Computers in Biology and Medicine 147 (2022) 105667

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Slice imputation: Multiple intermediate slices interpolation for anisotropic 3D medical image segmentation

Zhaotao Wu [a], Jia Wei [a,*], Jiabing Wang [a], Rui Li [b]

[a] School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
[b] Golisano College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, NY, 14623, USA

ARTICLE INFO

Keywords:
Slice imputation
Frame interpolation
Medical image segmentation
Multi-task learning
Deep learning

ABSTRACT

We introduce a novel frame-interpolation-based method for slice imputation to improve segmentation accuracy for anisotropic 3D medical images, in which the number of slices and their corresponding segmentation labels can be increased between two consecutive slices in anisotropic 3D medical volumes. Unlike previous inter-slice imputation methods, which only focus on the smoothness in the axial direction, this study aims to improve the smoothness of the interpolated 3D medical volumes in all three directions: axial, sagittal, and coronal. The proposed multitask inter-slice imputation method, in particular, incorporates a smoothness loss function to evaluate the smoothness of the interpolated 3D medical volumes in the through-plane direction (sagittal and coronal). It not only improves the resolution of the interpolated 3D medical volumes in the through-plane direction but also transforms them into isotropic representations, which leads to better segmentation performances. Experiments on whole tumor segmentation in the brain, liver tumor segmentation, and prostate segmentation indicate that our method outperforms the competing slice imputation methods on both computed tomography (1\% Dice improvement for CT liver tumor segmentation) and magnetic resonance images volumes (over 2\% Dice improvement for MRI prostate segmentation) in most cases.

1. Introduction

In the field of medical image processing and analysis, medical image segmentation is a difficult but critical task. It is a crucial step in image-guided surgery, computer-aided detection, and medical data visualization [1–3]. Its goal is to accurately segment medical images with semantic labels so that it can provide reliable meaningful information for clinical diagnosis and pathology research. Moreover, it aims to assist physicians in making correct diagnoses. Deep learning-based methods for medical image segmentation have been proposed in recent years and have demonstrated state-of-the-art performance [4–6].

For 3D medical image segmentation, deep-learning-based methods prefer to learn features from isotropic volume data as they can provide more anatomical details and metabolism information [7]. However, due to hardware limitations and time costs, isotropic volumes are difficult to obtain in clinical practice [8]. Anisotropic volumes are commonly available in most cases. In the through-plane directions, an anisotropic 3D volume is elongated, resulting unequal resolutions in the three dimensions [9]. For example, it may have high resolution (HR) in the in-plane (axial) but low resolution (LR) in the through-plane (sagittal and coronal) directions. Consequently, detailed structures in the through-plane direction are unclear, which leads to a negative impact on image analysis, visualization, and diagnosis of lesions [7].

In the field of medical image segmentation, segmentation on anisotropic 3D medical volumes remains a difficult task. The main problem is that anisotropic 3D medical volumes are too sparse to adequately fit functional representations or provide sufficient fine-scale information to recover the missing details [10]. Image clarity is reduced and anatomical structures are significantly distorted between consecutive slices when we directly increase apparent volume resolution, for example, using linear interpolation (Linear) [11], as presented in Fig. 1(B).

We address the challenge described above by proposing a method to synthesize intermediate slices between consecutive slices. 3D medical volumes are continuous slice sequences in the dimension of space, similar to videos, which are continuous image sequences in the dimension of time. As a result, we propose a multiple intermediate slices interpolation method, called slice imputation (SI), to generate isotropic 3D volumes and make them suitable for segmentation, inspired by the idea of frame interpolation [12]. The main contributions of this work are as follows:

* Corresponding author.
E-mail address: csjwei@scut.edu.cn (J. Wei).

https://doi.org/10.1016/j.compbiomed.2022.105667
Received 29 March 2022; Received in revised form 7 May 2022; Accepted 22 May 2022
Available online 31 May 2022

- To solve the anisotropy problem of 3D medical volumes for medical image segmentation, we use a frame interpolation method. The inter-slice distance of the interpolated volumes can become close to the x-y spacing distance by increasing the number of slices between every two consecutive slices, transforming them into isotropic volumes.
- To improve slice smoothness in the through-plane direction, we introduce a smoothness loss function to evaluate the smoothness of 3D medical volumes in the through-plane direction.
- To improve the authenticity of the interpolated slices, we employ a multitask learning mechanism. The model can use the interacting information between the classification and distinguishing tasks to generate more realistic slices by learning the object classifier and the local discriminator together.

This research is a significant extension of our previous work [13]. The smoothness in the in-plane direction is the only focus of the inter-slice image augmentation (IIA) method proposed in Ref. [13]. The interpolated 3D volumes of IIA tend to have blurry edges in the through-plane direction. Therefore, we utilize a smoothness loss function which can evaluate the smoothness of 3D medical volumes in the sagittal and coronal directions. Furthermore, we adopt the multi-task learning mechanism to learn the classification task and the distinguishing task, which helps the model make use of the correlation information between the two tasks.

The remainder of this paper is laid out as follows: In Section 2, we go over some related works. The SI method and its derivation are discussed in Section 3. We describe the implementation details in Section 4, as well as experimental results and an analysis of the algorithm's behavior. Section 5 performs a detailed ablation analysis of the network to validate the effect of the local and global discriminators and determine the optimal numbers of intermediate slices. Finally we present a conclusion in Section 6.

2. Related work

2.1. Medical image segmentation for anisotropic volumes

In the field of medical image segmentation, deep-learning-based medical image segmentation methods have recently achieved state-of-the-art performance [11,14–17]. However, due to the anisotropy of the 3D volumes, deep-learning-based methods still fall short in 3D medical image segmentation [18]. Some recent studies focused on addressing this anisotropy problem [19,20].

Delannoy et al. [21] proposed an end-to-end methodology dedicated to the analysis of low-resolution and anisotropic MR images. They used a GAN-based approach to estimate jointly an high-resolution (HR) image and its corresponding segmentation map from an low-resolution (LR) image. Although the proposed method performs well in segmentation task, it requires paired LR/HR images to learn the mapping from LR images to HR images. However, in practice, such training data are often unavailable. Lee et al. [19] proposed to avoid down-sampling feature maps along the z-dimension and used convolution kernels with a particular size, which transforms the volumes into almost isotropic ones. Based on 2D and 3D vanilla U-Nets, Isensee et al. [20] proposed a robust and self-adapting framework. This method makes the volumes isotropic by first down-sampling the HR axes of the anisotropic medical volumes until they match the LR axes and then up-sampling the volumes to the original voxel spacing. However, directly down-sampling the higher-resolution axes of the volumes may lead to information loss, which may lead to a negative impact on high quality segmentation. On the contrary, our method, namely, SI, preserves information in the original medical volumes by increasing the slices between every two consecutive slices while transforming the volumes into isotropic ones.

2.2. Super-resolution algorithm

Super-resolution (SR) methods aim to learn complex mapping relations between LR and HR images. Kim et al. [22] proposed a deep-convolutional-network based SR method. By increasing the network depth, this method significantly improves the accuracy of restoration. Although deep SR methods achieve accurate restorations of high frequency contents, effectively training a very deep SR CNN is challenging due to the vanishing gradient problem [7]. Du et al. [7] proposed an SR reconstruction method based on residual learning with long and short skip connections. Deep networks' vanishing gradient problem can be effectively addressed with the proposed method, which restores high-frequency details of magnetic resonance images (MRI). Most of the existing SR algorithms, according to Lim et al. [23], treat super-resolution of different scale factors as independent problems without considering mutual relationships among different scales in SR. As a result, they proposed the EDSR, an enhanced deep SR network, that transfers knowledge from a model trained at other scales. Zhang et al. [24] proposed a residual channel attention network (RCAN) to obtain very deep trainable networks and adaptively learn informative channel-wise features. Wang et al. [25] proposed a patch-free 3D medical image segmentation method, which can realize HR segmentation with LR input. The motivation of the proposed method is capturing global context while not introducing too much extra computational cost. They use super resolution method as an auxiliary task for the segmentation task to restore the HR details lost in the down-sampling procedure. To shorten the time of image reconstruction and optimize the structure for speed, Zhang et al. [26] proposed a fast medical image super-resolution (FMISR) method. FMISR is a combined sub-pixel convolutional layer and mini-network to shorten the time of super-resolution. Furthermore, FMISR implemented hidden layers to remain the information while training the images for improving the quality of the reconstruction. Iglesias et al. [27] proposed a method which can utilize LF-MRI T1-and T2-weighted scans to generate an image with 1 mm isotropic resolution and MPRAGE contrast. By incorporating a segmentation-based regularizer, the proposed method can improve the quality of reconstruction. Yan et al. [28] introduced microbubble image features into a Kalman tracking framework, and made the framework compatible with sparsity-based deconvolution, in order to address the key challenges of tracking bubbles of high concentration at low frame rate.

In the field of SR, generative Adversarial Networks (GANs) [29] are also used to improve the visual quality of the generated images, called super-resolution generative adversarial network (SRGAN) [30].

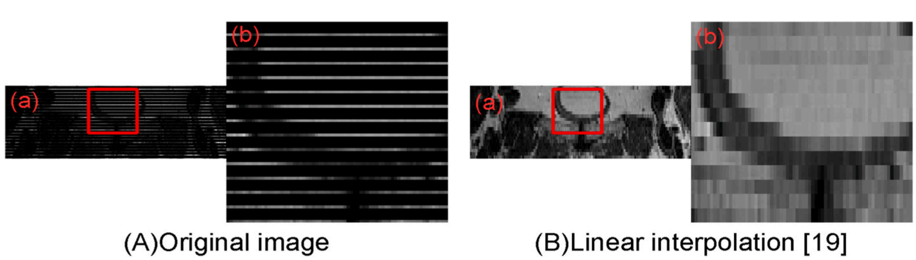

Fig. 1. Synthetic slices of linear interpolation (Linear) [11] in the through-plane direction.

However, the hallucinated details of SRGAN are often accompanied with unpleasant artifacts. For this reason, Wang et al. [31] proposed an enhanced super-resolution generative adversarial network (ESRGAN). They revisited the key components of SRGAN and improved the model by introducing the residual dense block without batch normalization as the basic network building unit.

Although the aforementioned methods perform well at reconstructing HR images from LR ones, they cannot generate corresponding segmentation labels, and some of them may change the original slices. On the contrary, SI can use bidirectional spatial transformations to generate intermediate slices and segmentation labels between every two consecutive slices. For this reason, SI can generate the HR volumes without changing the original slices, and the corresponding segmentation labels can be directly generated using the bidirectional spatial transformations.

2.3. Image augmentation based on spatial transformation

A large amount of training data is critical to the success of deep learning. However, in the field of medical imaging, a lack of training data is a significant challenge. Due to issues such as a lack of cases, insufficient medical resources, and costly labeling, researchers have turned to data augmentation to better utilize existing data [32,33]. For medical images, data augmentation is commonly preferred using random smooth flow fields to simulate anatomical variations [14]. Although this method can reduce overfitting and improve test performance [34,35], the selection of transformation functions and parameter settings tend to influence the improvement of performance [36].

Data augmentation methods based on learning spatial transformations from existing data have been proposed [37,36]. Hauberg et al. [37] aimed to improve MNIST digit classification performance through data enhancement. It learns digit-specific spatial transformations and samples training images and transformations to create new examples. Zhao et al. [36] proposed an automatic augmentation method that has the potential to improve the performance of brain MRI segmentation. The set of spatial and appearance transformations between the labeled atlas and unlabeled volumes is modeled using learning-based registration methods. It can use unlabeled images to synthesize diverse and realistic labeled samples by capturing anatomical and imaging diversity. However, these methods cannot be directly applied to our problem scenarios. The synthetic slices of these methods cannot be interpolated into the original volumes to make them isotropic because the slice smoothness is ignored. In our previous work [13], we proposed a method that generates synthetic inter-slice images based on frame interpolation and attention mechanism, called IIA. IIA makes use of the idea of frame interpolation to generate spatial transformation between two consecutive slices. The method can generate as many intermediate slices as needed by employing spatial transformations. However, IIA only focuses on smoothness in the in-plane direction, and it tends to perform poorly in generating 3D volumes with clear edges in the through-plane direction. SI proposes the use of a smoothness loss function that can evaluate the smoothness of 3D medical volumes in the through-plane direction to generate medical volumes with significantly clearer edges in order to improve slice smoothness in the through-plane direction.

3. Methods

By making 3D medical volumes isotropic and clear in both the in- and through-plane directions, we proposed a method, namely, SI, to improve the 3D medical image segmentation accuracy. Fig. 2 presents the proposed method.

SI's first step is to learn an inter-slice synthesis model, which is presented in detail in Fig. 2(a). Between every two consecutive slices in the through-plane direction, the model is used to generate intermediate slices. In this step, the slices of each volume in the through-plane direction are divided into multiple sets in sequence, each with $N+2$ slices, $\{I_n\}_{n=0}^{N+1}$, where N denotes the number of intermediate slices between two input slices. Given two input slices I_n and I_{N+1}, we synthesize the intermediate slices, $\{\widehat{I}_n\}_{n=1}^{N}$, which should be as close as possible to the ground-truth intermediate slices $\{I_n\}_{n=1}^{N}$.

In particular, the inter-slice synthesis network generates the bidirectional spatial transformations $\widehat{F}_{0\rightarrow N+1}$ and $\widehat{F}_{N+1\rightarrow 0}$. Then the intermediate spatial transformations $\widehat{F}_{n\rightarrow 0}$ and $\widehat{F}_{n\rightarrow N+1}$ can be approximated by combining the bidirectional spatial transformations as follows:

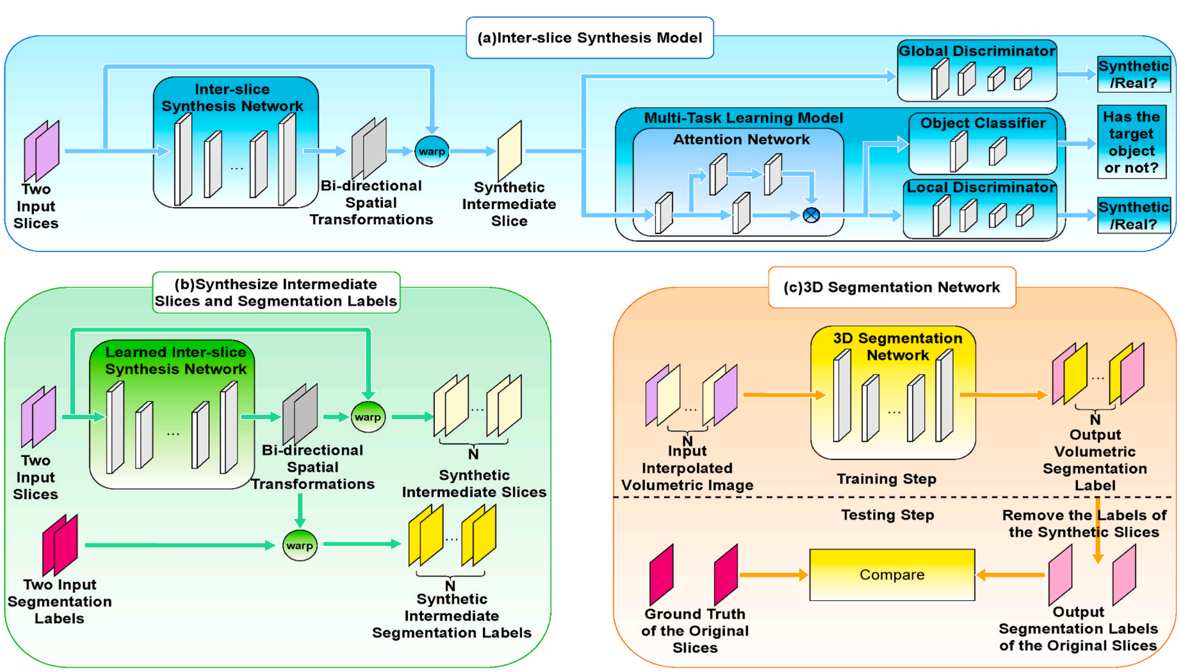

Fig. 2. An overview of the proposed method: (a) the architecture of the inter-slice synthesis model, (b) the process of synthesizing intermediate slices and their corresponding segmentation labels, and (c) the architecture of the 3D segmentation network. N is the number of interpolated slices between two consecutive slices.

$$\widehat{F}_{n\to 0} = -\left(1-\frac{n}{N+1}\right)\frac{n}{N+1}\widehat{F}_{0\to N+1} + \frac{n}{N+1}^{2}\widehat{F}_{N+1\to 0} \tag{1}$$

$$\widehat{F}_{n\to N+1} = \left(1-\frac{n}{N+1}\right)^{2}\widehat{F}_{0\to N+1} - \left(1-\frac{n}{N+1}\right)\frac{n}{N+1}\widehat{F}_{N+1\to 0} \tag{2}$$

The intermediate spatial transformations $\widehat{F}_{n\to 0}$ and $\widehat{F}_{n\to N+1}$ warp I_0 and I_{N+1}, respectively, to synthesize the inter slice $\widehat{I}_n$ as follows:

$$\widehat{I}_0 = \left(1-\frac{n}{N+1}\right)g(I_0,\widehat{F}_{n\to 0}) + \frac{n}{N+1}g(I_{N+1},\widehat{F}_{n\to N+1}) \tag{3}$$

where $g(\cdot,\cdot)$ denotes a backward-warping function, which is implemented with bilinear interpolation [38,39].

$\widehat{I}_n$ and I_n are further fed into the global discriminator and multitask learning model. The multitask learning model consists of an attention network, an object classifier, and a local discriminator. The object classifier detects whether the input slice contains the target object, whereas the two discriminators determine whether the slices are synthetic or real. The attention network focuses on the useful parts in the slice that help the object classifier and the local discriminator make predictions on their tasks. The object classifier and local discriminator are optimized together, and they share the same attention network, to take advantage of the interacting information between the two tasks.

The second step of SI is to create synthetic slices and their associated segmentation labels between two consecutive slices in the through-plane direction using the learned inter-slice synthesis network. The process of synthesizing intermediate slices and their corresponding labels is presented in Fig. 2(b). Given two consecutive input slices I_0 and I_1, and their corresponding segmentation labels L_0 and L_1, the learned inter-slice synthesis network generates the intermediate spatial transformations, $\widehat{F}_{n\to 0}$ and $\widehat{F}_{n\to 1}$, at position $n\in(0,1)$. I_0, I_1 and L_0, L_1 are warped by $\widehat{F}_{n\to 0}$ and $\widehat{F}_{n\to 1}$ to generate the intermediate slice and their corresponding segmentation label, $\widehat{I}_n$ and $\widehat{L}_n$, as follows:

$$\widehat{I}_n = (1-n)g(I_0,\widehat{F}_{n\to 0}) + ng(I_1,\widehat{F}_{n\to 1}) \tag{4}$$

$$\widehat{L}_n = (1-n)g(L_0,\widehat{F}_{n\to 0}) + ng(L_1,\widehat{F}_{n\to 1}) \tag{5}$$

The interpolated slices are used to train the segmentation network in the third step of SI. Fig. 2(c) depicts the segmentation networks in detail. The synthetic slices and segmentation labels are interpolated into the original volumes in the through-plane direction during the training of the 3D segmentation network to convert the 3D volumes into isotropic ones. We then train the 3D segmentation network with the isotropic volumes and the segmentation labels. After the training process, we use the learned inter-slice synthesis network to generate intermediate slices for test samples and interpolate the synthetic slices into the test samples in the through-plane direction to make them isotropic. We then remove the output segmentation labels of the synthetic intermediate slices after feeding the interpolated test volumes into the learned 3D segmentation network. The model's performance is evaluated by comparing the remaining parts of the segmentation labels with the ground-truth.

The inter-slice synthesis network: We construct the inter-slice synthesis network using the method proposed by Jiang et al. [12]. The loss function of the inter-slice synthesis network is defined as follows:

$$l = \lambda_{rec}l_{rec} + \lambda_{per}l_{per} + \lambda_{warp}l_{warp} + \lambda_{smooth}l_{smooth} + \lambda_{adv}l_{adv} + \lambda_{tp-smooth}l_{tp-smooth} \tag{6}$$

Equation (6) is a linear combination of six terms, where a set of coefficients $\{\lambda_{rec},\lambda_{per},\lambda_{warp},\lambda_{smooth},\lambda_{adv},\lambda_{tp-smooth}\}$ regularizes the contribution of the corresponding term.

The first term of (6) is l_{rec}. It is the reconstruction loss between the real slices and the synthetic slices. The loss function of l_{rec} is defined as follows:

$$l_{rec} = \frac{1}{N}\sum_{i=1}^{N}\|\widehat{I}_n - I_n\|_1 \tag{7}$$

The second term of (6) is l_{per}. It is the perceptual loss to measure perceptual difference between $\widehat{I}_n$ and I_n, which can preserve details of the predictions and make interpolated frames sharper [12]. The loss function of l_{per} is defined as follows:

$$l_{per} = \|\varphi(\widehat{I}) - \varphi(I_n)\|_2 \tag{8}$$

where φ means the conv4_3 features of an ImageNet pre-trained VGG16 model [40].

The third term of (6) is l_{warp}. It is the warping loss, which models the quality of the spatial transformation [12]. The loss function of l_{warp} is defined as follows:

$$l_{warp} = \|I_0 - g(I_{N+1},F_{N+1\to 0})\|_1 + \|I_{N+1} - g(I_0,F_{0\to N+1})\|_1 + \frac{1}{N}\sum_{i=1}^{N}\|I_n - g(I_0,\widehat{F}_{0\to n})\|_1 + \frac{1}{N}\sum_{i=1}^{N}\|I_n - g(I_{N+1},\widehat{F}_{1\to n})\|_1 \tag{9}$$

The fourth term of (6) is l_{smmoth}. It is the smoothness loss, which encourages neighboring pixels in the in-plan direction to have similar transformation values [12]. The loss function of l_{smmoth} is defined as follows:

$$l_{smooth} = \|\nabla F_{0\to N+1}\|_1 + \|\nabla F_{N+1\to 0}\|_1 \tag{10}$$

The fifth term of (6) is l_{adv}. It is the adversarial loss, which encourage the generator to synthesis image to confuse the discriminator. It can improve the authenticity of the synthetic images. The loss function of l_{adv} is defined as follows:

$$l_{adv} = \frac{-1}{N}\sum_{i=1}^{N}logLD(Att(\widehat{I}_n)) - \frac{1}{N}\sum_{i=1}^{N}logGD(\widehat{I}_n) \tag{11}$$

where *Att* means the attention network, *LD* means the local discriminator, *GD* means the global discriminator.

The fifth term of (6) is $l_{tp-smooth}$. It is the smoothness term to encourage adjacent pixels of the interpolated volumes in the through-plane direction to have similar values. The loss function of $l_{tp-smooth}$ is defined as follows:

$$l_{tp-smooth} = \frac{1}{LW}\sum_{i=1}^{L}\sum_{j=1}^{W}\left(\left(I_{tp}(i,j-1) - I_{tp}(i,j)\right)^2 + \left(I_{tp}(i+1,j) - I_{tp}(i,j)\right)^2\right) \tag{12}$$

where $I_{tp}\in\{I_{sagittal},I_{coronal}\}$ denotes the slices in the through-plane directions: $I_{sagittal}$, the volume slices in the sagittal direction; and $I_{coronal}$, the volume slices in the coronal direction. $I_{tp}(i,j)$ denotes the value in (i,j) of I_{tp}, L denotes the length of I_{tp} and W denotes the width of I_{tp}.

The global discriminator and the multitask learning model: An attention network, an object classifier, and a local discriminator are the components of the multitask learning model. While the object classifier detects whether the input slices contain the target objects, the global and local discriminators compete with the inter-slice synthesis network. The attention network can automatically focus on the interacting information between the classification and the distinguishing tasks by training the local discriminator and the object classifier simultaneously.

The loss function for the global discriminator is defined as follows:

$$l_{global} = \frac{-1}{N}\sum_{i=1}^{N} log(1 - GD(\widehat{I}_n)|) - \frac{1}{N}\sum_{i=1}^{N} logGD(I_n) \tag{13}$$

The loss function for the multitask learning model is defined as follows:

$$l_{mul} = \left(-\frac{1}{\mathrm{N}}\sum\nolimits_{i=1}^{\mathrm{N}}\log(1 - LD(Att(\widehat{I}_n))) - \frac{1}{\mathrm{N}}\sum\nolimits_{i=1}^{\mathrm{N}}\log LD(Att(I_n))\right) + \left(-\frac{1}{\mathrm{N}}\sum\nolimits_{i=1}^{\mathrm{N}}\widehat{Y}_n \log OC\,(Att(\widehat{I}_n)) - \frac{1}{N}\,Y_n\sum\nolimits_{i=1}^{N}\log OC\,(Att(I_n))\right) \tag{14}$$

where OC denotes the object classifier network. $Y_n \in \{0,1\}$ and $\widehat{Y}_n \in \{0,1\}$ denote whether the real and synthetic slices contain the target object. $Y_i = 1$ and $\widehat{Y}_i = 1$ if the input original and synthetic slices contain the target object, and $Y_i = 0$ and $\widehat{Y}_i = 0$ otherwise.

4. Experiment

We provide both quantitative and qualitative performance evaluations for SI on three 3D medical imaging datasets.

4.1. Datasets

Brain tumors segmentation (BTS) in Medical Segmentation Decathlon [41] is the first dataset in our experiment. All scans in BTS are co-registered to a reference atlas space using the SRI24 brain structure template [42], resampled to isotropic voxel resolution of $1mm^3$, and skull-stripped using various methods followed by manual refinements. We segment the whole tumors in our experiment by selecting 100 FLAIR modal MRI data in this dataset. We use training data from 56 scans, validation data from 14 scans, and test data from 30 scans. The second dataset is liver tumor segmentation (LTS) in Medical Segmentation Decathlon [41]. The LTS slices are generated by a variety of different scanning devices with intra-slice and inter-slice distances ranging from 0.5 to 1 mm and 0.45–6.0 mm, respectively. A total of 131 portal venous phase computed tomography (CT) scans with two annotated objects (liver and tumor) are selected. We use 74, 18, and 39 scans as training, validation, and test data, respectively. Prostate segmentation (PS) in Medical Segmentation Decathlon [41] is the third dataset. PS includes 32 transverse T2-weighted scans with two annotated objects (prostate peripheral zone and the transition zone), each with voxel resolution $0.6 \times 0.6 \times 4mm^3$. We use 17, 6, and 9 scans as training, validation, and test data, respectively.

4.2. Evaluation

Dice score: Dice score [43] quantifies the overlap between two segmentation labels. The formulation of the Dice score is shown as follows:

$$Dice(L,\widehat{L}) = 2 \times \left(\frac{|L \cap \widehat{L}|}{|L| + |\widehat{L}|}\right) \times 100\% \tag{15}$$

where L denotes the ground truth of the real image and $\widehat{L}$ denotes the predicted segmentation label. If the Dice score is 0, the two labels have no overlap. With the Dice score increasing, the two labels have more overlap. When the Dice score is 1, the two labels have completely overlap. A better model will have a higher Dice score.

Relative absolute volume difference: The relative absolute volume difference (RAVD) [44] reveals if a method tends to over- or under segment. The formulation of RAVD is defined as follows:

$$RAVD(L,\widehat{L}) = \left(\frac{|\widehat{L}| - |L|}{|L|}\right) \times 100\% \tag{16}$$

A value of 0 means both volumes are identical. A better model will have a lower RAVD.

Average symmetric surface distance: Average symmetric surface distance (ASSD) [44] is given in millimeters and based on the surface voxels of two segmentations L and $\widehat{L}$. The formulation of ASSD is defined as follows:

$$ASSD(L,\widehat{L}) = \frac{1}{|S(L)| + |S(\widehat{L})|}\left(\sum_{l \in S(L)} \min_{\widehat{l} \in S(\widehat{L})} \|l - \widehat{l}\|_2 + \sum_{\widehat{l} \in S(\widehat{L})} \min_{l \in S(L)} \|l - \widehat{l}\|_2\right) \tag{17}$$

where $S(\cdot)$ denote the set of surface voxels of volumes. For each surface voxel of L, the Euclidean distance to the closest surface voxel of $\widehat{L}$ is calculated. In order to provide symmetry, the same process is applied from the surface voxels of $\widehat{L}$ to L. ASSD is then defined as the average of all distances, which is 0 for a perfect segmentation. A better model will have a lower ASSD.

Maximum symmetric surface distance: Maximum symmetric surface distance (MSSD) [44] is given in millimeter and based on the surface voxels of two segmentations L and $\widehat{L}$. The formulation of MSSD is defined as follows:

$$MSSD(L,\widehat{L}) = max\left\{\max_{l \in S(L)}\min_{\widehat{l} \in S(\widehat{L})} \|l - \widehat{l}\|_2, \max_{\widehat{l} \in S(\widehat{L})}\min_{l \in S(L)} \|l - \widehat{l}\|_2\right\} \tag{18}$$

Different from ASSD, surface voxels of MSSD are determined using Euclidean distances, and the maximum value yields MSSD. For a perfect segmentation MSSD is 0. A better model will have a lower MSSD.

4.3. Implementation

To implement SI, we divide the training data into multiple sets in sequence, each with $N + 2$ slices. We will discuss the setting of N in Section 5.2. The first four hyper-parameters of (6) are determined according to Ref. [13]. Furthermore, we use five-fold cross-validation to select λ_{adv} and $\lambda_{tp-smooth}$. λ_{adv} is set to 0.050 and $\lambda_{tp-smooth}$ is set to 0.467. As presented in Fig. 2(a), we optimize the attention network, object classifier, and local discriminator together. Two fully connected layers comprise the object classifier, whereas, three convolutional layers, a fully connected layer, and a sigmoid function comprise the global and local discriminators. Two branches make up the attention network. The features of the slices are extracted by one convolutional layer, and the corresponding attention masks are generated by two convolutional layers in the other branch. The inter-slice synthesis network is trained in 30 epochs using the basic learning rate of 0.001, and the batch size is set to 2. To optimize all of the networks in the inter-slice synthesis model, we use Adam optimizer [45]. To better introduce the detail of hyper parameters, we show the value of hyper parameters in SI in Table 1.

In the experiment of BTS, EDSR [23], ESRGAN [31], RCAN [24], Linear [11], and IIA [13] are compared with our method. The first three methods transform the LR images in the through-plane direction into HR ones and do not generate the corresponding segmentation labels of the synthetic slices. To leverage unlabeled slices, the 2D uncertainty aware self-ensembling mean teacher model [46], which is a semi-supervised segmentation model, is employed to segment the medical volumes

Table 1
Hyper parameters of SI.

	λ_{rec}	λ_{per}	λ_{warp}	λ_{smooth}	λ_{adv}	$\lambda_{tp-smooth}$
value	2.000	0.005	1.000	1.000	0.050	0.467

augmented by the first three methods. The 2D uncertainty aware self-ensembling mean teacher model contains two models, the teacher model and the student model. The backbone of the teacher and student models is U-Net [14]. For a fair comparison, U-Net is employed to segment the medical volumes augmented by the other methods. In the experiments of LTS and PS, Linear [11], IIA [13] are compared with our proposed method. All methods transform the anisotropic medical volumes into isotropic ones and use nnU-Net [20] to segment the medical volumes.

4.4. Comparison with SR algorithms

In this session, we compare our method's performance with the baselines on whole tumor segmentation of BTS. To make the data in BTS anisotropic, the R-th slices of the volumes in the through-plane direction are removed. R is defined as follows:

$$R = \{r \vee r\%4 \neq 0, 0 \leq r < H\} \tag{19}$$

where H denotes the number of slices of the volumes in the through-plan direction.

4.4.1. Visualization performance of synthetic slices

Synthetic slices of different methods and their difference maps with the original slices in BTS are presented in Fig. 3. The difference maps of EDSR and ESRGAN in Fig. 3 indicate that, compared with other methods, the synthetic slices of these methods are significantly different from the original slices (columns 3, 5 in Fig. 3). For Linear, the synthetic slices are less different from the original slices compared with EDSR and ESRGAN. However, the synthetic slices of Linear have aliasing on the edges of the object (columns 9 in Fig. 3). For RCAN, although it can recover the slices that are similar to the ground-truth, it changes the original slices (column 7 in Fig. 3): thus, it may cause a negative impact on the segmentation and cannot be directly applied to our problem scenario. By contrast, since IIA, SI w/o (without) $l_{tp-smooth}$, and SI interpolate slices into the volumes to make it isotropic, the original slices do not change. Although IIA and SI w/o $l_{tp-smooth}$ can recover slices that are less different from the original slices on the object's edge, they are unable to recover the texture of the object (columns 11, 13 in Fig. 3). For SI, its performance of recovering on the edges and texture is better than those of IIA and SI w/o $l_{tp-smooth}$, which means that the smoothness function of SI can improve the authenticity of the synthetic slices (columns 15 in Fig. 3).

4.4.2. Segmentation performance evaluations

The segmentation accuracies obtained by different methods are presented in Table 2. In all tables, ↑ denotes higher is better, whereas ↓ denotes lower is better. The best results are in **bold**, and the second-best results are <u>underlined</u>. As can be seen from Table 2, all the SR algorithms

Table 2
Performance on the BTS dataset for whole tumor segmentation.

Method	Dice(%) ↑	RAVD(%) ↓	ASSD(%) ↓	MSSD(%) ↓
U-Net [14]	82.91	18.97	2.97	13.68
EDSR [23]	80.25	20.87	3.09	14.83
ESRGAN [31]	81.79	19.03	3.47	16.74
RCAN [24]	82.00	19.37	3.07	14.99
Linear [11]	68.69	38.33	7.75	38.06
IIA [13]	85.72	38.33	2.28	8.50
SI w/o $l_{tp-smooth}$	<u>86.36</u>	<u>14.71</u>	**2.14**	<u>8.35</u>
SI	**87.00**	**13.16**	<u>2.15</u>	**8.27**

perform worse than U-Net. Although SR algorithms can recover slices that are less different from the original slices, they are not suitable for helping with segmentation, because the original slices are changed. Moreover, the training of the SR algorithm needs LR/HR pairs; thus, these methods cannot be trained with LR slices only. Hence, the SR algorithm is not suitable for our problem scenario. Linear has the worst performance in the experiments. One possible reason is that the distance of the original slices is large. It's difficult to recover the feature of the missing slices using linear interpolation in the through-plane direction. By contrast, IIA, SI w/o $l_{tp-smooth}$, and SI perform better than U-Net. Because all of the three methods generate intermediate slices and labels without changing the original slices, and can provide more labeled training data. Furthermore, all of the three methods use the local discriminator, which makes the methods pay more attention to the authenticity of the target object. Thus, the segmentation model can better learn the feature of the target object, and ultimately improve the segmentation performance. SI w/o $l_{tp-smooth}$ performs better than IIA, which means that the multitask learning mechanism is useful for boosting segmentation performance. In Addition, due to the smoothness loss function, SI performs better than SI w/o $l_{tp-smooth}$ overall.

4.5. Comparison with algorithms that transform the data into isotropic data

In this section, we compare our method's performance with the baselines on 3D segmentation tasks of LTS and PS. For isotropic volume, the intervals of the in-plane and through-plane directions are equal, namely the inter-slice distance of an isotropic volume is equal to its intra-slice distance. To transform the interpolated volumes into isotropic ones, the synthetic slices can be interpolated into the anisotropic volume in the through-plane direction to reduce its intra-slice distance. In the experiment, the number of interpolated slices between two consecutive slices is calculated using the following equation:

$$N_A = \left\lfloor \frac{D_{inter}}{D_{intra}} \right\rfloor - 1, \tag{20}$$

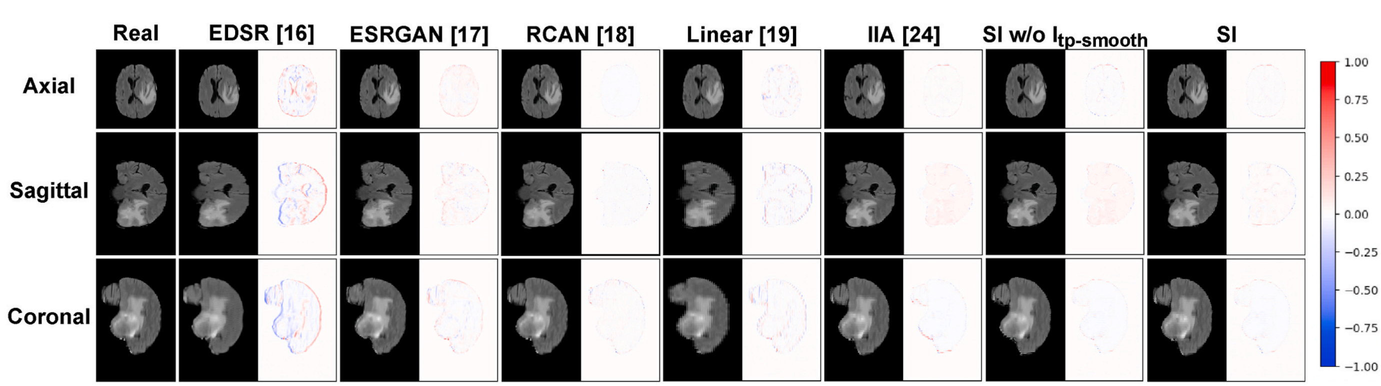

Fig. 3. Synthetic slices of different methods in BTS from different directions: the in-plane (axial view), and through-plane (sagittal and coronal views) directions. The difference maps are provided to the right of the results for better visualization.

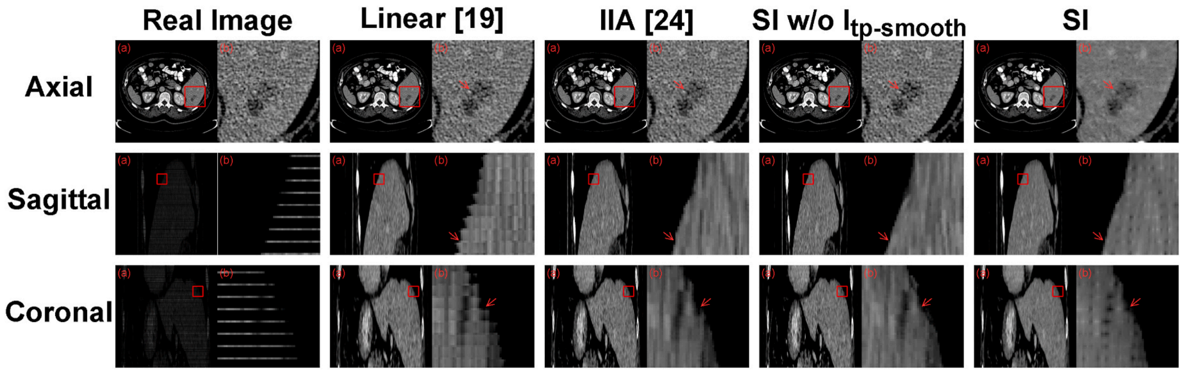

Fig. 4. Synthetic slices of different methods in LTS from different directions: the in-plane (axial view), and through-plane (sagittal and coronal views) directions.

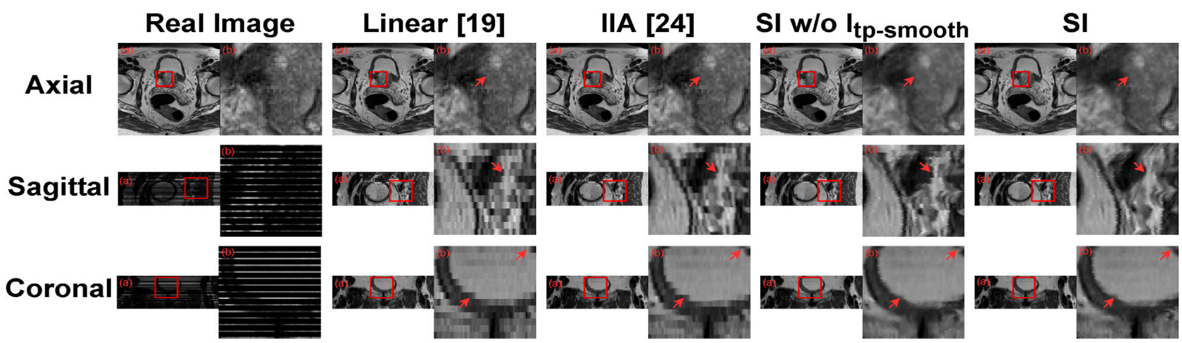

Fig. 5. Synthetic slices of different methods in PS from different directions: the in-plane (axial view), and through-plane (sagittal and coronal views) directions.

where D_{inter} denotes the inter-slice distance of the volume; D_{intra} denotes the intra-slice distance of the volume; and $\lfloor . \rfloor$ denotes a floor function that outputs the greatest integer that is less than or equal to the input.

4.5.1. Visualization performance of synthetic slices

The synthetic slices of different methods in LTS and PS are presented in Fig. 4 and Fig. 5. The synthesized slice of SI w/o $l_{tp-smooth}$ has a clearer outline of tumor (red arrow) than Linear and IIA, as shown in the first row of Fig. 4. The second and third rows of Figs. 4 and 5 show that SI w/o $l_{tp-smooth}$ can generate volumes with clearer edges than Linear and IIA. The results described above indicate that the multitask learning mechanism helps the model capture more details of the object and generate more realistic slices. The synthesized slices of SI are less noisy than those of the other methods in Figs. 4 and 5. This is because the smoothness loss function of SI aims at encouraging neighboring pixels to have similar values. The proposed method can generate slices with more spatial smoothness in the through-plane directions by reducing noise in the synthetic slices using the smoothness loss function. On the contrary, since Linear, IIA, and SI w/o $l_{tp-smooth}$ do not incorporate slice smoothness in the through-plane direction, their synthetic slices suffer aliasing on the edges of the object (rows 2 and 3 in Figs. 4 and 5).

4.5.2. Segmentation performance evaluations

Table 3 and Table 4 present the 3D segmentation performance of various methods. We also visualize the activation maps extracted by SI's attention networks to demonstrate the effectiveness of the multitask mechanism in SI. Fig. 6 presents the activation maps of two specific slices. Since SI w/o ltp−smooth and SI use the same attention mechanism, we only show the activation maps of SI. Tables 3 and 4 show that in most cases, nnU-Net and Linear perform worse than IIA, SI w/o $l_{tp-smooth}$, and SI. Since nnU-Net directly down-samples the HR axes of the volumes to transform the data into isotropic ones, it fails to fully exploit the information of the volumes. Furthermore, Linear directly increases slices resolution, which substantially changes the anatomical structure between consecutive slices. For most cases, SI w/o $l_{tp-smooth}$ performs better than IIA, suggesting that the multitask learning mechanism helps boosting segmentation performance. Fig. 6 demonstrates that IIA only focuses on the local target object, but SI highlights the whole target object (red arrows in Fig. 6). The results indicate that the multitask learning model enables the attention network to capture more detailed information of the slices and leads to better segmentation performances. Overall, benefiting from the smoothness loss function, SI achieves the best performance.

Table 3
Performance of different methods over two annotated objects, namely, liver and tumor, and the mean scores of two annotated objects in LTS.

Method	Dice(%) ↑			RAVD(%) ↓			ASSD(%) ↓			MSSD(%) ↓		
	Liver	Tumor	Mean	Liver	Tumor	Mean	Liver	Tumor	Mean	Liver	Tumor	Mean
nnU-Net [20]	94.11	55.81	74.96	6.69	49.01	27.85	3.00	13.15	8.08	60.53	60.48	60.51
Linear [11]	94.01	55.54	74.78	7.10	46.60	26.85	2.60	12.91	7.76	51.09	60.19	55.64
IIA [13]	94.30	56.88	75.79	**5.76**	47.22	26.49	1.97	15.41	8.69	**36.76**	68.70	53.23
SI w/o $l_{tp-smooth}$	**95.14**	56.37	75.76	6.49	46.48	26.48	2.44	**10.73**	**6.59**	53.43	**49.15**	51.29
SI	94.57	**57.38**	**75.98**	6.14	**44.55**	**25.35**	**1.95**	12.69	7.32	52.09	49.59	**50.84**

Table 4
Performance over two annotated objects, namely, prostate peripheral zone (PZ) and the transition zone (TZ), and the mean scores of two annotated objects in PS.

Method	Dice(%) ↑			RAVD(%) ↓			ASSD(%) ↓			MSSD(%) ↓		
	PZ	TZ	Mean	PZ	TZ	Mean	PZ	TZ	Mean	PZ	TZ	Mean
nnU-Net [20]	87.79	83.81	85.81	9.66	18.98	14.32	0.57	0.68	0.63	4.39	**5.92**	5.15
Linear [11]	89.78	84.74	87.26	8.39	17.29	12.84	0.49	0.64	0.56	3.99	6.12	5.06
IIA [13]	89.97	85.72	87.84	7.49	15.86	11.68	0.47	0.59	0.53	3.90	6.38	5.14
SI w/o $l_{tp-smooth}$	89.96	86.19	88.08	7.50	**15.63**	11.56	0.47	**0.58**	**0.52**	3.76	5.96	**4.86**
SI	**90.29**	**86.33**	**88.31**	**6.31**	16.46	**11.39**	**0.46**	**0.58**	**0.52**	**3.56**	6.27	4.91

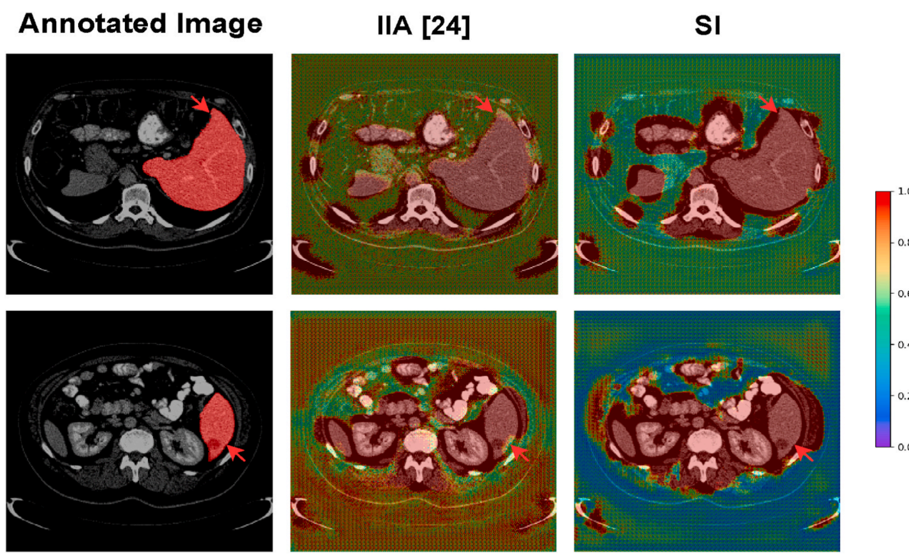

Fig. 6. Visualization of the activation map.

5. Ablation study

In this section, we perform a detailed ablation analysis of the network to validate the effect of the local and global discriminators and determine the optimal numbers of intermediate slices. In section 5.1, we investigate the effect of the global and local discriminators. In section 5.2, the effect of the number of intermediate slices is studied.

5.1. Effect of GD and LD

To validate the effect of the local and global discriminators, we conduct detailed ablation studies in LTS and PS. (1) Only the intermediate slice synthesis network is used, without using the two discriminators (global discriminator, GD, and local discriminator, LD). (2) To highlight the effect of GD, the intermediate slice synthesis network is used to battle with the GD. (3) To highlight the effect of LD, the intermediate slice synthesis network is used to battle with the LD. (4) The intermediate slice synthesis network is used to battle with both the GD and LD.

Tables 5 and 6 show the segmentation performance under different module configurations. As expected, SI with discriminators outperforms SI without any discriminators. It is worth noting that rows 2 to 4 in Table 5 and Table 6 have lower ASSD and MSSD than row 1 in Tables 5 and 6. This shows that GD and LD are effective to help SI to generate slices with more realistic edges. Compared with only using GD or LD, using GD and LD simultaneously has the best performance.

5.2. Effect of the numbers of intermediate slices

We conduct four segmentation experiments using the inter-slice synthesis network to generate different numbers of intermediate slices between two consecutive slices to investigate the effect of intermediate slice numbers on our method. The results are presented in Table 7 and Table 8. N_A, the number of automatically interpolated slices, is determined by the original data according to (20).

Tables 7 and 8 show that in most cases, with N_A slices added between every two consecutive slices, our method yields the best results (row 4 in Tables 7 and 8). This is because the distance between two consecutive slices of different volumes is not the same, so if we interpolate a fixed number of slices between two consecutive slices, not all the volumes become isotropic, some may remain anisotropic. By using N_A as the number of interpolated slices between two consecutive slices, SI can

Table 5
The segmentation over two annotated objects, namely, liver and tumor, and the mean scores of two annotated objects in LTS for the detailed ablation studies of each modules in our framework.

No.	Module		Dice(%) ↑			RAVD(%) ↓			ASSD(%) ↓			MSSD(%) ↓		
	GD	LD	Liver	Tumor	Mean	Liver	Tumor	Mean	Liver	Tumor	Mean	Liver	Tumor	Mean
1			94.27	56.26	75.27	6.49	53.51	30.00	2.79	12.20	7.50	59.77	55.72	57.75
2	✓		94.30	56.27	75.29	6.40	54.86	30.63	2.62	**11.07**	**6.85**	54.43	53.47	53.95
3		✓	94.35	**57.51**	75.93	**5.98**	45.50	25.74	2.10	12.23	7.17	58.22	50.75	54.59
4	✓	✓	**94.57**	57.38	**75.98**	6.14	**44.55**	**25.35**	**1.95**	12.69	7.32	**52.09**	**49.59**	**50.84**

Table 6
Performance over two annotated objects, namely, prostate peripheral zone (PZ) and the transition zone (TZ), and the mean scores of two annotated objects in PS.

No.	Module		Dice(%) ↑			RAVD(%) ↓			ASSD(%) ↓			MSSD(%) ↓		
	GD	LD	PZ	TZ	Mean	PZ	TZ	Mean	PZ	TZ	Mean	PZ	TZ	Mean
1			89.94	83.94	86.94	7.64	20.32	13.98	0.57	0.70	0.64	7.76	7.12	7.44
2	✓		90.00	84.53	87.27	6.67	18.78	12.73	0.49	**0.48**	**0.49**	3.70	6.78	5.24
3		✓	90.16	84.95	87.56	6.54	18.12	12.33	0.52	0.62	0.57	3.84	6.66	5.25
4	✓	✓	**90.29**	**86.33**	**88.31**	**6.31**	**16.46**	**11.39**	**0.46**	0.58	0.52	**3.56**	**6.27**	**4.91**

Table 7
Performance of interpolating different numbers of intermediate slices over two annotated objects, namely, liver and tumor, and the mean scores of two annotated objects in LTS.

Interpolated Slices	Dice(%) ↑			RAVD(%) ↓			ASSD(%) ↓			MSSD(%) ↓		
	Liver	Tumor	Mean	Liver	Tumor	Mean	Liver	Tumor	Mean	Liver	Tumor	Mean
1	94.21	55.15	74.68	6.63	47.72	27.18	2.38	12.34	7.36	**47.70**	60.11	53.91
2	90.90	48.59	69.75	12.63	56.07	34.35	3.60	16.25	9.93	61.54	81.49	71.52
3	**95.11**	54.98	75.05	**5.74**	50.51	28.12	2.44	**10.28**	**6.36**	59.44	54.65	57.05
N_A	94.57	**57.38**	**75.98**	6.14	**44.55**	**25.35**	**1.95**	12.69	7.32	52.09	**49.59**	**50.84**

Table 8
Performance of interpolating different numbers of intermediate slices over two annotated objects, namely, prostate peripheral zone (PZ) and the transition zone (TZ), and the mean scores of two annotated objects in PS.

Interpolated Slices	Dice(%) ↑			RAVD(%) ↓			ASSD(%) ↓			MSSD(%) ↓		
	PZ	TZ	Mean	PZ	TZ	Mean	PZ	TZ	Mean	PZ	TZ	Mean
1	89.73	85.42	87.58	9.16	21.73	15.44	0.93	1.11	1.02	4.83	6.72	5.77
2	89.22	85.21	87.21	9.51	21.60	15.55	0.77	0.92	0.84	4.39	**5.91**	5.15
3	89.49	85.65	87.57	8.65	20.35	14.50	0.63	0.77	0.70	4.06	5.93	5.00
N_A	**90.29**	**86.33**	**88.31**	**6.31**	**16.46**	**11.39**	**0.46**	**0.58**	**0.52**	**3.56**	6.27	**4.91**

adaptively transform all the volumes into isotropic ones, which improves the 3D segmentation performance.

6. Conclusion

In this paper, we propose a novel multitask frame-interpolation-based method for slice imputation whereby the number of slices and corresponding labels is increased between two consecutive slices. SI can generate more realistic 3D medical volumes by evaluating the smoothness of the interpolated 3D medical volumes in the through-plane direction with a smoothness loss function, which improves the accuracy of 3D medical image segmentation. Furthermore, SI can generate as many intermediate slices as needed between two consecutive slices to transform the 3D medical volumes into isotropic ones. The experiments comparing with the baseline methods on BTS, LTS, and PS demonstrate the superior performances of SI. The smoothness loss function and the multitask learning model are shown to be effective in ablation experiments. 3D segmentation networks can achieve the best performance if the number of interpolated slices is appropriate to transform the anisotropic 3D medical volumes into isotropic ones.

There are still a lot of works to improve the proposed method. The performance of SI may degrade when encountering domain shift, i.e, attempting to apply the learned models on different modalities or organs that have different distributions from the training data.

The exploration of the above limitations leads us to future directions. Domain adaptation can be used in our future work to model the shift between datasets of different modalities or organs [47]. Furthermore, for different modalities, the idea of cross-modality translation can be applied in our model. The idea of CycleGAN [48] can be added in our model to learn the feature of different modalities [49].

Acknowledgements

This work is supported in part by the Natural Science Foundation of Guangdong Province (2020A1515010717), the Fundamental Research Funds for the Central Universities (2019MS073), NSF-1850492 (to R.L.) and NSF-2045804 (to R.L.).

References

[1] G. Litjens, T. Kooi, B. Ehteshami, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W. M. van der Laak, B. van Ginneken, C.I. Sanchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88. Dec.
[2] D.D. Ruikar, K.C. Santosh, R.S. Hegadi, Automated fractured bone segmentation and labeling from CT images, J. Med. Syst. 43 (No. 3) (2019) 1–13. Feb.
[3] L. Canalini, J. Klein, D. Miller, R. Kikinis, Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery, Int. J. Comput. Assist. Radiol. Surg. 14 (No. 10) (2019) 1697–1713. Aug.
[4] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: a nested U-Net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, pp. 3–11. Sep.
[5] X. Li, H. Chen, X. Qi, Q. Dou, C. Fu, P. Heng, H-DenseUNet, Hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE Trans. Med. Imag. 37 (12) (2018) 2663–2674. Jun.
[6] A.G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 2018, pp. 421–429. Sepp. 16–20.
[7] J. Du, Z. He, L. Wang, A. Gholipour, Z. Zhou, D. Chen, Y. Jia, Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network, Neurocomputing 392 (2020) 209–220. Jun.
[8] J.V. Manjón, P. Coupé, A. Buades, V. Fonov, D.L. Collins, M. Robles, Non-local MRI upsampling, Medical image analysis 14 (6) (2010) 784–792. Dec.
[9] S. Liu, D. Xu, S.K. Zhou, O. Pauly, S. Grbic, T. Mertelmeier, J. Wicklein, A. Jerebko, W. Cai, D. Comaniciu, 3D anisotropic hybrid network: transferring convolutional features from 2D images to 3D anisotropic volumes, in: International Conference

on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 2018, pp. 851–858. Sepp. 16–20.
[10] A.V. Dalca, K.L. Bouman, W.T. Freeman, N.S. Rost, M.R. Sabuncu, P. Golland, Medical image imputation from image collections, IEEE Trans. Med. Imag. 38 (2) (2019) 504–514. Feb.
[11] X. He, S. Yang, G. Li, H. Chang, Y. Yu, Non-local context encoder: robust biomedical image segmentation against adversarial attacks, Proc. AAAI Conf. Artif. Intell. 33 (2019) 865–872. Jul.
[12] H. Jiang, D. Sun, V. Jampani, M. Yang, E. Learned-Miller, J. Kautz, Super SloMo: high quality estimation of multiple intermediate frames for video interpolation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9000–9008. Salt Lake City, America, Jun. 18-22.
[13] Z. Wu, J. Wei, W. Yuan, J. Wang, T. Tasdizen, Inter-slice image augmentation based on frame interpolation for boosting medical image segmentation accuracy, in: 24th European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 2020, pp. 130–139. Aug. 29- Sep. 8.
[14] O. Ronneberger, P. Fischer, T. Brox, U-Net, Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer Assisted Intervention, 2015, pp. 234–241. Munich, Germany, Oct. 5-9.
[15] Y. Qin, K. Kamnitsas, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi, A. Nori, Autofocus layer for semantic segmentation, in: Medical Image Computing and Computer Assisted Intervention, Granada, Spain, 2018, pp. 603–611. Sepp. 16–20.
[16] C. Chen, C. Biffi, G. Tarroni, S. Petersen, W. Bai, D. Rueckert, Learning shape priors for robust cardiac MR segmentation from multi-view images, Medical Image Computing and Computer Assisted Intervention (2019) 523–531. Shenzhen, China, Oct. 13-17.
[17] C. Chen, D. Qi, H. Chen, J. Qin, P. Heng, Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 865–872. Jul.
[18] S. Liu, D. Xu, S.K. Zhou, S. Grbic, W. Cai, D. Comaniciu, Anisotropic hybrid network for cross-dimension transferable feature learning in 3D medical images, in: Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics, 2019, pp. 199–216. Sep.
[19] K. Lee, J. Zung, P. Li, V. Jain, H.S. Seung, Superhuman Accuracy on the SEMI3D Connectomics Challenge, 2017 arXiv preprint arXiv:1706.00120, May.
[20] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Kohler, T. Norajitra, S. Wirkert, K.H. Maier-Hein, nnU-Net: Self-Adapting Framework for U-Net-Based Medical Image Segmentation, 2018 arXiv preprint arXiv:1809.10486, Sep.
[21] Q. Delannoy, C.H. Pham, C. Cazorla, et al., SegSRGAN: super-resolution and segmentation using generative adversarial networks—application to neonatal brain MRI[J], Comput. Biol. Med. 120 (2020), 103755.
[22] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654. Las Vegas, Nevada, Jun. 26- Jul. 1.
[23] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144. Venice, Italy, Oct. 22-29.
[24] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 286–301. Munich, Germany, Sepp. 8–14.
[25] H. Wang, L. Lin, H. Hu, et al., Patch-free 3D medical image segmentation driven by super-resolution technique and self-supervised guidance[C], in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2021, pp. 131–141.
[26] S. Zhang, G. Liang, S. Pan, et al., A fast medical image super resolution method based on deep learning network[J], IEEE Access 7 (2018) 12319–12327.
[27] J.E. Iglesias, R. Schleicher, S. Laguna, et al., Accurate Super-resolution Low-Field Brain MRI[J], 2022 arXiv preprint arXiv:2202.03564.
[28] J. Yan, T. Zhang, J. Broughton-Venner, et al., Super-Resolution Ultrasound through Sparsity-Based Deconvolution and Multi-Feature Tracking[J], IEEE Transactions on Medical Imaging, 2022.
[29] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks, 2014 arXiv preprint arXiv:1406.2661, Jun.
[30] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic Single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690. Honolulu, Hawaii, Jul. 22-25.
[31] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C.C. Loy, ESRGAN: enhanced super-resolution generative adversarial networks, in: Proceedings of the European Conference on Computer Vision Workshops, 2018. Munich, Germany, Sepp. 8–14.
[32] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6 (No. 1) (2019). Jul.
[33] N. Tajbakhsh, L. Jyeaseelan, Q. Li, J.N. Chiang, Z. Wu, X. Ding, Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation, Med. Image Anal. 63 (Jul) (2020).
[34] H.R. Roth, C.T. Lee, H.C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, R.M. Summers, Anatomy-specific classification of medical images using deep convolutional nets, in: 2015 IEEE 12th International Symposium on Biomedical Imaging, ISBI), New York, NY, USA, 2015, pp. 101–104. Apr. 16-19.
[35] F. Milletari, N. Navab, S.A. Ahmadi, V-Net, Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV), Stanford, CA, USA, 2016, pp. 565–571. Oct. 25-28.
[36] A. Zhao, G. Balakrishnan, F. Durand, J.V. Guttag, A.V. Dalca, Data augmentation using learned transformations for one-shot medical image segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8543–8553. Long Beach, California, Jun. 16-20.
[37] S. Hauberg, O. Freifeld, A.B.L. Larsen, J.W. Fisher, L.K. Hansen, Dreaming more data: class-dependent distributions over diffeomorphisms for learned data augmentation, in: Artificial Intelligence and Statistics, 2016, pp. 342–350.
[38] T. Zhou, S. Tulsiani, W. Sun, J. Malik, A.A. Efros, View synthesis by appearance flow, in: European Conference on Computer Vision, 2016, pp. 286–301. Amsterdam, The Netherlands, Oct. 11-14.
[39] Z. Liu, R.A. Yeh, X. Tang, Y. Liu, A. Agarwala, Video frame synthesis using deep voxel flow, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4463–4471. Venice, Italy, Oct. 22-29.
[40] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015 arXiv preprint arXiv:1409.1556, Apr.
[41] A.L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. van Ginneken, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, P. Bilic, P.F. Christ, R.K.G. Do, M. Gollub, J. Golia-Pernicka, S. H. Heckers, W.R. Jarnagin, M.K. McHugo, S. Napel, E. Vorontsov, L. Maier-Hein, M.J. Cardoso, A Large Annotated Medical Image Dataset for the Development and Evaluation of Segmentation Algorithms, 2019 arXiv preprint arXiv:1902.09063, Feb.
[42] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, 2019 arXiv preprint arXiv: 1811.02629, Nov.
[43] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302. Jul.
[44] T. Heimann, B. van Ginneken, M.A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, et al., Comparison and evaluation of methods for liver segmentation from CT datasets, IEEE Trans. Med. Imag. 28 (8) (2009) 1251, 1265, Feb.
[45] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: 3rd International Conference on Learning Representations, 2015. San Diego, CA, USA, May 7-9.
[46] L. Yu, S. Wang, X. Li, C.W. Fu, P.A. Heng, Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation, in: Medical Image Computing and Computer Assisted Intervention, 2019, pp. 605–613. Shenzhen, China, Oct. 13-17.
[47] A. Rozantsev, M. Salzmann, Beyond sharing weights for deep domain adaptation, IEEE Trans. Pattern Anal. Mach. Intell. 41 (4) (2019) 801–814. Apr.
[48] J. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232. Venice, Italy, Oct. 22-29.
[49] W. Yuan, J. Wei, J. Wang, Q. Ma, T. Tasdizen, Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images, Med. Image Anal. 64 (2020), 101731, Aug.