mask of a target object from multi-view MRIs, image registration and alignment are essential. Traditional medical image registration is often effective for the more subtle deformations of objects during the examination, and the targeted images are often of relatively similar resolution. The images we have to register vary in resolution in different dimensions, and the target objects have large displacements, which greatly increases the difficulty of registration.

In this paper, we introduce a novel framework called xxx that can automatically register low-resolution, unaligned medical images gotten from different orientations of MRIs and produce high-resolution masks with enhanced accuracy for representing the precise high-resolution shapes of target objects. This is the first time, to our knowledge, that an attempt

has been made to solve this problem. Our approach utilizes deep learning techniques to learn the deformations required to align and register the images, along with the segmentation of target subjects on different views of low-resolution scans. The xxx consists of three parts: the first stage learns registration and segmentation networks separately; the second stage is intertwined learning; and the third part is the fusion of segmentation results. By introducing two regularization terms, our innovative design of intermingled learning can well facilitate each part's learning: the segmentation results are used to boost the registration of the target objects, and more information gotten from the other views is provided to the current view's out-of-plane segmentation. This complementary learning approach can achieve a more optimal solution for both registration and target object segmentation. In addition, our newly designed single-dimensional uncertainty Gaussian fusion can fuse information from multiple views simply and efficiently. It takes inspiration from GMM (), which not only takes care of the characteristic of MRIs' out-of-plane uncertainty but is also faster than other commonly used image interpolation integration methods and has been especially validated in our experiments for high resolution generation of MRI images.

We demonstrate the effectiveness of our framework on a variety of datasets with varying distances between slices. Our experiments show that xxx is able to outperform previous methods, particularly in cases where the images are low-resolution or poorly aligned. In addition, different from previously generative based methods () and image-to-image translation methods (), our method does not require any high-resolution images to be involved in the training and design process, which also provides the extensibility and effectiveness of our method on MRI of body parts where other high-resolution images are almost inaccessible.

2. Methods

We introduce our framework for xxx in this section. Our xxx is made to effectively auto-register low-resolution, unaligned images, segment the targets, and generate high-resolution masks that are more accurate representations of the target objects' shapes. In the following section, we will explain how they are achieved by our 2-phase training: a coarse pre-train for image registration and image segmentation in **??**; an intertwined training for fine-tuning in **??** and finally, how the high-resolution masks are generated by our novel 1D-Gaussian fusion in **??**.

2.1. Pre-processing

Suppose we have the input of three low-resolution anisotropic MR scans from three different views (axial, coronal, and sagittal) as $R_{i,1} \in \mathbf{R}^{X \times Y \times Z_l}$, $R_{i,2} \in \mathbf{R}^{X \times Y_l \times Z}$,$R_{i,3} \in \mathbf{R}^{X_l \times Y \times Z}$ from the same patient $I = i$ at the same body part, where X_l, Y_l, Z_l represent the sizes of the low-resolution dimensions, which are much smaller than the high resolution sampled dimensions X, Y, and Z. The ground-truth masks for these three images are $Y_1 \in \mathbf{R}^{X \times Y \times Z_l}$, $Y_2 \in \mathbf{R}^{X \times Y_l \times Z}$,$Y_3 \in \mathbf{R}^{X_l \times Y \times Z}$. For MRI, which is often not isotropic, its voxel size can be expressed in terms of pixel-spacing (p) and distance between slices (d) $p \times p \times d$ (take R1 as an example), d equals to $\frac{Z_l}{Z}$. Thus, in order to better register R1, R2, and R3, we first need to match their voxels to the same unit, i.e., by up-sampling (nearest) all three images and masks into a voxel size of $p \times p \times p$, and we get $R'_1, R'_2, R'_3, Y'_1, Y'_2, Y'_3 \in \mathbf{R}^{X \times Y \times Z}$.

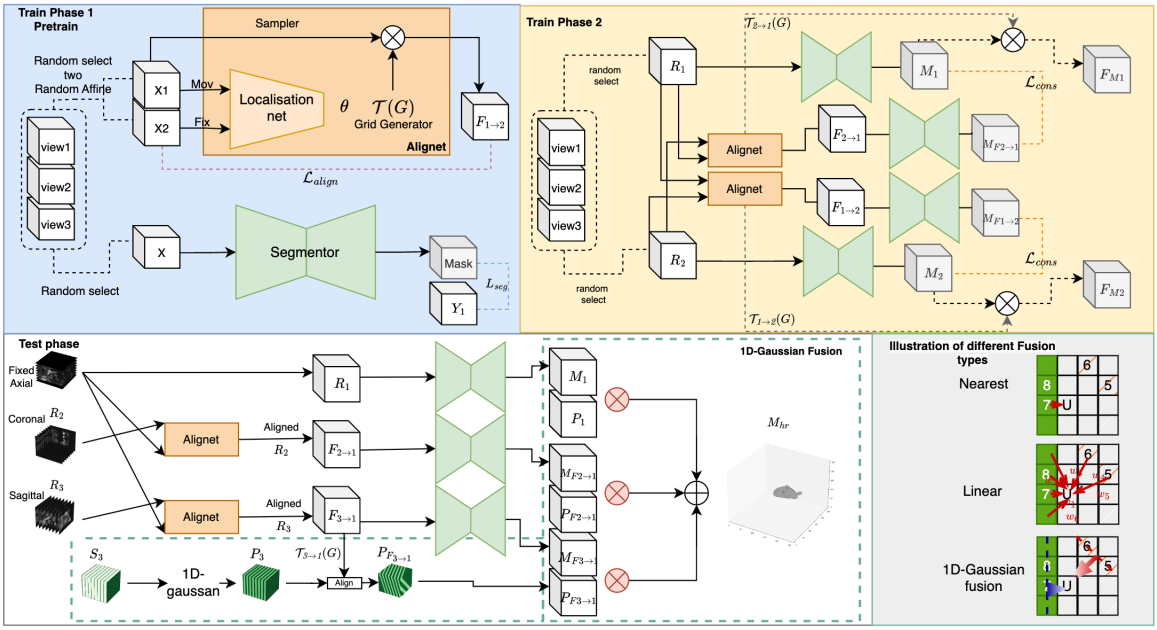

Figure 2: This is the pipeline of our proposed method xxx. In training phase 1, an image registration network (Alignet) and a segmentation network are trained separately. In the training phase2, the Alignet and segmentor are interwine to fine-tune with two newly introduced regularization loss. In the text phase, Cornoal and Sagittal view images are aligned into the Axial view location first, and a 1-D Gaussian fusion are applied to fuse the final predicted masks into a high-resolution mask.

2.2. Coarse segmentation and registration

In *phase 1* training, we train the coarse segmentation and registration separately, seen as figure **??** (upper-left). For segmentation, we assembled all of the available up-sampled images and masks into a training set $\mathbf{X} = \mathbf{R'_1} + \mathbf{R'_2} + \mathbf{R'_3}$, $\mathbf{Y} = \mathbf{Y'_1} + \mathbf{Y'_2} + \mathbf{Y'_3}$ and feed them into a 3D U-net to () train a coarse segmentation network $\mathcal{S}_1$. Let x as the input volume, $m = \mathcal{S}_1$ as the predicted mask and y as the corresponding ground truth mask. Dice loss () was applied for penalizing the segmentation objective for each input pair (x, y),

$$L_{seg_1} = -2 \cdot \frac{|y \cap m|}{|y| + |m|}. \tag{1}$$

For registration, we implemented a coarse 3D spatial transformer network () $\mathcal{G}_1$. For each patient i, we sampled two views at one time, $x_1, x_2 \in R'_{i,1}, R'_{i,2}, R'_{i,3}$ and random affined each of them to build more spatial transformations. The localization network takes two 3D volumes, x_1 as the moving image and x_2 as the fixed image, and generates 12 affine parameters that reflect the transformation from x_1 to x_2: three rotation angles, three translation parameters, three scaling parameters, and three shearing parameters. The details of this architecture are shown in Appendix **??**. Considering that MRI responds to real-life objects and does not have the deformation and scaling of objects, we reduced the number of affine parameters by setting the shearing and scaling parameters all to 1. After sampling grid G based on these affine parameters $\theta_{1\to2}$ to get registration field $\phi_{1\to2} = \mathcal{T}_\theta(G)$, we applied a differentiable image sampling on the input moving image x_2 to generate a moving image $f_{1\to2} = \mathcal{G}(x_1, x_2) = x_1 \circ \phi_{1\to2}$ that registered into the pose of x_2.

The unsupervised registration loss combines two components: $\mathcal{L}_{sim}$ that penalizes the difference between the moved image and the fixed image, and $\mathcal{L}_{ID}$ that regularizes that two identical inputs would not be transformed,

$$\mathcal{L}_{align_1} = \mathcal{L}_{sim}(x_1, f_{1\to2}) + \lambda \mathcal{L}_{ID}, \tag{2}$$

where we applied L_{sim} as local cross-correlation () of x_1 and $f_{1\to2}$, which is more robust to intensity variations across scans and datasets, and $L_{ID} = ||x_1, \mathcal{G}(x_1, x_1)||_2 + ||x_2, \mathcal{G}(x_2, x_2)||_2$, where $|| \cdot ||$ denotes L2-norm. We could also extend the unsupervised registration loss into a segmentation-driven version since the ground-truth masks y_1 and y_2 are available for the inputs, which were demonstrated to achieve a better within-subject image registration (), the registration loss now is:

$$\mathcal{L}_{align_1*} = \mathcal{L}_{sim}(x_1, f_{1\to2}) + \lambda_1 \mathcal{L}_{ID} + \lambda_2 \mathcal{L}_{sim}(y_1, y_1 \circ \phi_{1\to2}). \tag{3}$$

In section xx, we analyzed and compared these two versions of registration loss, and both of them are effective for our algorithm.

2.3. Intertwined-tuning segmentation and registration

After getting the pre-trained segmentation S_1 and G_2, we fine-tuned them by intertwining the segmentation and registration steps. Similarly, in this stage, we randomly select two

different views for each patient i at one time, $x_1, x_2 \in R'_{i,1}, R'_{i,2}, R'_{i,3}$. Besides the previous described L_{seg} and L_{align_m}, we introduced a cross-view supervision L_{cons_1} and L_{cons_2} as

$$\begin{aligned} \mathcal{L}_{cons_1} &= ||M_{f2\to1} - M_1||_2 + ||M_{f1\to2} - M_2||_2, \\ \mathcal{L}_{cons_2} &= ||F_{M1} - M_1||_2 + ||F_{M1} - M_1||_2, \end{aligned} \tag{4}$$

where $M_{f2\to1} = \mathcal{S}(f_{2\to1})$ represents the segmentation masks gotten from the aligned fake image $f_{2\to1}$, and $M_{f1\to2} = \mathcal{S}(f_{1\to2})$ denotes the segmentation masks gotten from the aligned fake image $f_{1\to2}$. $F_M1 = S(x1) \circ \phi_1 \to 2$ was obtained by first segmenting the original image x_1 and then registering the segmented mask into the pose of image x_2, whereas $F_M2 = S(x_2) \circ \phi_2 \to 1$ is obtained by first segmenting the original image x_1 and then registering the segmented mask into the pose of image x_1. $\mathcal{L}_{cons_1}$ encourages the segmentation network $\mathcal{S}$ to predict masks from a single view of the image that look similar to masks from aligned images, and $\mathcal{L}_{cons_2}$ encourages the registration network $\mathcal{G}$ to align the segmented masks to have more overlap. These two intertwined regularizations were applied separately to $\mathcal{S}$ and $\mathcal{G}$ to achieve more accurate and precise segmentation and registration.

In training stage 2, the alignment and segmentation networks are updated iteratively, and the final objectives to update the segmentation network and registration network are $\mathcal{L}_{seg_2} = L_{seg_1} + \alpha_1 L_{cons_1}$, $\mathcal{L}_{align_2} = L_{align_1} + \alpha_2 L_{cons_2}$, respectively. The details of the algorithm are shown in Appendix **??** todo.

2.4. 1-D Gaussian fusion to generate high-resolution mask

During inference, we take the three views of scans R'_1, R'_2 and R'_3, and align the last two views (R'_2 R'_3 into the axial scans (R_1) first; then feed these three aligned scans into the segmentation networks to get three predicted masks.

Meanwhile, take the sagittal view as an example. We introduced a sampling matrix called S_3, that reflects the sampling slices and the voxel accuracy in the up-sampled image matrix, where $S_3[i, j, k] = 1$ when this voxel point is actually taken from the original low-resolution image R_1 instead of a interpolated value. Thus we can see that the matrix S3 is like a matrix with the xy-plane assigned to 1 at a certain distance apart in the x-dimension, as shown in Figure 2. Then we applied a 1D-gaussian along the out-of-plane dimension (x dimension for sagittal views) with a Gaussian kernel size of $d_s/2$, we could get a probability matrix (P_3) that reflects the confidence of this scan at each voxel, and this probability matrix can be aligned into the registered images as $P_{3\to1} = S_3 \circ \phi_{3\to1}$. The final mask is fused by the following equation:

$$M_{hr} = \frac{1}{\omega_1} M_1 * P_1 + \frac{1}{\omega_2} M_{F2\to1} * P_{F2\to1} + \frac{1}{\omega_3} M_{F3\to1} * P_{F3\to1}; \tag{5}$$

where w_1, w_2, w_3 are the normalization factors with $\omega_i = \frac{P_i}{\sum_1^3 P_k}$.

3. Experiments

3.1. datasets

According to the same as we described before, most of the body parts usually do not operate high-resolution (HD) MRI, so to be able to better prove the effectiveness of our method, we

test our method by the body parts that HD MRI can provide. By building the algorithm on the low-resolution images, we can evaluate our method's on the high-definition images for the same subject. In this work, we obtained two MRI datasets, *Brats* () and *Heart* ().

Brats dataset xxx todo. For the 273 patient studies, we split them as 0.8:0.2, where there are 164 studies in the training set and 42 studies in the testing set. Considering the original volumes are isotropic with a voxel spacing of 1mm*1mm*1mm, we started by shifting all objects to the center and filling the 3D object into a cube of 256*256*256 voxels. To get the low-resolution images, we first applied a random 3D affine of a rotation ranging from -30 degrees to 30 degrees and a translation ranging from -30 voxels to 30 voxels in each direction to all of the volumes. Then, we resampled each dimension with a slice distance as 8mm (noted as dataset *Brain_8*) and 16mm (noted as dataset *Brain_16*), where the resampled images were sized in $R_{i,1} \in \mathbf{R}^{256\times256\times32}$ (Axial view), $R_{i,2} \in \mathbf{R}^{256\times32\times256}$ (Coronal view),$R_{i,3} \in \mathbf{R}^{32\times256\times256}$ (Sagittal view) for voxel spacing as 8mm and $R_{i,1} \in \mathbf{R}^{256\times256\times16}$, $R_{i,2} \in \mathbf{R}^{256\times16\times256}$,$R_{i,3} \in \mathbf{R}^{16\times256\times256}$ for voxel spacing as 16mm. It is worth noting that for $\mathbf{R}_{i,1}$, $\mathbf{R}_{i,2}$, and $\mathbf{R}_{i,3}$, they were subjected to different 3D affine operations before resampling operation, i.e., the images sampled from these three views are completely unaligned.

Heart dataset todo. For the 20 patient studies, we split them into 15 studies for the training set and 5 studies for the test set. The original images have a voxel spacing of 1.25mm*1.25mm*1.37mm, which has a slightly lower but acceptable resolution in the third dimension. To make a consistent 3D input size for networks, we resampled and padded the volumes into 256*256*256 voxels with a 1mm*1mm*1mm voxel spacing first. Then, as with the previous *Brats* dataset, we resampled in each direction with voxel spacings of 8 mm and 16 mm to create the *Heart_8* and *Heart_16* datasets.

3.2. Baseline methods

We compared our methods with the following methods: (1) Unet-LR, the first baseline is the segmentation algorithms (3D Unet) with a combination of all the low-resolution images with all views available. (2), Unet-cc-img-nst, is the Unet-LR plus a traditional, non-learning-based iterative registration method from the SimpleITK package [] that uses image correlation as a similarity metric and has a convergence minimum value of 1E-6 to register three-view images. Then the segmented masks are aligned by the affine transformation gotten from image registration and fused by a 3D nearest neighbor interpolation (Unet-cc-img-nst). (3) Unet-cc-msk, is a similar version of (2) but applies registration directly to the segmented masks. (4) Unet-cc-img-vote, a similar version of (2) but applying majority voting method () commonly used in ensemble learning as the final fusion algorithm. (5) Unet-HD, which is trained on 1 mm* 1 mm* 1 mm voxel-spaced HD images. It is not fair to consider it one of the methods to compare, but it could serve as an "upper bound" for all methods.

3.3. Metrics

We compare the predicted masks with the ground-truth high-resolution masks Y_{HD}. To make the comparison fair, we assume that the high-resolution masks are perfectly aligned with the axial view images (the Affine operation was not used to make the axial view testing

images and masks). By directly calculating the metrics on the predicted axial view masks, we could compare the predicted low resolution masks M_1 with HD ground truth masks Y_{HD} for baselines when only one view is available during inference. Baselines (2)-(4) fuse registered low-resolution masks by different strategies, and when three views are available, coronal and sagittal view masks are aligned to the axial view by non-learning methods or neural networks. We quantify the generated high-resolution masks with Y_{HD} by (1) the dice score (DSC) which quantifies the overlap of the predicted masks and HD ground truths. (2) the US (),OS (), RMS () which quantify the precision of the predicted masks. (3) In light of the fact that the dice-score evaluates overlap but places little emphasis on the shape information of segmented volumes, we also developed a boundary accuracy () that focuses on the shape details.

3.4. Implementation details

Pytorch was used to implement the method (see citation). For our methods, during training stage 1, Segmentor and Alignet are trained separately with a learning rate of 0.001 for 40 epochs; and during training stage 2, the learning rate is set at 0.0001 with a step-wise learning rate decay for every 200 iterations as gamma = 0.9, and the stage 2 training is set for 60 epochs. For all the training, the batch size is set to 2, and 2 GPUs with the Geforce A6000 are used. For the hyper-paramters, $\alpha 1 = \alpha 2 = 0.05$, $\lambda_1 = 0.1$ and $\lambda_2 = 0.05$. For the other baseline methods, all the methods are set for training over 100 epochs with the same learning rate and batch size as our method.

Table 1: Quantitative results of the high-resolution mask generation methods.

Methods			Heart 8					Brain 8		
	DSC	US	OS	RMS	mBA	DSC	US	OS	RMS	mBA
Unet-LR	0.863	0.118	0.15	0.141	0.85	0.84	0.127	0.173	0.166	0.819
Unet-img-cc-nst	0.819	0.202	0.156	0.182	0.803	0.634	0.194	0.397	0.195	0.789
Unet-img-cc-vote	0.879	0.063	0.174	0.13	0.859	0.813	0.058	0.262	0.196	0.807
Ours-1seq	0.871	0.098	0.117	0.108	0.855	0.849	0.142	0.151	0.157	0.821
Ours	**0.896**	**0.093**	**0.113**	**0.105**	**0.87**	0.883	0.109	0.116	0.126	0.848
			Heart 16					Brain 16		
	DSC	US	OS	RMS	mBA	DSC	US	OS	RMS	mBA
Unet-LR	0.812	0.2	0.173	0.189	0.78	0.81	0.134	0.224	0.195	0.797
Unet-img-cc-nst	0.664	0.338	0.33	0.336	0.687					
Unet-img-cc-vote	0.683	0.319	0.302	0.310	0.703	0.638	0.504	0.042	0.359	0.715
our-1seq	0.819	0.2	0.157	0.182	0.791	0.82	0.156	0.19	0.184	0.801
Ours	0.875	0.139	0.131	0.137	0.842	0.86	0.108	0.165	0.152	0.823
dataset type			HD					HD		
Unet-HD	0.923	0.102	0.051	0.082	0.908	0.9	0.068	0.117	0.11	0.877

3.5. Results and discussion

Table 1 shows the results for the high-resolution mask generation on various datasets. It is shown that our xxx method outperforms all the baselines on all the datasets when only low-resolution images are available. Compared with the drop in segmentation performance

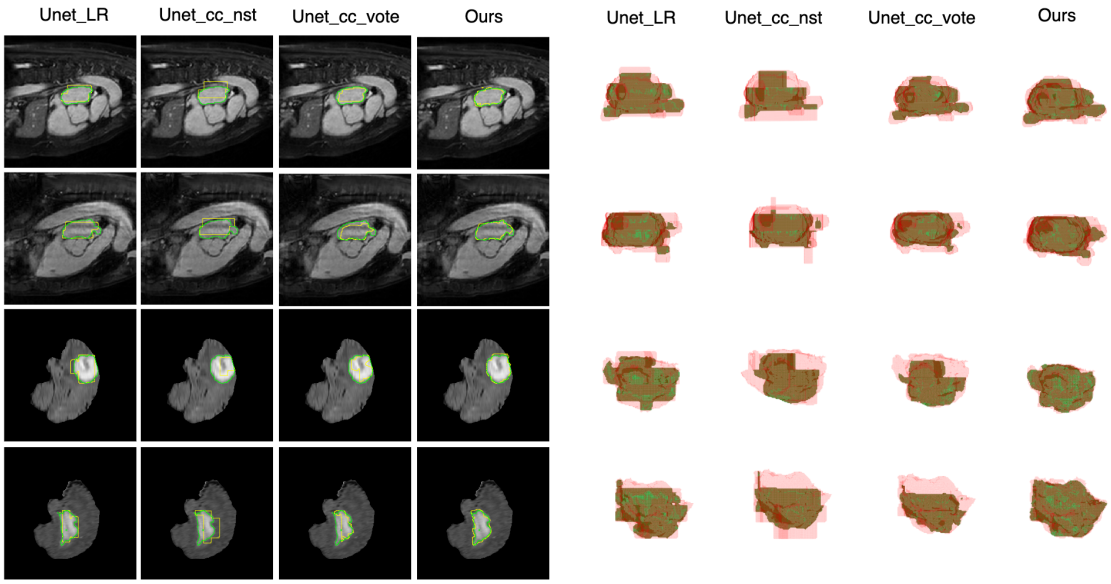

Figure 3: Qualitative results of the high-resolution mask generation methods. The methods depicted here are Unet-LR, Unet-cc-nst, Unet-cc-vote, and Ours, in that order. The left columns are the 2D-slice view of the HD images with masks; the green curves are the ground truths, and the yellow curves are the predictions. The right columns are the 3D-view of the objects, where green voxels are the overlapping areas where $M \wedge Y_{HD}$, and red areas are the disagree areas where $MxorY_{HD}$.

Table 2: Ablation study of each components of our model.

Ablation methods	componenets			heart-16				
	two-stage	Alignet	fusion	DSC	US	OS	RMS	mBA
Our-1-stage	0	1	our fusion					
Our - voting	1	1	voting	0.826	0.154	0.161	0.157	0.821
Our - nearest	1	1	nearest	0.787	0.047	0.329	0.235	0.783
Our - cc	1	cc align	our fusion					

from Unet-HD to Unet-LR, there is a drop in DSC of 0.06 and 0.04 for slice distances of 8 mm, and a drop of 0.112 and 0.09 for slice distances of 16 mm. We can see that only training with low resolution is not enough to achieve acceptable high-resolution masks, and the loss of information between out-of-plane slices is detrimental. And the more distant the slices are from each other, the lower the resolution of the image is, and the more significant this performance reduction is. This could also be seen from Figure 3, where, despite applying up-sampling to the real-world coordinate system, the segmented masks from Unet-LR have a very distinct ladder-like appearance.

Even when only getting low-resolution images, our methods get an increase in DSC of 0.03 and 0.043 for slice thickness of 8 mm and a significant increase of 0.052 and 0.09 for slice thickness of 16 mm, for the heart and brain datasets, compared with Unet-LR. This confirms the ability and effectiveness of our method to combine more details by extracting information from low-resolution images. Comparing this parameter to OS, it can be reduced by about 40% on the dataset and even 100% on the brain dataset with a slice distance of 16. Our approach can substantially reduce the over-segmentation due to the out-of-plane uncertainty. Using mBA as a metric to evaluate the performance of the segmentation boundary details, we could see an obvious increase in mBA for our methods, especially with a slice thickness of 16 mm, which can compensate for a large part of the mBA lost due to low resolution.

Seen from the drop of performance for baselines Unet-img-cc-nst and Unet-img-cc-voting, it shows the dice of Unet-img-cc-nst as 0.664 for the Heart 16 dataset and the dice of Unet-img-cc-voting as 0.763. This implies that generating high-resolution masks from multiple views requires highly precise image registration. The traditional image registration methods based on image correlation can not handle the very low out-of-plane resolution registration well, and an unsuccessful image registration not only does not help to improve the resolution of the segmented targets, but may also have the exact opposite effect. The learning of our registration network based on segmentation and target object's mask can circumvent this problem and make the connection between views to contribute to the fusion of target object's segmentation, more details of the effectiveness of each component are shown in appendix xxx Table 2.

Also seen from the illustration in Figure 3, our final generated masks, though free from any high resolution images or masks during the training and evaluation, can preserve the details of the object boundaries and have a high consensus with the high-resolution masks (the disagreement is shown in red voxels in Figure 3 (right)).

4. Conclusion

In this work, we proposed a method xxx that could generate high-resolution target segmentation from multiple unaligned low-resolution images. The experiments demonstrate the effectiveness of our methods for improving segmentation performance when only low-resolution masks are available. Our work is free of the involvement of HD images during the training stage, which implies its generalization ability to extend to other body parts; it could also be extended to high-resolution image generation by some small modifications.

Acknowledgments

Acknowledgments withheld.

Algorithm 1: Computing Net Activation

Input: $x_1, \ldots, x_n, w_1, \ldots, w_n$
Output: y, the net activation
$y \leftarrow 0$;
for $i \leftarrow 1$ **to** n **do**
  $y \leftarrow y + w_i * x_i$;
end

References

Yaqiong Chai, Botian Xu, Kangning Zhang, Natasha Lepore, and John C Wood. Mri restoration using edge-guided adversarial learning. *IEEE Access*, 8:83858–83870, 2020.

Michael Ebner, Guotai Wang, Wenqi Li, Michael Aertsen, Premal A Patel, Rosalind Aughwane, Andrew Melbourne, Tom Doel, Steven Dymarkowski, Paolo De Coppi, et al. An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain mri. *NeuroImage*, 206:116324, 2020.

Ali Gholipour, Judy A. Estroff, and Simon K. Warfield. Robust super-resolution volume reconstruction from slice acquisitions: Application to fetal brain mri. *IEEE Transactions on Medical Imaging*, 29(10):1739–1758, 2010. doi: 10.1109/TMI.2010.2051680.

Yuanyuan Jia, Ali Gholipour, Zhongshi He, and Simon K. Warfield. A new sparse representation framework for reconstruction of an isotropic high spatial resolution mr volume from orthogonal anisotropic resolution scans. *IEEE Transactions on Medical Imaging*, 36 (5):1182–1193, 2017. doi: 10.1109/TMI.2017.2656907.

Jakub Jurek, Marek Kociński, Andrzej Materka, Marcin Elgalal, and Agata Majos. Cnn-based superresolution reconstruction of 3d mr images using thick-slice scans. *Biocybernetics and Biomedical Engineering*, 40(1):111–125, 2020.

Chi-Hieu Pham, Carlos Tor-Díez, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Passat, and François Rousseau. Multiscale brain mri super-resolution using deep 3d convolutional networks. *Computerized Medical Imaging and Graphics*, 77:101647, 2019.

Esben Plenge, Dirk HJ Poot, Monique Bernsen, Gyula Kotek, Gavin Houston, Piotr Wielopolski, Louise van der Weerd, Wiro J Niessen, and Erik Meijering. Super-resolution methods in mri: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magnetic resonance in medicine*, 68(6):1983–1993, 2012.

K Somasundaram and P Kalavathi. Analysis of imaging artifacts in mr brain images. *Oriental J Comput Sci Technol*, 5(1):135–141, 2012.

Yao Sui, Onur Afacan, Ali Gholipour, and Simon K Warfield. Isotropic mri super-resolution reconstruction with multi-scale gradient field prior. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–11. Springer, 2019.

Eric Van Reeth, Ivan WK Tham, Cher Heng Tan, and Chueh Loo Poh. Super-resolution in magnetic resonance imaging: a review. *Concepts in Magnetic Resonance Part A*, 40(6): 306–325, 2012.