



URO-GAN: An untrustworthy region optimization approach for adipose tissue segmentation based on adversarial learning

Kaifei Shen¹ · Hongyan Quan² · Jun Han³ · Min Wu¹

Accepted: 2 November 2021 / Published online: 12 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Automatic segmentation of adipose tissue from CT images is an essential module of medical assistant diagnosis. A large scale of abdominal cross-section CT images can be used to segment subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT) with deep learning method. However, the CT images still need to be professionally and accurately annotated to improve the segmentation quality. The paper proposes a semi-supervised segmentation network based on adversarial learning. The model is called URO-GAN and consists of two paths used to segment SAT and VAT, respectively. An SAT-to-VAT transmission mechanism is set up between these two paths, where several inverse-SAT excitation blocks are set to help the SAT segmentation network guide the VAT segmentation network. An untrustworthy region optimization mechanism is proposed to improve the segmentation quality and keep the adversarial learning stable. With the confidence map output from the discriminator network, an optimizer network is used to fix the error in the masks predicted by the segmentation network. The URO-GAN achieves good results by training with 84 annotated images and 3969 unannotated images. Experimental results demonstrate the effectiveness of our approach on the segmentation of adipose tissue in medical images.

Keywords CT image · Adversarial learning · Adipose tissue · Semantic segmentation · Semi-supervised learning

1 Introduction

Adipose tissue is an important component of the human body, mostly composed of subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT). They are closely related to human health, and their association with overweight, obesity, and cardiovascular disease have long been worldwide issues [1]. Studies have shown that the adipose tissue area in cross-sectional computed tomography (CT) images located at the abdominal region are indicative of the body's whole body fat profile [4]. So sufficiently accurate adipose tissue segmentation can assist

clinicians in diagnosing related diseases. Recently, several studies have shown a stronger relationship between the accumulation of VAT and adverse metabolic syndrome as well as inflammatory features than SAT [5, 6]. Therefore, in addition to the SAT segmentation, accurate and rapid VAT segmentation is also important. Based on the above mentioned background, we propose a semi-supervised deep learning approach for synchronized SAT and VAT segmentation based on adversarial learning.

In the first chapter, related works is described in Section 1.1. The contributions of this paper is provided in Section 1.2. Finally, We illustrate the organization of the paper in Section 1.3.

1.1 Related works

We illustrate the related work in this section. Section 1.1.1 illustrates the deep learning network, which is the foundation of our method. As adipose tissue segmentation is the research subject in this paper, its related work is shown in Section 1.1.2. Semi-supervised learning and adversarial learning are both features of our method. So we illustrate the segmentation method based on semi-supervised learning in Section 1.1.3 and discuss adversarial learning in Section 1.1.4.

✉ Hongyan Quan
hyquan@cs.ecnu.edu.cn

¹ Department of Software Science and Technology, Software Engineering Institute, East China Normal University, North Zhongshan Road Campus: 3663 N. Zhongshan Rd., Shanghai 200062, China

² School of Computer Science and Technology, East China Normal University, North Zhongshan Road Campus: 3663 N. Zhongshan Rd., Shanghai 200062, China

³ Department of General Surgery, Zhongshan Hospital, Fudan University, Shanghai, China

1.1.1 Deep learning networks

In the last decade, many deep-learning-based approaches are applied to implement semantic segmentation. The first and foremost model is FCN [7], which opened up semantic segmentation research. UNet [8], as an improvement of FCN, used the encoder-decoder structure with skip-connections to supplement the information lost in the downsampling process. The UNet was applied to the medical image field from the beginning with good results. Jégou et al. [9] used UNet as the basic structure, proceeding from the backbone of the model and using DenseNet to deepen the depth of the model to get better semantic segmentation. DenseNet is an efficient network structure that fully connects the nodes in the model for the purpose of improving the gradient flow. Cui et al. [10] used attention gates between the encoder and decoder of UNet to improve the information lost during the downsampling phase. The gates focus on the more important areas of the transmitted information. The dual attention mechanism proposed by DANet [11] adaptively integrated local features and global dependencies. Two modules can be attached to the segmentation model to model the semantic interdependencies in the spatial and channel dimensions. SENet [12] proposed a more lightweight block called 'squeeze & excite' (SE) that automatically obtained the importance of each feature channel by deep learning. The SE then re-calibrated the original of channel dimension by element-wise multiplication, weighing them to the previous features. In other words, SENet enhanced useful features and suppressed features that were of little use to the current task depending on their importance. Roy et al. [13] extended this approach to the spatial dimension and proposed a module that used SE in both space and channel. Deep learning methods have played an important role in various fields [2, 3]. Deep learning methods have played a significant role in the field of medical image segmentations. Many researchers have used various state-of-the-art neural network methods to segment certain human organs [14–16].

Deep learning is the basis of our approach. We use DenseNet as backbone of the proposed adversarial learning method. We also interspersed the model with some attention mechanisms and SE modules.

1.1.2 Adipose tissue segmentation

Some researchers have worked on segmenting adipose tissue from CT images or nuclear magnetic images (MRIs) [17–21]. There exist two main approaches for adipose tissue segmentations: traditional image processing methods [20, 21] and machine learning methods [17–19]. In traditional methods, researchers usually used thresholding as the basic method to obtain segmentation [20, 21].

However, CT images inevitably suffer from image quality problems such as artifacts, noise, and low contrast. The complex human organ structure also affects the accuracy of traditional methods. In contrast, machine learning methods can increase segmentation accuracy by learning a large amount of training data [17–19].

In a multi-category segmentation study with only one model [17, 19], we found that SAT segmentation accuracy was significantly better than VAT segmentation accuracy. Based on this phenomenon, we design a parallel structure with the aim of allowing the easier SAT task to guide the VAT task to some extent. We also design SAT-to-VAT transmission mechanism to implement the mentioned guidance. The proposed dual-task parallel path framework and the transmission mechanism are proved to enhance the VAT segmentation accuracy.

1.1.3 Semi-supervised segmentation

The labeling of medical images is very time-consuming and demands a tremendous level of expertise. All the models mentioned in the previous paragraphs were required to be implemented under full-supervised training with many labeled data. To ease the burden of image labeling work, many researchers have used semi-supervised training strategies for semantic segmentation tasks [22–24, 27–31], where only a small or limited number of annotated images can be used for training, in contrast to the large amount of unannotated data that can be used. For example, Luc et al. [23] trained an FCN-like segmentation model using a semi-supervised approach, which updated the model parameters using the probability maps predicted from the unannotated data directly as new ground truth. Goodfellow et al. [22], on the other hand, optimized this approach by restricting only the more reasonable predictions to be used as new annotated data for training the model. The drawback of mentioned semi-supervised methods is obvious, they cannot guarantee that the results of unannotated data can be good enough to be used as ground truth, which puts a high demand on the distribution and quality of unannotated data. Furthermore, a series of semi-supervised learning methods using adversarial learning have been proposed [24, 27–31]. Such methods enhance the usefulness of unlabeled data by training two networks adversarial to each other, and related work will be presented in detail in the next subsection. In adversarial training, the poor quality of data tends to make training unstable, thus limits the improvement of model performance from unannotated data.

To make the best use of unannotated data, we design a semi-supervised adversarial learning method to segment adipose tissues. The experiments demonstrate that unlabeled data can improve SAT and VAT segmentation performance.

1.1.4 Adversarial learning

The emergence of the generative adversarial network (GAN) led to the boom of unsupervised methods [25], whose core is the adversarial training of generator network and discriminator network. Fei et al. [26] pioneered the attempt to apply adversarial training to semantic segmentation. In this work, the generator network was implemented as a segmentation network. A discriminator network should distinguish between the generated segmentation and ground truth. Miller et al. [27] extended this model to semi-supervised training and proposed a new implementation idea for semi-supervised training. In this model, the discriminator network discriminated the segmentation of labeled and unlabeled images. The segmentation network is trained so that the predictions generated from the annotated and unannotated data are realistic enough to fool the discriminator network. The same type of semi-supervised approach was also used in the medical field [28]. Huang et al. [29] implemented another semi-supervised training method. In this method, the model further improved the output of the discriminator network to predict areas of high confidence on input images. The output was called confidence map and was used for semi-supervised training with unlabeled images. According to the confidence map, only the segmentation regions with high confidence can be further exploited to guide the learning process, forming pseudo-labeling learning. Similar approaches were extended to industrial and medical fields [24, 30, 31]. Lee et al. [30] replaced the basic network with DenseNet so that the model has fewer parameters and better performance, which was used for pavement crack detection. Mao et al. [31] applied the adversarial learning model to the segmentation of the left ventricle. The confidence map was operated the binarization before being used to supervise the unlabeled images.

Confidence maps might be affected by scattering noise of unannotated data [28]. Unstable confidence maps might influence the whole learning process, leading to undesirable segmentation results. The methods mentioned above all used the confidence map for semi-supervised training, prone to error amplification, and did not address the fundamental problem of high requirements for unlabeled images. The adversarial learning models are usually challenging to train and not stable enough to ensure the quality of the confidence maps.

This paper proposes an optimizer network after the adversarial training network to utilize the confidence map more effectively and reliably. Proposed method does not use the confidence map directly to supervise the unannotated data. Instead, by utilizing the third network, the optimizer network effectively uses the information of the confidence map to improve the predicted results of the segmentation network. The method is called untrustworthy

region optimization mechanism (UROM). This approach improves the robustness of the model by counteracting the effect of instability caused by unlabeled data. More importantly, with only a small amount of professionally labeled data, the proposed model can get almost as good results as other models trained with a large amount of labeled data.

1.2 Contributions

Our main contributions in this paper can be summarized as follows:

1. A semi-supervised network architecture based on the adversarial training is designed. The proposed optimizer network can utilize the confidence map to optimize the result of the segmentation network. The proposed approach can effectively prevent the emergence of unstable adversarial training.
2. A dual-task parallel path framework for adipose tissue segmentation is proposed, where an SAT-to-VAT transmission mechanism improves the VAT segmentation network performance.
3. An untrustworthy region optimization mechanism (UROM) is proposed, in which the optimizer network is connected to two adversarial training networks through the untrustworthy region excitation (URE) blocks. The mechanism can help the optimizer network focus on optimizing the unreliable region of the segmentation network's output.
4. The proposed model is trained and tested on the abdomen adipose tissue dataset. Our method achieve the best segmentation accuracy compared to other abdomen adipose segmentation methods.

1.3 Organization

The paper is divided into five chapters, with Chapter 1 focusing on the motivation, related work and contributions. Chapter 2 describes our proposed approach in detail. Chapter 3 explains the theory of the proposed method in mathematical terms, including the loss function and the training strategy used to train the model. Chapter 4 is the experimental part where we evaluate and discuss the performance of the proposed model.

2 Method

The proposed framework for the semi-supervised semantic segmentation of SAT and VAT is illustrated in Fig. 1. The model contains two parallel paths, which are used to segment SAT and VAT, respectively. The SAT path

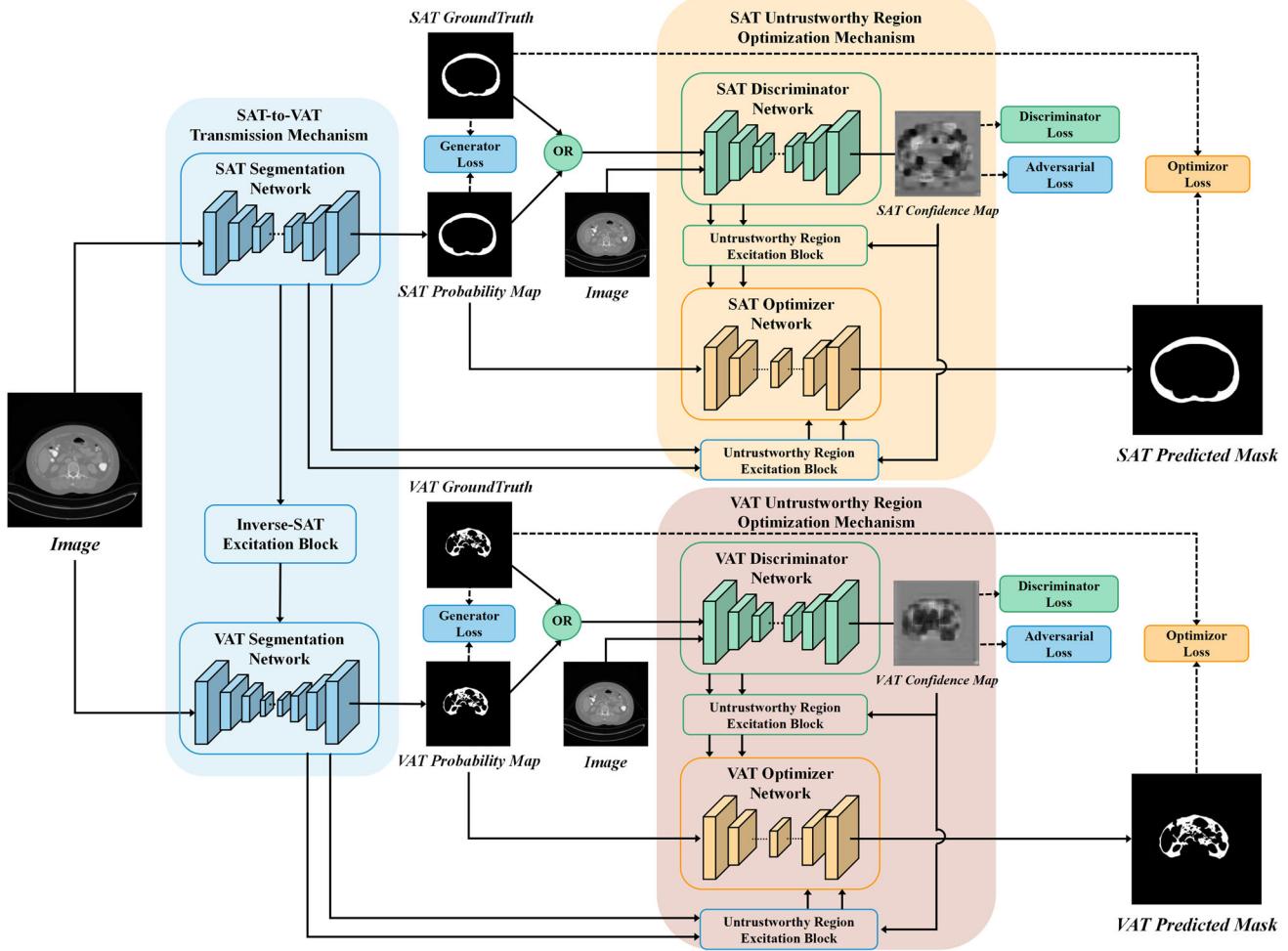


Fig. 1 Proposed dual-task parallel path framework for segmenting SAT and VAT concurrently. The SAT-to-VAT transmission mechanism consists of the SAT segmentation network, the VAT segmentation network and the inverse-SAT excitation blocks. The two paths have their

own UROM. Each UROM contains a discriminator network, an optimizer network and URE blocks. Finally, the two branches output the prediction results for SAT and VAT respectively

is composed of three networks: the SAT segmentation network, the SAT discriminator network and the SAT optimizer network. Similarly, the VAT path consists of the VAT segmentation network, the VAT discriminator network and the VAT optimizer network. These two paths are connected through some inverse-SAT excitation blocks between the SAT segmentation network and VAT segmentation network. The inverse-SAT excitation blocks and two segmentation networks form the SAT-to-VAT transmission mechanism, which will be described in details in Section 2.1. The disctiminator network, the optimizer network and the untrustwrothy region excitation (URE) blocks form the untrustwrothy region optimization mechanism (UROM). Each of the two paths has its own UROM. More details about the UROM will be discussed in Section 2.2.

For each branch, there exist three stages:

Firstly, we input the abdomen's cross-sectional CT image to the segmentation network, which will output the probability map of the target tissue. The closer the pixel value of the probability map is to 1, the more likely the segmentation network considers the matching area as the SAT or VAT.

Secondly, the discriminator network takes the probability map or ground truth labels as input, combines the original CT image's information, and yields the confidence map as output. The closer the pixel value is to 1, the more the discriminator considers the result of the corresponding position in the probability map to be reliable. On the contrary, the pixel with the value closer to 0 means that the discriminator network considers the predicted result at the corresponding position to be untrustwrothy. The goal of the segmentation network is to make the discriminator network regard the probability map as plausible, which means each

pixel value in the confidence map is close to 1. In contrast, the discriminator network aims to discriminate between the ground truth label and the probability map. If the input of the discriminator network is ground truth label, each pixel value of the confidence map has a tendency to converge to 1. If the input is a probability map, each pixel value tends to become 0.

Thirdly, the probability map and confidence map will be adopted together to the proposed optimizer network. The URE blocks receive feature maps from the segmentation and discriminator network, then transmit them to the corresponding layers of the optimizer network. Simultaneously, the confidence map is passed into the URE blocks, aiming to make features focus more on the untrustworthy region of the probability map.

Finally, the optimizer model outputs predicted masks.

2.1 SAT-to-VAT transmission mechanism

In this work, we notice that the task of segmenting SAT is simpler than that of segmenting VAT. Most of the subcutaneous fat has smooth edges, well-filled areas, prominent image features that are not difficult to distinguish, while the visceral fat is very detailed and has sophisticated image features. The size of the visceral fat area vary significantly between patients, and even for the same patient, the visceral fat size of different cross-sectional CT images can vary dramatically.

The proposed SAT-to-VAT transmission mechanism is designed to enable the SAT segmentation task to gain knowledge from the VAT segmentation task. To demonstrate knowledge transfer, we adopt dual-task parallel path framework in the overall model design. The essence of the proposed mechanism is the interaction between the SAT model and the VAT model, which is often used in few-shot learning [32, 33]. Few-shot learning is a machine learning technique that seeks to address the situation where an existing model needs to be extended to an unknown semantic class with a few samples at a fast rate [34–36]. A lightweight, strongly connected interaction between models is provided in [32], using ‘squeeze & excite’ modules [12, 13] to improve gradient flow and simplify network training with little increase in model complexity and computing time. We take examples from [32] and design a module called inverse-SAT excitation (ISE) blocks to transmit the activated feature map.

The SAT-to-VAT transmission mechanism is implemented by the ISE blocks between two task paths’ segmentation networks. The entire mechanism is shown in Fig. 2. To express the role of the ISE blocks more clearly, we will first illustrate the structure of two segmentation networks. We choose DenseNet[37] as their backbone network. Due to the dense connection method, DenseNet improves

gradients’ backpropagation, making the network easier to train. Since each layer can go straight to the last error signal, an implicit “deep supervision” is achieved [38]. More detail about SAT and VAT segmentation networks will be shown in Sections 2.1.1 and 2.1.2. The ISE blocks will be described in Section 2.1.3.

2.1.1 SAT segmentation network

In the specific implementation of the SAT segmentation network, the encoder uses two DenseNet-BC structures [37]. The model uses a 7×7 convolution of the beginning process before setting the first DenseBlock. The DenseBlock contains four dense layers, each of which uses a bottleneck layer to reduce the computational effort. The bottleneck layer consists of two Conv with stride=1. The obtained feature maps are densely connected with the feature maps obtained in each previous layer. The DenseBlock then passes through a transition layer with a compression rate of 0.6. After that, the DenseBlock sets a 1×1 Conv with stride=2 to reduce the feature map size. Each Conv in the DenseBlock is preceded by a batch normalization layer (BN) and a leaky rectified linear unit (LeakyReLU). Next, the feature map is fed into a second DenseBlock, which contains six bottleneck layers. The second DenseBlock also uses a transition layer with a compression rate of 0.6.

A channel attention module (CAM) [11], which applies the self-attention mechanism to obtain the correlations of any two feature maps, is set up between the encoding and decoding phases of the SAT segmentation network. The CAM emphasizes the feature maps of interdependent channels using the dependencies between different channels.

As shown in Fig. 3, consider an original feature map $A \in R^{H \times W \times C}$. First, the CAM deforms A into two features: reshaping and transposing A to $R^{C \times HW}$ and reshaping A to $R^{HW \times C}$. Then, The CAM performs a matrix multiplication between the feature map with the shape of $R^{C \times HW}$ and the feature map with the shape of $R^{HW \times C}$, followed immediately by a softmax layer to obtain the channel attention map X with the shape of $R^{C \times C}$.

$$X_{(i,j)} = \frac{\exp(A_{(i)} \cdot A_{(j)})}{\sum_{i=1}^C \exp(A_{(i)} \cdot A_{(j)})} \quad (1)$$

where $X_{(i,j)}$ denotes the channel attention map. The value of $X_{(i,j)}$ will become larger as the connection between the i -th channel and the j -th channel. After obtaining the channel attention map, we utilize a matrix multiplication between the reshaped A and the channel attention map to get a feature map with the shape of $R^{H \times W \times C}$. Finally, we do an element-wise sum operation between the original features A and the last feature map with the shape of $R^{H \times W \times C}$ to obtain the output.

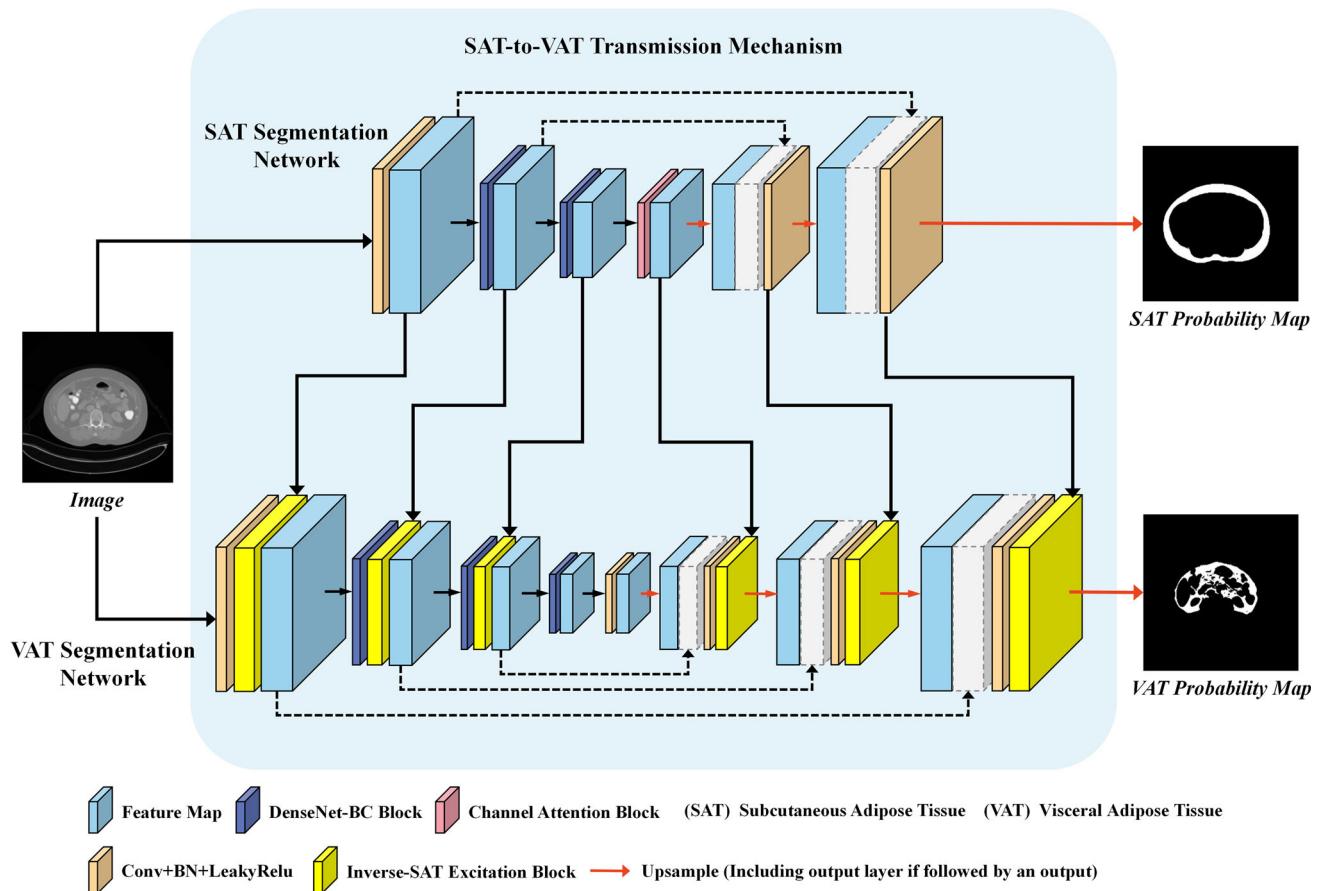
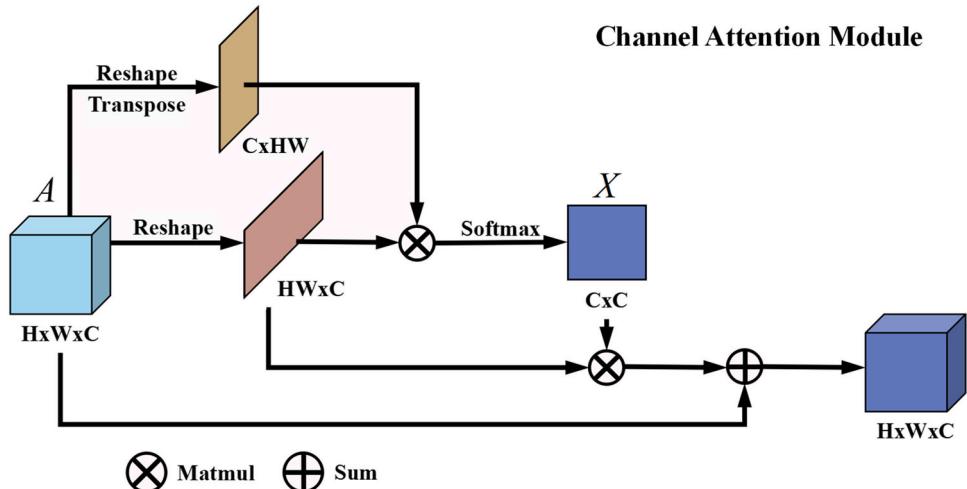


Fig. 2 Structure of the proposed SAT-to-VAT transmission mechanism. The inverse-SAT excitation block enables the VAT segmentation network to gain knowledge from the SAT segmentation network through passages on each layer

After the CAM is applied, the SAT segmentation network enters the decoder stage, which upsamples the feature maps and gradually recovers to the original size at a 2x per layer zoom rate. We set a 3×3 Conv after each upsampling

phase. Each Conv is followed by BN and LeakyReLU. The decoder uses skip-connections to obtain the feature maps from the encoder to enhance the semantic information of low levels. In the output layer, the network sets a 5×5

Fig. 3 Channel Attention Module



Conv+BN+LeakyReLU and then uses a 1×1 Conv to reduce the number of feature maps to 1. Finally, a sigmoid layer is used to obtain the SAT probability map.

2.1.2 VAT segmentation network

The VAT segmentation network is illustrated in Fig. 2. Since it is more difficult to segment the VAT, the model needs a broader and deeper network. The encoder of the VAT segmentation network uses three DenseBlock, and the number of dense layers they contain is 4, 6, 12, respectively. The broader and deeper model perceives much more semantic information and expands the perceptual field. As DenseNet accomplishes short-circuit connectivity by concatenating features, the network can effectively combine both high-level and low-level features to accommodate targets of different sizes, coping with the fact that fat sizes can vary significantly between CT slices of distinct patients. To avoid too much extra time cost overall, we set a 3×3 Conv between encoder and decoder instead of CAM in the SAT segmentation network. Since there exists an extra DenseBlock, the decoder correspondingly contains an extra layer to upsample the feature map. Like the SAT segmentation network, the VAT segmentation network sets BN and LeakyRelu after each Conv and includes similar skip-connections and the same output layer.

2.1.3 Inverse-SAT excitation block

Considering that SAT and VAT are mutually independent regions, it is necessary for the VAT segmentation network to pay more attention to the regions beyond SAT, helping the

network learn VAT features better. The VAT segmentation network sets the ISE blocks at the end of each layer to activate the region other than SAT. The structure is shown in Fig. 4. We assume that the feature map to be activated in the VAT segmentation network is $F_{VAT} \in R^{H \times W \times C'}$, while $F_{SAT} \in R^{H \times W \times C}$ is the feature map of the corresponding size in the SAT segmentation network, and is used to segment SAT. The ISE block first inverses F_{SAT} and squeezes it to only one feature by using a 1×1 Conv with stride=1. Then the ISE block utilizes a sigmoid layer to obtain the squeezed feature. To match the number C' of F_{VAT} to be activated, the block copies the squeezed feature and stacks it C' times to get $E \in R^{H \times W \times C'}$. An element-wise multiplication between E and F_{VAT} is applied to achieve activation of all regions outside the SAT partition.

2.2 Untrustworthy region optimization mechanism

The goal of the untrustworthy region optimization mechanism (UROM) is to efficiently utilize the confidence map generated by the discriminator network for robustly improving the segmentation quality of the segmentation network. The mechanism is composed of both the discriminator network and the optimizer network. The optimizer network receives information from the discriminator network encoder and the segmentation network decoder.

Since the proposed model is a dual-task parallel structure, there exist two UROMs: the SAT-UROM and the VAT-UROM. Each UROM contains both an SAT/VAT discriminator network and an SAT/VAT optimizer network. The UROM follows the corresponding SAT/VAT segmentation

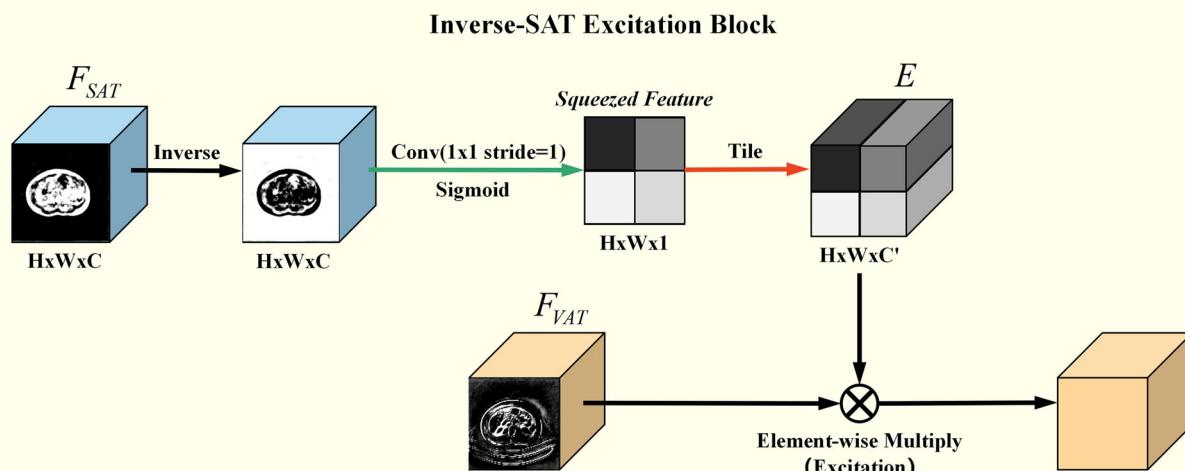


Fig. 4 Proposed inverse-SAT excitation block, used to activate regions that are weakly responsive to SAT segmentation on features at each layer of the VAT segmentation network

network and serves to segment SAT or VAT, respectively. In this section, we describe the specific structure of the discriminator network and optimizer network in both parallel paths. There exist some subtle differences between the SAT-UROM and VAT-UROM, mainly in the optimizer network. Therefore, we first describe the discriminator network of SAT-UROM and VAT-UROM together in Section 2.2.1, as their model structures in this part are identical. Then we show the structure of the SAT optimizer network in Section 2.2.2. The architecture of the VAT optimizer network is illustrated in Section 2.2.3 for more complex VAT optimization scenarios. Finally, we introduce the untrustworthy region excitation (URE) block used in both UROM in Section 2.2.4.

2.2.1 SAT/VAT Discriminator Network

The SAT discriminator network contains the same structure as the VAT discriminator network. Both discriminator networks are shown in Figs. 5 or 6. The discriminator network, the segmentation network, the probability map, and the confidence map mentioned in this section refer to the model elements in two task paths.

The original CT image and the probability map output from the segmentation network is first fed together to the discriminator network. The probability map is utilized as attention features to make the discriminator network focus on those segmented regions predicted by the segmentation network. In more details, the discriminator network uses a 7×7 Conv with stride=2 to expand the number of original image features to 32, which will be taken as the initial features into the network. In this case, a 3×3 Conv with stride=1 is used for both the original CT image features and the probability map features. The number of CT image features are compressed to 16 to reduce the computational effort, while the number of probability map features are extended to 16 and followed by a sigmoid function. These two processed features are bilinearly interpolated to match the size, and then element-wise multiplication is applied to obtain the attentive rated map. Finally, the attentive rated map is added to the initial features. Then we obtain the fusion of two input features.

The discriminator network is composed of an encoder-decoder architecture. In the encoding phase, the network applies three 3×3 Conv with stride=2 to downsample the mixed features, followed by one 3×3 Conv with stride=1. In

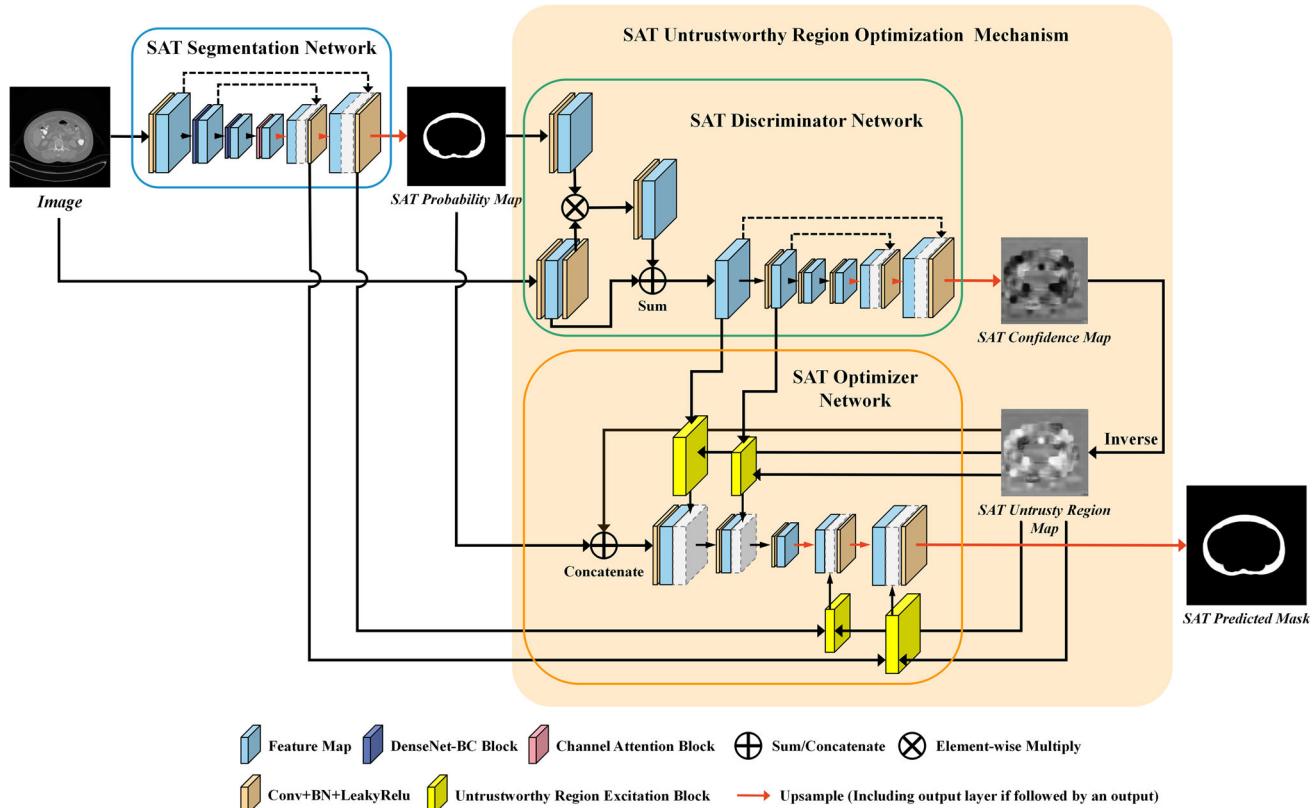


Fig. 5 Structure of the proposed task path for segmenting SAT, consisting of an SAT segmentation network that outputs an SAT probability map, an SAT discriminator network that outputs an SAT confidence

map, and an SAT optimizer network that uses the output of the former two networks to output an SAT predicted mask

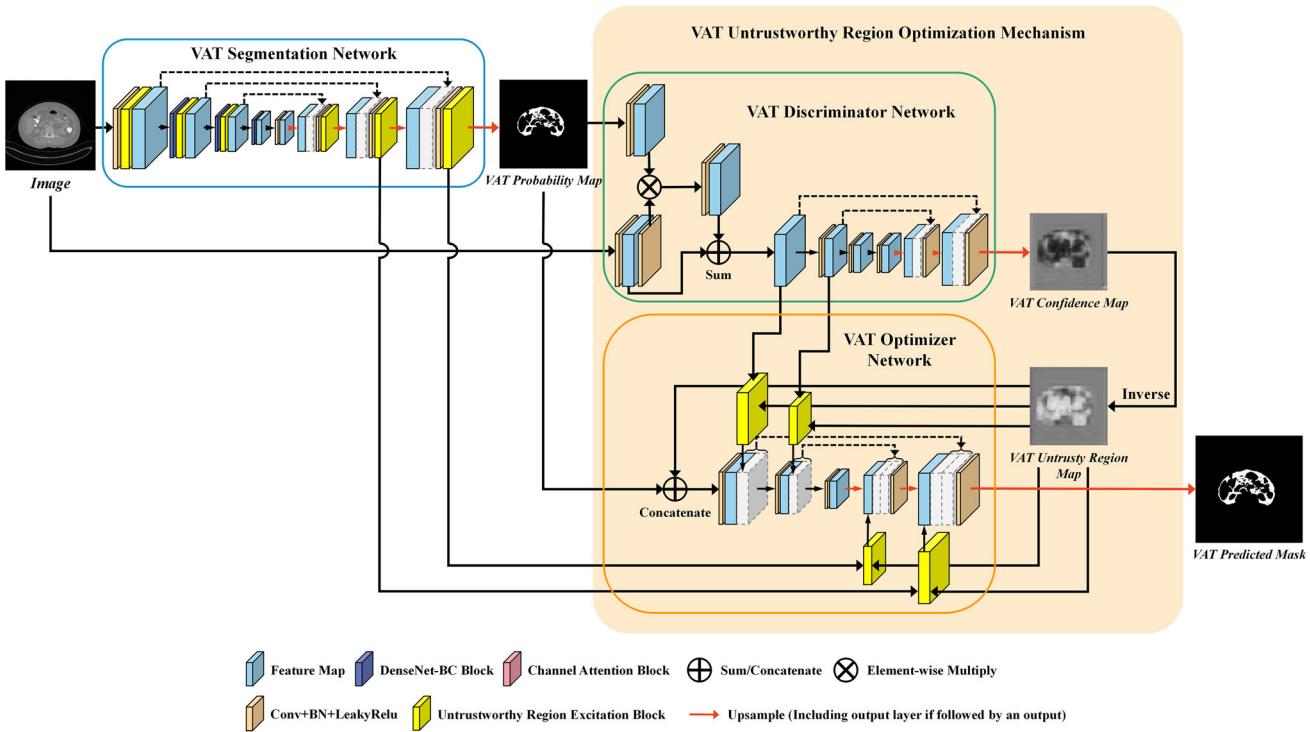


Fig. 6 Structure of the proposed task path for segmenting VAT, consisting of a VAT segmentation network that outputs a VAT probability map, a VAT discriminator network that outputs a VAT confidence map, and a VAT optimizer network that uses the output of the former two networks to output a VAT predicted mask

the decoding phase, three bilinear interpolations are used to complete the upsampling process. After each upsampling, a skip-connection is applied, where the concatenation is implemented with the corresponding feature map of the downsampling process. A 3x3 Conv with stride=1 is then applied to each feature fused by the skip-connection. In the output phase, the discriminator network sets a 3x3 Conv to output the confidence map. The network does not end up using sigmoid layer activation because [39] pointed out that the sigmoid function used by the traditional GAN discriminator network saturates very rapidly, causing the function to quickly ignore the distance of the sample to the decision boundary even for minimal data points. In addition, the discriminator network sets a LeakyRelu after each Conv.

In the final confidence map obtained, a pixel closer to 1 means that the more likely the discriminator thinks the pixel is from the ground truth label. On the contrary, a pixel closer to 0 means that the more likely the discriminator thinks the pixel is from the segmentation network, which is untrustworthy.

2.2.2 SAT optimizer network

The goal of the SAT optimizer network is to optimize the predicted mask generated by the SAT segmentation network. In real-world problems where we optimize some flawed problems, one of the most intuitive and efficient

approaches is focusing on those parts that are considered unreliable. The SAT optimizer network and the VAT optimizer network are proposed based on such intuition. The confidence map is an interpretative feature that points out the unreliable part of the probability map.

The network first inverses the SAT confidence map to obtain the SAT untrustly region map. The closer the pixel is to 1, the less reliable it is. As the last layer of the discriminator network is not a sigmoid function, it is also necessary to clip SAT untrustly region map so that its value range is limited between 0 and 1. The structure of the SAT optimizer network is shown in the eponymous part of Fig. 5. A concatenation is first applied to the SAT probability map and SAT untrustly region map. The SAT optimizer network utilizes their fused features as inputs and then is composed of an encoder-decoder structure like two previous networks.

To achieve better optimization results, we think it necessary for the optimizer network to parse information from as many sources as possible. In terms of this aspect, the SAT optimizer network has the similar input with the discriminator network because both of them take the probability map as input. So we consider that the encoder flow of the discriminator network can guide the optimizer network. The encoder of the SAT optimizer network receives the semantic information sent from the SAT discriminator network's encoder. Correspondingly, since the optimizer network outputs a predicted mask that

is the same as the output of the segmentation network, the decoder of the SAT optimizer network receives the semantic information passed by the decoder of the SAT segmentation network. All these external input feature maps will first enter the URE block, whose specific principle is explained in Section 2.2.4. Through URE block, the transferred feature maps can focus on the unreliable area indicated in the untrustworthy region map.

The SAT optimizer network first sets up a 7×7 Conv to obtain the initial feature of input. Then we apply a concatenation between the obtained feature map and the output from the URE block. After that, a 3×3 Conv with stride=2 is utilized for further downsampling. The resulted feature map is concatenated with the URE block's output containing the discriminator network information, followed by another 3×3 Conv with stride=2 to complete the encoding process. In the decoding stage, the SAT optimizer network first upsamples the feature map through bilinear interpolation. The scaled feature map is then concatenated with the output of a URE block, which processes the network feature map of the corresponding size in the SAT segmentation network decoder. Afterward, the SAT optimizer network applies a 3×3 Conv with stride=1. Further, the feature map is upsampled and repeats the first decoding process until the feature map is the same size as the initial CT image. The SAT optimizer network uses a 3×3 and a 1×1 Conv to obtain the final predicted mask.

2.2.3 VAT optimizer network

The VAT optimizer network belongs to the VAT-UROM, which improves the predicted segmentation results of the VAT segmentation network for VAT. Because of the feature complexity of VAT, the decoder of the VAT optimizer network adds skip-connections, which can integrate the low-level information in encoding phase. Thus the model can locate the part that needs to be enhanced more accurately. The structure is shown in the eponymous part of Fig. 6.

In the specific design, similar to the SAT optimizer network, a concatenation is first applied between the VAT probability map and the VAT untrustworthy region map. Their fused features are used as inputs. In the beginning, a 7×7 Conv is set to obtain the initial feature of input, which is then concatenated with the feature map output from the URE block. The VAT optimizer network sets up a 3×3 Conv with stride=2 and concatenates a feature map obtained from the URE block. Consequently, a 3×3 Conv with stride=2 is used to complete the encoding. At this point, the URE block of the encoder collects the attentive rated feature maps of the VAT discriminator network. Next, the VAT optimizer network upsamples the feature map using bilinear interpolation and then sums the feature map

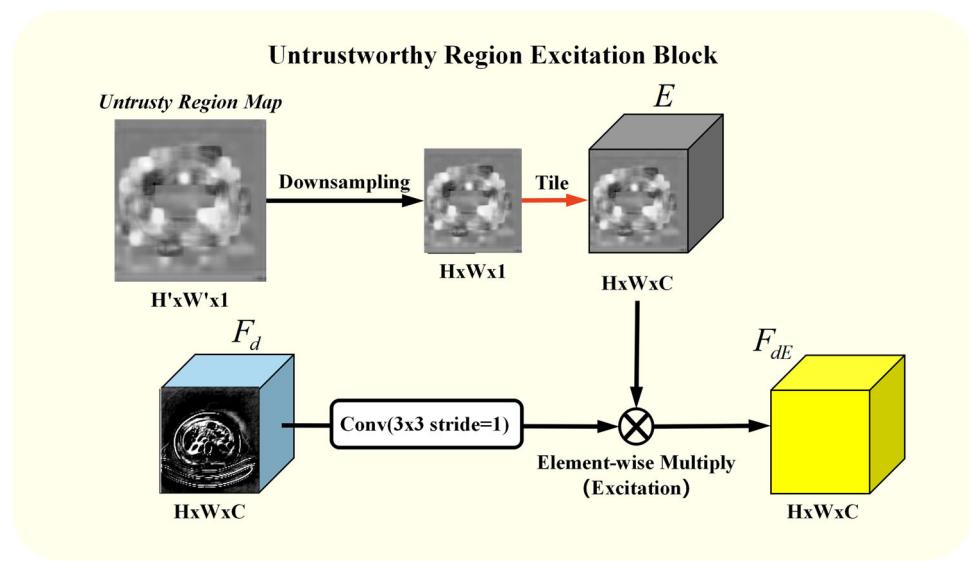
with the URE block output. The URE block processes the information of the corresponding layer from the VAT segmentation network and uses the VAT untrustworthy region map to emphasize the unreliable regions. The model then performs a concatenation between the summed feature map and the low-level features passed by the skip-connection. Thus, the model continues to upsample the feature map and repeats the process until the feature map is the same size as the initial CT image. Finally, the model utilizes a 3×3 and a 1×1 Conv to output the final predicted mask.

2.2.4 Untrustworthy region excitation block

The role of the URE block is to establish the passages between two models. From the perspective of the global information flow, the output of the optimization network and the segmentation network are both predicted masks of SAT or VAT, so the feature stream near the output of their network are similar. We assume that the information transmitted from the decoder of the segmentation network could help the optimizer network construct segmentation predictions better. The optimizer network aims to optimize the segmentation network results, so the transmitted information also informs the optimizer network how to avoid inaccurate predictions. On the other hand, one of the major tasks of the discriminator network's encoder is to compare the difference between the probability map and the original map to output the confidence map. So passing the discrimination process information to the optimizer network allows the optimizer network to locate and process the discrepancy accurately.

The URE block shown in Fig. 7 is similar to the ISE block in Section 2.1. The URE block directly takes the untrustworthy region map as a squeezed feature and then applies excitation to each layer of the segmentation network and discriminator network. Then the layers of the optimizer network receive these activated preceding feature maps. We suppose $F_d \in R^{H \times W \times C}$ to be the output feature map of some layer in the segmentation network or discriminator network, which is the feature to be activated for this process. The untrustworthy region excitation (URE) block first downsamples the untrustworthy region map for matching the feature size and obtains a new untrustworthy region map $\in R^{H \times W \times 1}$. The untrustworthy region map is duplicated in C copies and stacked to obtain $E \in R^{H \times W \times C}$. A 3×3 Conv is set on F_d . The URE block then applies element-wise multiplication, which is an excitation operation, between E and F_d . The final obtained $F_{dE} \in R^{H \times W \times C}$ is the activated feature map. The F_{dE} in the SAT optimizer network and the encoder of the VAT optimizer network is concatenated to the target feature map. In contrast, the F_{dE} obtained in the decoder of the VAT optimizer network is summed with the target feature map.

Fig. 7 Proposed untrustworthy region excitation block, used to activate areas of the input feature map that are considered untrustworthy by the discriminator work



3 Theory

We illustrate the theory of the proposed method in this section. The URO-GAN is trained by minimizing the loss functions listed in Section 3.1. To keep the adversarial learning stable, we propose a training strategy with three steps, as shown in Section 3.2.

3.1 Loss functions

The parallel paths of the model segment SAT and VAT separately, and their loss functions are synonymous. When the training model segments SAT, the ground truth label is annotated with the SAT region. When the training model segments VAT, the ground truth label is annotated with the VAT region. Therefore, the segmentation network, discriminator network, and optimizer network mentioned in this chapter represent the three networks of each SAT and VAT at the same time. We use $S(\cdot)$ to refer to the SAT segmentation network or VAT segmentation network. $S(X_a)$ represents the one-hot encoded probability map output from the segmentation network when the input is an annotated sample. $D(\cdot)$ denotes the SAT discriminator network or VAT discriminator network. $D(X_a, S(X_a))$ represents the confidence map output from the discriminator network when the inputs of the discriminator network are an annotated sample and a probability map. Similarly, X_u represents the unannotated sample. $OPT(\cdot)$ denotes the one-hot encoded segmentation prediction output by the SAT optimizer network or the VAT optimizer network, which is also the final result of the whole model.

Segmentation network The segmentation network is trained by minimizing the following loss function.

$$L_S = L_{seg} + \lambda_{adv} L_{adv} \quad (2)$$

where L_{seg} and L_{adv} represent the supervised segmentation loss and adversarial loss. λ_{adv} refers to the weight of the adversarial learning and is set to 0.02.

L_{seg} indicates whether the probability map output by the segmentation network is close to the ground truth label, which is defined as follows.

$$L_{seg}(X_a, Y_a; \theta_S) = L_{bce}(S(X_a), Y_a) + L_{dice}(S(X_a), Y_a) \quad (3)$$

where X_a denotes annotated samples, Y_a denotes corresponding ground truth labels, θ_S denotes the model parameters of the segmentation network. L_{bce} and L_{dice} are binary cross-entropy loss and dice cross-entropy loss, and they are defined in (4), (5).

$$\begin{aligned} L_{bce}(\hat{Y}, Y) = & - \sum_{h,w} \sum_{c \in C} Y^{(h,w,c)} \log(\hat{Y}^{(h,w,c)}) \\ & + \sum_{h,w} \sum_{c \in C} (1 - Y^{(h,w,c)}) \log(1 - \hat{Y}^{(h,w,c)}) \end{aligned} \quad (4)$$

$$L_{dice}(\hat{Y}, Y) = 1 - 2 \frac{\sum_{h,w} \sum_{c \in C} (Y^{(h,w,c)} \hat{Y}^{(h,w,c)})}{\sum_{h,w} \sum_{c \in C} (Y^{(h,w,c)} + \hat{Y}^{(h,w,c)})} \quad (5)$$

where the size of images processed by the network is $h \times w \times c$, C is the number of output image categories. Binary cross-entropy loss and dice cross-entropy loss are both utilized for training the segmentation network in the supervised mode, making the segmentation prediction more accurate.

In the meantime, the adversarial loss L_{adv} is used in adversarial training to align the probability map output by the segmentation network closer to the ground truth label by continuously decreasing its value, which is mathematically defined as follows.

$$\begin{aligned} L_{adv}(X_a, X_u; \theta_S) = & \lambda_a \cdot L_{bce}(D(X_a, S(X_a)), 1) \\ & + \lambda_u \cdot L_{bce}(D(X_u, S(X_u)), 1) \end{aligned} \quad (6)$$

where X_a denotes annotated samples. X_u denotes unannotated samples. θ_S denotes the parameters of the segmentation network. L_{bce} denotes binary cross-entropy loss, λ_a, λ_u denote the correlation weight between annotated samples and unannotated samples, which determines how much contribution they make in L_{adv} .

For full-supervised training, we set $\lambda_a = 1, \lambda_u = 0$ to train the model with annotated samples. When we add unannotated samples for semi-supervised training, the model performs best when trained with $\lambda_a = \lambda_u = 0.5$. The adversarial loss L_{adv} aims to train the segmentation network to drive the discriminator network to output a confidence map that is all 1 when judging the generated probability map. In other words, the segmentation network is trained to generate the probability map as real as possible so that the discriminator network cannot distinguish whether the input image is the original ground truth label or the generated prediction map.

Discriminator network The discriminator network is trained by minimizing the following loss function.

$$\begin{aligned} L_D(X_a, X_u, Y_a; \theta_D) = & L_{bce}(D(X_a, Y_a), 1) \\ & + \lambda_a \cdot L_{bce}(D(X_a, S(X_a)), 0) \\ & + \lambda_u \cdot L_{bce}(D(X_u, S(X_u)), 0) \end{aligned} \quad (7)$$

where X_a denotes annotated samples, X_u denotes unannotated samples, Y_a refers to the ground truth labels corresponding to the annotated samples, and θ_D refers to the parameters of the discriminator network. L_{bce} is binary cross-entropy loss, λ_a, λ_u represent the weights of annotated samples and unannotated samples in the loss function.

The L_D is intended to train the discriminator network for adversarial training. The goal of the discriminator network is to distinguish whether the input is from the ground truth label image or the probability map generated by the segmentation network. When the input of the discriminator network is the ground truth label, each pixel of its confidence map should be close to 1. On the contrary, the discriminator network should output a confidence map with each pixel close to 0 when its input is a probability map. Similar to the segmentation network, the discriminator network can add unannotated samples to perform semi-supervised training.

Optimizer network The optimizer network takes the probability map and confidence map as inputs. The probability map is generated from the segmentation network, and the confidence map is obtained from the discriminator network. Several URE blocks are set between three networks. The optimizer network receives partial feature maps from the previous two network and is trained by minimizing the following loss function.

$$L_O(X_a, Y_a; \theta_O) = L_{bce}(O(X_a), Y_a) + L_{dice}(O(X_a), Y_a) \quad (8)$$

where X_a denotes annotated samples, Y_a denotes the ground truth labels corresponding to the annotated samples, θ_O refers to the parameters of the optimizer network. L_{bce}, L_{dice} are binary cross-entropy loss and dice cross-entropy loss, which are jointly used in the supervised training of the optimizer network so that its output predictions are closer to the ground truth label. $O(\cdot)$ is an abbreviation for $OPT(\cdot)$, which stands for the output mask of the optimizer network and is defined as follows.

$$O(X) = OPT(S(X), \tilde{D}(X, S(X)), F_S(X), F_D(X, S(X))) \quad (9)$$

where X refers to the original annotated samples, $\tilde{D}(\cdot)$ denotes the untrust region map obtained from the discriminator network through inverting the confidence map. $F_S(\cdot)$ represents the feature map of the segmentation network passed to the optimizer network, while $F_D(\cdot)$ represents those of the discriminator network passed to the optimizer network. In particular, the optimizer network is only trained with annotated samples for full-supervised training.

3.2 Training strategy

The proposed model uses a staged training strategy. The segmentation network and the discriminator network constitute a GAN. They are trained in a semi-supervised way with limited annotated data and relatively large amount of unannotated data. In contrast, the optimizer network uses only annotated data for training. The training process consists of following three steps.

Step1: To keep the GAN steady until the adversaries are balanced, we first trained the segmentation network and the discriminator network with annotated samples in a full-supervised manner. In this way, authoritative data prevent the model from receiving unlabeled data prematurely, thus avoiding the influence of initial noise and poor predictions. Specifically, we train the GAN by minimizing (2) and (7), where $\lambda_a = 1, \lambda_u = 0$.

Step2: Additional unannotated data is added to train the GAN together with the original annotated data. If we take the SAT segmentation as an example, this process is shown in Fig. 8. We input both the unannotated image and the annotated image into the segmentation network to obtain two probability maps. The probability maps are then inputted together with the ground truth label into the discriminator network to acquire the confidence map. In step2, We train the GAN by minimizing (2) and (7), where $\lambda_a = \lambda_u = 0.5$.

Step3: After the first two steps, the segmentation network has been able to output a relatively accurate segmentation

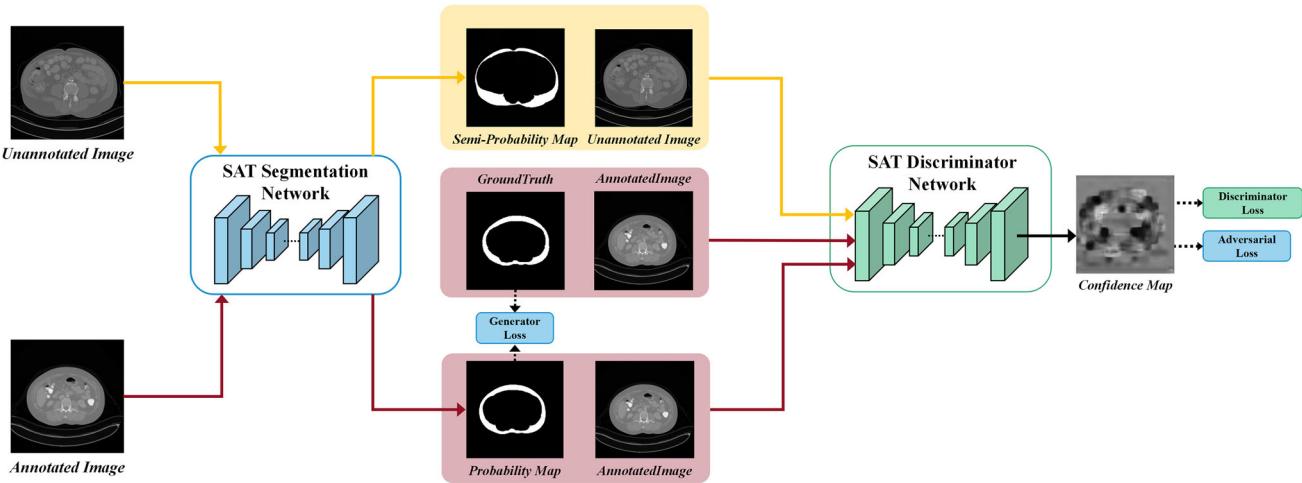


Fig. 8 Illustration of the semi-supervised training strategy when segmenting SAT in step2, where the red line shows the training process using annotated images, and the yellow one shows the process using unannotated images

prediction image, the discriminator network has been able to output a well-grounded confidence map. The confrontation between the segmentation network and the discriminator network has reached a stable equilibrium. In step3, the optimizer network is trained with annotated data by minimizing (8). The optimizer network improves the output of the segmentation network and obtains the final prediction results.

4 Experiment

This section presents our experiments to evaluate the proposed model. Section 4.1 provides the implementation details of our experiment. Section 4.2 describes the dataset used in evaluation. In Section 4.3, we compare the proposed method with various models quantitatively. Then we analyze the visualization results in Section 4.4. In Section 4.5, a series of ablation studies is conducted to investigate the effectiveness of main novel modules.

4.1 Implementation details

We use TensorFlow [40] to implement the proposed approach. The devices for training and testing include Intel(R) Core(TM) i7-9700KF CPU @3.6GHz and a single 2080S GPU with 8 GB memory.

The network weights are initialized by the random normal initializer whose standard deviation is 0.02. To train the model, we use the Adam optimizer [41] with different learning rates and exponential decay rates. The learning rate of the segmentation network is 1e-4. To keep the adversarial training stable, we set different learning rates among three networks. The learning rate of the discriminator network

is 3e-4, the exponential decay rate for the 1st-moment estimates is 0, the exponential decay rate for the 2nd-moment estimates is 0.9. The learning rate of the optimizer network is 1e-4 without any exponential decay.

We start training with step1 mentioned in Section 3.2, where we train the segmentation network and the discriminator network for 100 epochs with annotated data. We then proceed to step2 and start semi-supervised training of both networks for a continuous period of 20,000 iterations, equivalent to 238 epochs for the labeled data or five epochs for the unlabeled data. Notably, the annotated data and unannotated data are fed into the model at the same time. When they complete one epoch respectively, they are disrupted before beginning a new epoch again. Finally, we proceed to step3, where we train the optimizer network for 20,000 iterations under full supervision. We set the batch size as 2.

As ground truth labels are initially one-hot encoded, the probability map are one-hot encoded before being fed into the discriminator network to bring the data distribution closer to the ground truth label. This operation prevents the discriminator network from easily distinguishing between real and fake images.

The evaluation metrics applied in this paper are utilized to evaluate the quality of the generated segmentation mask, including the Jaccard Score [18, 28], Dice Coefficient [17, 18, 24, 28, 42].

4.2 Abdomen Adipose tissue dataset

We use two datasets, a public dataset for training and testing, and a clinical dataset for validating generalizability. After filtering and labeling processes, these two datasets form the abdomen adipose tissue dataset used in our experiments. The public dataset used in our experiment is filtered and

produced from CT Lymph Nodes (CTLN) [43], which is downloaded from TCIA [44]. This dataset consists of CT images of the mediastinum and abdomen in which radiologists mark lymph node positions at the National Institutes of Health, Clinical Center. The dataset contains a total of 110,103 CT images in DICOM format. The labels provided by the dataset are not the target ones needed for this paper. So we adopt CT images from the CTLN dataset without any provided annotations. The clinical dataset is collected by an experienced expert from a specialized institution that is ready to apply our proposed methodology in the future to aid clinical diagnosis. The clinical dataset is used to validate the generalization performance and application capability of our model.

In our experiment, an experienced expert has filtered 4089 clinically standard cross-sectional CT images of the abdomen from the CTLN dataset. The images capture the majority of realistic subcutaneous and visceral adipose tissue in the human body. Also, they contain similar noise features to the clinical CT images, ensuring that the model trained with this dataset can be practically implemented in the clinic. The expert further select 120 CT images from the filtered 4089 images to obtain full-supervised data, which represent different body types and adipose levels. A trained researcher label the SAT and VAT regions of the screened 120 images. These images are saved as JPEG images in one-hot encoding. 70% of the labeled data are used for network training, and the remaining 30% of the labeled data are used to evaluate the model in the testing phase. The other 3969 unlabeled data are used for semi-supervised training. In addition, a total of 50 expert-vetted standardized image sets are available in the clinical dataset, and these images are used for testing purposes only. The abdomen adipose tissue dataset used by the model is shown in Table 1.

As the CT images are stored in DICOM format, the DICOM images are converted to grayscale JPEG images and resized to 256x256 before training. All images are standardized so that their pixel values lie in the range of 0 to 1.

Table 1 Diagnosis for the abdomen adipose tissue dataset

Target	CTLN dataset			Clinic dataset
	Training (Annotated)	Training (Unannotated)	Testing	Testing
SAT	84	3969	36	50
VAT	84		36	50

The networks that segment SAT and VAT use the same unannotated images, and their annotated data comes from different annotations of the same image

To reduce the possibility of over-fitting and to improve the generalization of the network, the model apply some random transformations to the inputs and the corresponding labels during the training period, where the random variation of rotation, gamma, crop, zoom, noise are used.

4.3 Comparison

In this section, we compare the proposed URO-GAN with the state-of-the-art approaches. We divide the compared methods into two categories: full-supervised training and semi-supervised training.

In the comparison of full-supervised training, we compare the proposed approach with UNet [8], DANet [11], Dense-UNet [9], and the model proposed by Dabiri [18]. To further validate the effectiveness of the proposed optimizer network, we also compare the model's performance before and after adding the optimizer. As is shown in Table 2, “S” denotes the proposed SAT/VAT segmentation network, and “+D” represents the proposed SAT/VAT discriminator network. “+O” indicates that the model incorporates the SAT/VAT optimizer network to optimize the output of “S” further. The six compared full-supervised models are trained on labeled data from the abdomen adipose tissue dataset, which contains only 84 annotated images.

In the semi-supervised training, we compare the proposed method with semi-supervised methods that also used GAN in the medical field, including ASDNet [24], BUS-GAN [28] and DT-GAN [31]. The proposed method controls the use of semi-supervised data by changing λ_a , λ_u , which are only relevant for the segmentation network and the discriminator network. “Semi” after “D” indicates that the model is trained under semi-supervision. We likewise add “+O” to the semi-supervised training comparison to further validate the SAT/VAT optimizer network. The four compared semi-supervised models are each trained on the abdomen adipose tissue dataset, which consists of 84 annotated images and 3969 unannotated images.

Our proposed semi-supervised method (proposed S+D+Semi+O) achieves the best results as shown in Table 2. The average dice coefficient reaches 95.208%, which is significantly higher than other three methods in the semi-supervised method. The Jaccard score metric of the proposed method also outperformed the compared methods across the board, demonstrating the outstanding performance on the CTLN testing dataset. In addition, the proposed method (proposed S+D+O) performs best among the full-supervised methods. The average segmentation dice coefficient reaches 95.004%, demonstrating the excellent performance of the proposed URO-GAN on the small amount of annotated data. Notably, the model achieves better results using only the proposed segmentation network and the discriminator network (proposed S+D) compared

Table 2 SAT/VAT segmentation results evaluation on the CTLN test set

	SAT		VAT		Average	
	Jac.	Dic.	Jac.	Dic.	Jac.	Dic.
Full-supervised method						
UNet [8]	88.825	94.021	74.076	84.933	81.451	89.477
DANet [11]	93.948	96.866	84.254	91.351	89.101	94.109
Dense-UNet [9]	92.845	96.262	84.737	91.646	88.791	93.954
Dabiri [18]	90.442	94.939	78.219	87.540	84.331	91.240
Proposed S+D	94.054	96.925	86.461	92.684	90.258	94.805
Proposed S+D+O	94.346	97.082	86.887	92.926	90.617	95.004
Semi-supervised method						
ASDNet [24]	93.217	96.475	84.986	91.825	89.102	94.150
BUS-GAN [28]	93.727	96.743	77.159	86.931	85.443	91.837
DT-GAN [31]	93.881	95.932	85.163	92.078	89.522	94.005
Proposed S+D+Semi	94.080	96.929	86.489	92.689	90.285	94.809
Proposed S+D+Semi+O	94.238	97.015	87.700	93.400	90.969	95.208

to the rest of the methods, reflecting the good design of the adversarial learning component.

Considering the vertical comparison, we find that the method using the optimizer network (+O) is effective in improving the accuracy of the model. In full-supervised learning, the Jaccard score of the proposed S+D is 90.258%, while the score of the proposed S+D+O is up to 90.617%. The average dice coefficient increases from 94.805% to 95.004%. In semi-supervised learning, the average Jaccard score rises from 90.285% to 90.969%, and the average dice coefficient rises from 94.809% to 95.208%. It can be observed that the proposed optimizer network can improve the accuracy of adversarial learning in both data conditions.

On the other hand, the results of the proposed S+D and the proposed S+D+Semi are very close. The inclusion of unlabeled data improves the model to a limited extent without the optimizer network. In contrast, the proposed S+D+Semi+O achieves a Jaccard score of 87.700% on VAT and is significantly higher than the proposed S+D+O on VAT at 86.887%. The result demonstrates that the optimizer network stimulates the positive impact of unlabeled data on the model.

Table 3 shows the test results of the proposed method on the clinical dataset. Two test metrics of SAT are similar to the results on the CTLN test set. The result of VAT presents a slight decrease, with a Jaccard score of 87.423% and a dice coefficient down to 93.129%. Compared with other three adversarial learning methods, the proposed URO-GAN still reaches the highest value of both metrics, which demonstrates the good generalization performance of our approach.

Furthermore, we respectively use 1/8, 1/4, 1/2, and all of the annotated data in the abdomen adipose tissue dataset to validate the performance of URO-GAN on limited samples. The results are shown in Table 4. As can be seen, the proposed model obtains the best segmentation results in all four data cases. The dice coefficient of the model for SAT segmentation results reaches 95.391% for 1/8 data volume. Meanwhile, the method of Dabiri [18] is just close to 90%, the method of ASDNet [24] is only 85%, and Dense-UNet [9] is closer to the proposed method, reaching 94%. The comparison can prove the excellent performance of Dense-UNet as an ideal backbone network choice, but still 1.365% away from the proposed method. The above

Table 3 Proposed method results on the clinic test set

Method	SAT		VAT		Average	
	Jac.	Dic.	Jac.	Dic.	Jac.	Dic.
ASDNet [24]	92.803	96.121	84.001	90.930	88.402	93.526
BUS-GAN [28]	93.196	95.996	79.365	87.667	86.281	91.832
DT-GAN [31]	93.018	96.357	86.862	92.092	89.940	94.225
Proposed S+D+Semi+O	94.307	97.002	87.423	93.129	90.865	95.066

Table 4 Testing results by training with different ratios of annotated images

Method	1/8		1/4		1/2		Full	
	SAT	VAT	SAT	VAT	SAT	VAT	SAT	VAT
Dense-UNet [9]	94.026	88.181	95.893	89.681	96.209	91.159	96.262	91.646
Dabiri [18]	89.921	82.837	93.086	84.348	92.121	84.274	94.939	87.540
ASDNet [24]	85.397	81.736	88.927	83.819	92.894	84.284	96.475	91.825
Proposed	95.391	88.449	96.302	90.666	96.804	92.099	97.015	93.400

can demonstrate the excellent performance of URO-GAN on a tiny number of samples in the abdomen adipose tissue dataset.

4.4 Visualization analysis

This section provides the visualization results that are output by the above-mentioned methods. In Section 4.4.1, we compare the visualization outputs of different approaches. Section 4.4.2 provides an analysis around the visual output of URO-GAN, while Section 4.4.3 compares the output between the semi-supervised approaches.

4.4.1 Visualization results of different models

Figure 9 illustrates four CT images' SAT and VAT segmentation results from different methods visually. Compare with the other methods, the proposed method outputs the least number of red and blue pixels, implying that the URO-GAN output is the closest to the ground truth label. Especially in the VAT region, other methods exist large omissions (red) or misclassifications (blue). In contrast, the proposed method results exist tiny errors of this kind, demonstrating the excellent segmentation ability for both adipose tissues.

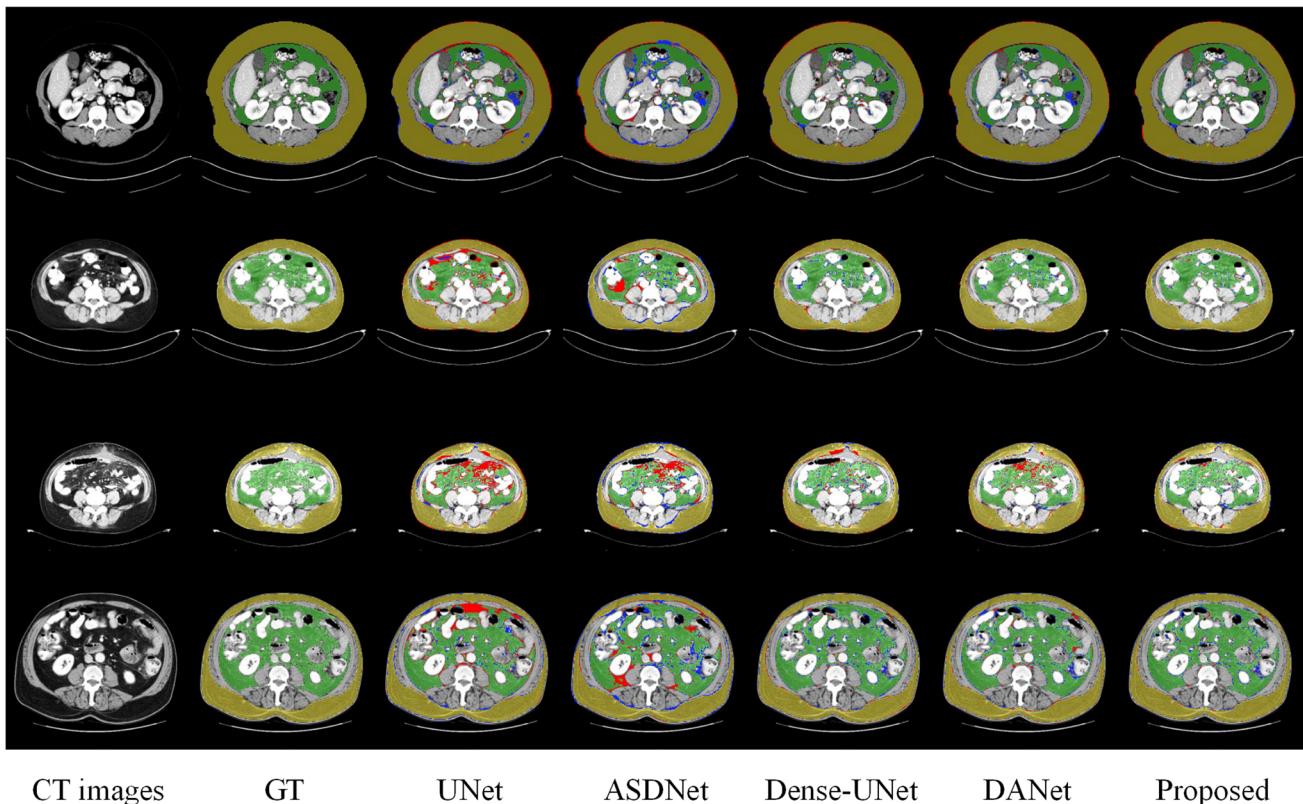


Fig. 9 Illustration of SAT and VAT segmentation resulted from different methods. The comparative tests results on the four images are presented by row, and each column indicates the output of the different models. The yellow pixels indicate the SAT regions correctly segmented by the model, and similarly, the VAT regions in green. The red

pixels indicate the regions where the model missed the segmentation. The blue pixels indicate the regions where the model misclassified as targets

4.4.2 URO-GAN visualization analysis

The outputs of each network in the proposed method are shown in Fig. 10. As can be observed in the figure, we can see a degree of error in the segmentation network's output, mainly presented as red pixels, indicating that the model misses this part of the segmentation. In contrast, the red pixels in the mask output from the optimizer network are reduced, indicating that the optimizer network corrects the error and improves the segmentation accuracy. In addition, we can observe from the confidence map that the discriminator network is trained to judge not only inaccurate areas of the segmentation network output as unreliable but also larger regions as false.

We compared the model visualization results with full-supervised training and semi-supervised training, as shown in Fig. 11. The top two rows provide the results of the proposed semi-supervised model, while the bottom two rows present the results of the proposed full-supervised model. The figure reveals that the discriminator network with full-supervised training is relatively more conservative than that with semi-supervised training. However, the optimizer network does not fix the significant errors in the segmentation network output, and extensive areas of red and blue regions still exist in the final prediction. With

semi-supervised training, the discriminator network is more stringent. The output confidence map tends to discriminate a broader range of regions as untrustworthy. Meanwhile, the O-output has few red or blue pixels, demonstrating that the optimizer network is significantly useful. It seems that the unlabeled data increase the likelihood of the discriminator network discriminating larger regions as untrustworthy. The seemingly unreliable confidence map is beneficial to the optimizer network.

To explain the relationship between the semi-supervision and the optimizer network, we have to review the principle of related networks. Discriminator network only strives to discriminate the labeled data as trustworthy, and the labeled data is undoubtedly matched with the original image. When /southe discriminator network encounters encountering more mismatch cases between the probability map and the original image, the discriminator network tends to discriminate a larger region of the segmentation network output as untrustworthy. The unlabeled data may cause the prediction to mismatch the original image in two ways. The first is that the image quality may be unstable. Even though the segmentation network may segment quite accurately, the poor quality original image makes the discriminator network judge those mismatched regions as untrustworthy. The second reason is that when the segmentation network

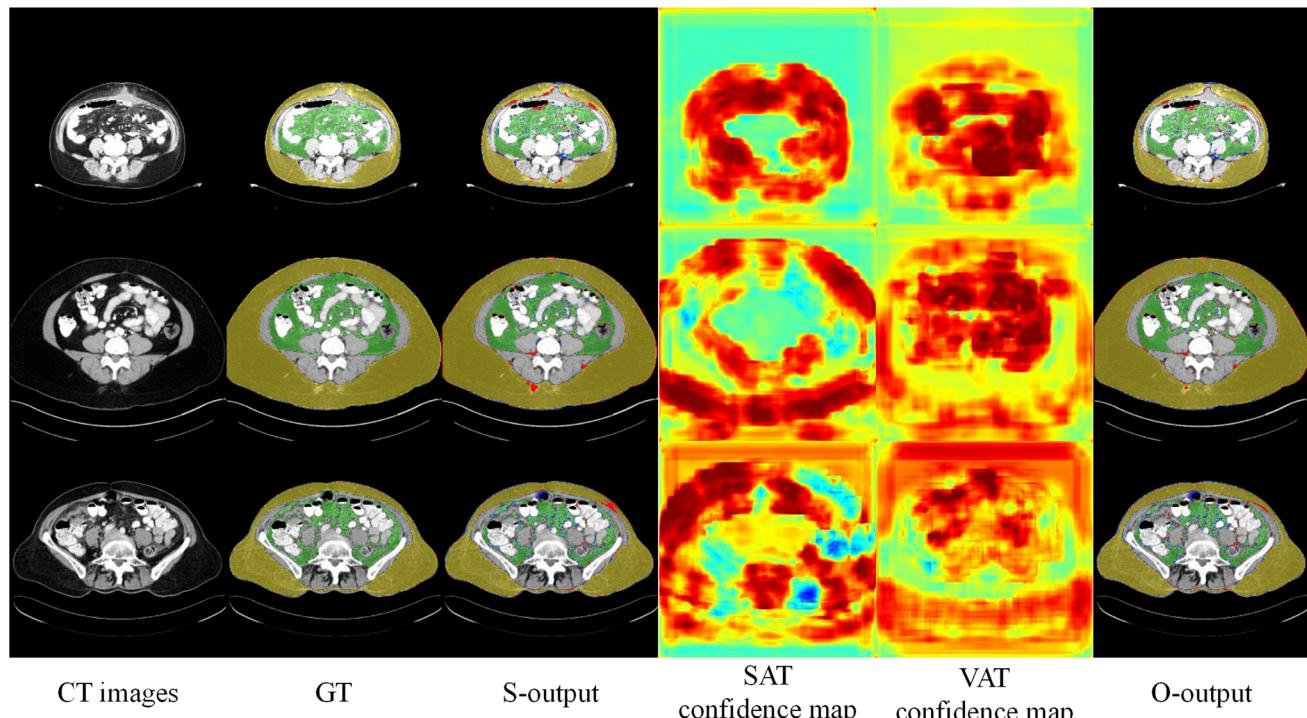


Fig. 10 Illustration of three networks' output of the proposed method. Rows present the test results for the three images, and each column shows the output of each network of the proposed model. The S-output, SAT/VAT confidence map, and O-output are the outputs of the segmentation network, discriminator network, and optimizer network.

The segmentation masks are classified according to yellow for SAT and green for VAT. The blue pixels refer to the misclassified region, while the red pixels refer to the missed region. A redder pixel in the confidence map means that the region is less trustworthy, while a bluer pixel means the region is more trustworthy

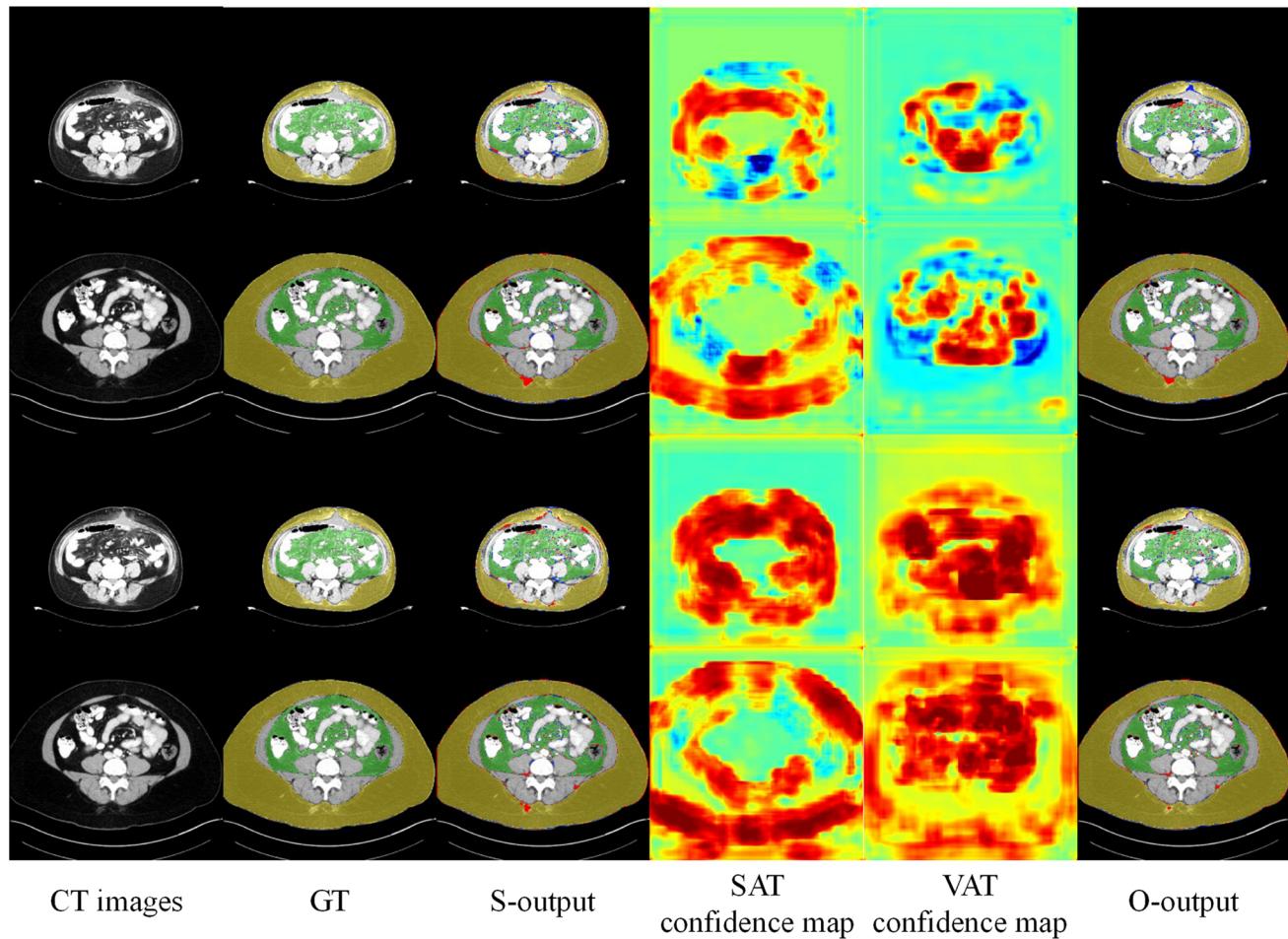


Fig. 11 Illustration of the proposed model with fully supervised and semi-supervised training. The top two rows provide the results of the proposed semi-supervised model, while the bottom two rows present the results of the proposed full-supervised model. The pixel color rules are consistent with Fig. 10

segments the unlabeled image, it is not yet accurate enough, causing the large error in the predicted mask. The discriminator network naturally considers the prediction as not matching the original image, thus judging more regions as unreliable. In the approach without the optimizer network, the increase of these unreliable regions is not fully utilized. Instead, the instability of the adversarial learning can easily lead to unpredictable results. On the contrary, the optimizer network uses these unreliable regions to optimize the results. More unreliable information can enhance the optimization and repair performance of the relevant regions. Therefore, for complex features such as VAT, the combination of unlabeled data and optimizer network can effectively improve the performance of the proposed method. They can also improve the stability and robustness of the overall model, making the unlabeled data more beneficial for adversarial learning models.

However, it is not easy for the proposed URO-GAN to distinguish the area where the tissue is shown too thin in the CT image. As shown in Fig. 12, URO-GAN misses the

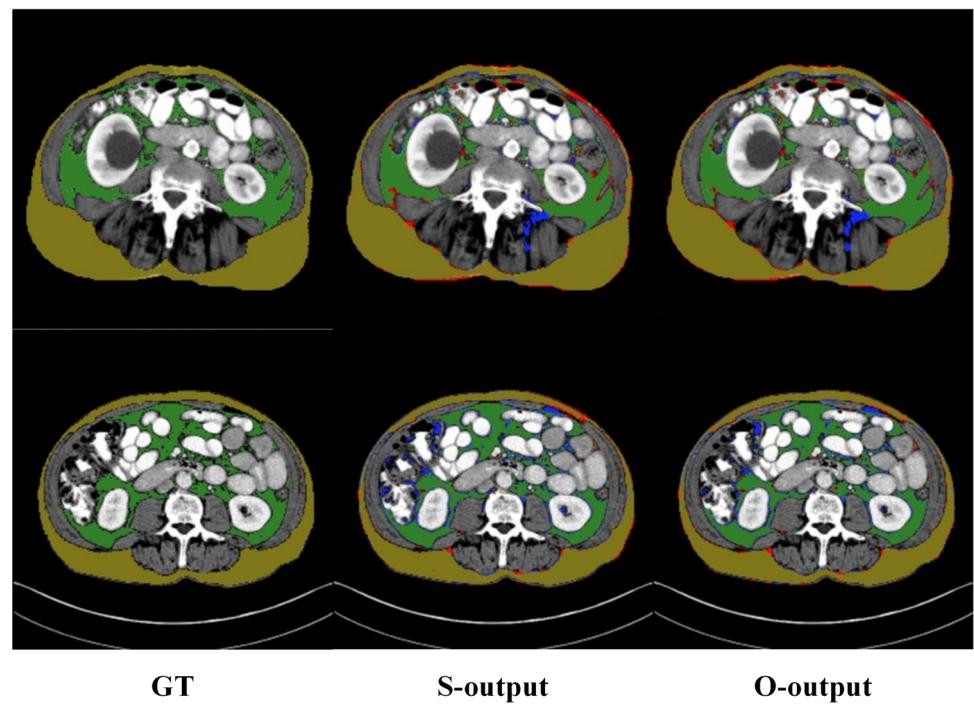
area that is too thin in the SAT. Furthermore, there still exist some wrong segmentation at the edge of SAT segmentation and the internal part of VAT segmentation.

4.4.3 Visualization analysis of semi-supervised methods

By using adversarial learning, our proposed semi-supervised learning method performs better than ASDNet and BUS-GAN on this task. To explain this phenomenon, we have to distinguish between these semi-supervised learning methods.

ASDNet uses adversarial learning to generate confidence maps, which are used to augment the supervised training set. The authors add unlabeled data, let the model read these unlabeled data to obtain a confidence map, and then use the regions of the confidence map with high confidence as labels for the unlabeled images, thus increasing the training set. The problem of this approach is that the quality of the expanded labels is difficult to be guaranteed. In the original paper, the authors only use thresholding to adjust the labeled

Fig. 12 Illustration of the proposed URO-GAN results for bad case



regions as unlabeled images, which makes the training more difficult. If the model cannot output a suitable confidence map from unlabeled images, the expanded labeled dataset will be very limited and may even cause a decrease in accuracy. In the Fig. 13 we show the output confidence map, which is obviously not accurate enough if new masks are extracted by thresholding. Those high confidence regions that are shown in blue in the confidence map are actually

still partially mis-segmented. Inaccurate labeling will make unexpected results.

The semi-supervised learning method proposed by BUS-GAN does not expand the supervised dataset. In this method, the semi-supervised data is present to increase the adversarial strength. Due to the inclusion of unlabeled data, the segmenter generates a set of fake images for the unlabeled images. The discriminator then needs to

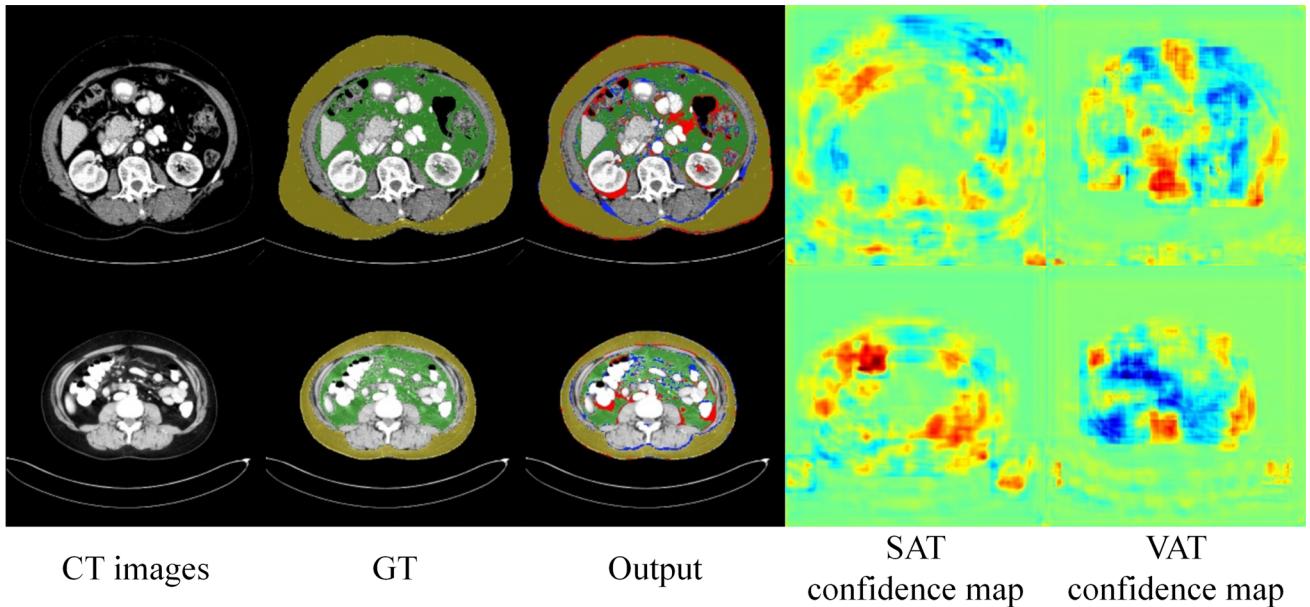


Fig. 13 Illustration of ASDNet outputs. These confidence maps are not suitable as new labels, which make it difficult to expand the supervised dataset

additionally discriminate this set of fake images. After adversarial learning, the segmenter becomes better and better at segmenting unlabeled images. In this approach, the discriminator only determines the authenticity of a whole image, and the adversarial loss has a limited impact on the performance, which makes it difficult to find the optimization direction. This approach is not as good as the confidence map approach to discriminate the truth or falsity of each pixel.

Our semi-supervised approach is derived from methods like BUS-GAN. Unlabeled data is used to enhance the adversarial training, but we let the discriminator determine true and false at pixel granularity and output a confidence map. At the same time, we do not use the resulting confidence maps to augment the supervised dataset, avoiding the addition of inaccurate labels as in ASDNet. Instead, we fed the confidence map into a third network and used it only to optimize the segmentation results obtained from the previous network. In summary, our method outperforms these semi-supervised methods by design, and the experimental results validate the effectiveness of the proposed semi-supervised method.

4.5 Ablation study

In this part, we implement ablation experiments on several components of URO-GAN, including the untrustworthy region excitation block, the inverse-SAT excitation block and the channel attention module.

4.5.1 Impact of the untrustworthy region excitation block

To demonstrate the positive effect of the URE blocks on the optimizer network and the overall model, we compare the segmentation results composed of the URE blocks at different locations, including the case without the URE block. We first label the positions of the “Untrustworthy Region Excitation Block” in Figs. 5 or 6 as L1, L2, L3, and L4 in order from input to output. The URE blocks of L1, L2 are used to fuse the layer information of the discriminator network. Those of L3 and L4 indicate the fusion process of

the segmentation network. A total of six networks consisting of different URE block topologies are compared. In addition to the cases where URE blocks are not used, there are cases where URE blocks are set all over the encoder or decoder, respectively, corresponding to positions (L1, L2) or (L3, L4). Next, we try the symmetric distribution by setting the URE blocks at the positions (L1, L4) or (L2, L3), respectively. The last distribution is to set the URE blocks above all layers from L1 to L4.

Table 5 shows the overall model results with different locations of URE blocks. The six compared models are identical except for the URE block. They are all trained on the abdomen adipose tissue dataset (84 annotated samples, 3969 unannotated samples). The results show that the topology with URE blocks in all four positions gives the best segmentation results. The average Jaccard score achieves 90.969%, which is nearly 2.5% higher than the layout without URE blocks. The result proves the positive impact of URE blocks on the model.

The performance of placing URE blocks at (L3, L4) positions is very close to the best results, which shows that the knowledge obtained from the segmentation network plays a larger role than that obtained from the discriminator network. The optimizer network intends to optimize the output of the segmentation network, so more information from the segmentation network can help the optimization network learn how to improve the accuracy. The URE blocks at (L1, L2) positions enable the optimizer network to match the probability map and the original map. Since the discriminator network has already made a judgment based on this match, the encoding phase of the discriminator network can guide the optimization network to do special processing on the mismatches, thus improving the accuracy of the result .

4.5.2 Impact of the inverse-SAT excitation block

The fact that the SAT results are much better than VAT results validates the difficulty difference between two tasks. The motivation for setting up the SAT-to-VAT transmission mechanism is sufficient. The ISE block is the core module

Table 5 Untrustworthy region excitation block analysis

Location of URE blocks	SAT		VAT		Average	
	Jac.	Dic.	Jac.	Dic.	Jac.	Dic.
No use	92.182	95.903	84.910	91.756	88.546	93.830
L1, L2	93.575	96.658	84.362	91.444	88.969	94.051
L3, L4	94.380	97.096	87.099	93.052	90.740	95.074
L1, L4	94.389	97.104	86.648	92.785	90.519	94.945
L2, L3	94.149	96.974	86.947	92.959	90.548	94.967
L1, L2, L3, L4	94.238	97.015	87.700	93.400	90.969	95.208

Table 6 SAT-to-VAT transmission mechanism analysis

Location of ISE blocks	VAT probability map		VAT predicted mask	
	Jac.	Dic.	Jac.	Dic.
No use	86.306	92.592	86.628	92.779
L1, L2, L3	85.435	92.084	86.258	92.569
L4, L5, L6	85.884	92.349	86.413	92.651
L1, L6	84.903	91.758	85.819	92.306
L1, L2, L5, L6	87.468	93.263	87.452	93.250
L1, L2, L3, L4, L5, L6	86.489	92.689	87.700	93.400

to extract information from the SAT segmentation network and thus deliver the processed information to the VAT segmentation network. To investigate whether the ISE block can effectively help the VAT segmentation network predict the target mask accurately, we tried the ISE block with different numbers and positions. The first and foremost is the model without any ISE blocks, where two paths of the model are not correlated at all. We define the network layers from input to output using the ISE block in Fig. 2 as L1, L2, L3, L4, L5, L6. Similar to the previous section, we then try the case where the blocks are all located on the encoder (L1, L2, L3) or decoder (L4, L5, L6), respectively. We then try three symmetric distributions: two blocks set on (L1, L6), four blocks set on (L1, L2, L5, L6), and blocks set on all six network layers (L1, L2, L3, L4, L5, L6).

Table 6 shows the Jaccard score and dice coefficient of the VAT probability map and the VAT predicted mask with different quantities and positions of the ISE blocks. They are both trained on the abdomen adipose tissue dataset (84 annotated samples, 3969 unannotated samples) with the same hyperparameters, iteration, and batch size. If we only consider the VAT probability map output from the VAT segmentation network, we find that setting blocks at (L1, L2, L5, L6) positions performs the best. However, in terms of the VAT result of the entire model, the layout that setting blocks at all positions performs the best. Compared with not using any ISE blocks, the mechanism with the ISE blocks in all positions helps the model increase the Jaccard score of segmenting VAT by 1.072% and the dice coefficient by 0.621%. It can be seen that the ISE blocks have a positive impact on the VAT segmentation model and achieve an overall improvement in the data flow, which can improve the accuracy of the final output of the model.

4.5.3 Impact of the channel attention module in the SAT segmentation network

To fully demonstrate each of the modules used, the experiment also verify the effectiveness of the CAM module. Table 7 shows the results of the CAM ablation

Table 7 Channel attention module analysis

CAM	SAT probability map		SAT predicted mask	
	Jac.	Dic.	Jac.	Dic.
No use	93.808	96.372	93.861	96.409
Use	94.080	96.929	94.238	97.015

experiments. The SAT probability map in the table denotes the performance of the SAT segmentation network, while the SAT predicted mask indicates the results of the SAT optimizer network. It is obvious that the model with CAM gets better performance. The dice coefficient indexes of both networks are improved by about 0.6%. It can be seen that CAM, as a mature module, is well compatible with our model.

5 Conclusions

In the field of adipose tissue segmentation, the current deep learning segmentation models still need a large amount of professional and accurate annotation data to improve the segmentation quality. In this paper, we propose a semi-supervised learning model, URO-GAN, based on an adversarial learning structure. The URO-GAN can read the abdominal CT cross-sectional image and output the segmented images of subcutaneous adipose tissue and visceral adipose tissue simultaneously.

In order to cope with the situation of few labeled samples, a semi-supervised segmentation method based on adversarial learning is proposed, which uses a small number of labeled samples and more unlabeled data to train the adversarial network. Compared to other semi-supervised adversarial learning methods, the average results of the proposed model are improved by more than 1% on the abdomen adipose tissue dataset.

We propose untrustworthy region optimization mechanism in the URO-GAN to further improve the model performance on semi-supervised and small sample data. The mechanism can utilize the confidence map output from the discriminator network to improve the segmentation quality through the third network called optimizer network. With the optimizer network, the proposed method obtains an average improvement of 0.684% on the Jaccard score in the evaluation of the abdomen adipose tissue dataset.

For the fact that the SAT segmentation task is easier than the VAT segmentation task, we propose the SAT-to-VAT transmission mechanism. The mechanism enables the SAT segmentation network to effectively guide the VAT segmentation network, further improving the VAT

segmentation result by 1.072% of Jaccard score on the abdomen adipose tissue dataset.

In the future work, the URO-GAN will continue to be researched to assist clinicians in scanning and localizing patients' abdominal adipose tissue, automatically providing more accurate subcutaneous and visceral fat segmentation masks as a reliable basis. With more research, the model can also be further used to calculate patients' fat areas to diagnose diseases related to patients' fat levels, avoiding errors caused by doctors' subjective judgments on CT images. In addition, the proposed training segmentation strategy is portable. As only a small amount of annotated data is required, it is potential to extend URO-GAN to other studies, where the medical CT datasets cannot be labelled by enough experts with sufficient effort.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. GBD 2015 Obesity Collaborators (2017) Health effects of overweight and obesity in 195 countries over 25 years. *N Engl J Med* 377(1):13–27
2. Wu Y, Jiang X, Fang Z, Gao Y, Fujita H (2021) Multi-modal 3D object detection by 2D-guided precision anchor proposal and multi-layer fusion. *Appl Soft Comput* 108:107405
3. Zhu K, Jiang X, Fang Z, Gao Y, Fujita H, Hwang JN (2021) Photometric transfer for direct visual odometry. *Knowl-Based Syst* 213:106671
4. Mourtzakis M, Prado CM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE (2008) A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Appl Physiol Nutrit Metabolism* 33(5):997–1006
5. Després JP, Lemieux I (2006) Abdominal obesity and metabolic syndrome. *Nature* 444(7121):881–887
6. De Larocheillière E, Côté J, Gilbert G, Bibeau K, Ross MK, Dion-Roy V, ... Larose É (2014) Visceral/epicardial adiposity in nonobese and apparently healthy young adults: association with the cardiometabolic profile. *Atherosclerosis* 234(1):23–29
7. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440
8. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241
9. Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y (2017) The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 11–19
10. Cui H, Yuwen C, Jiang L, Xia Y, Zhang Y (2021) Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images. *Comput Methods Prog Biomed* 106142
11. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2019) Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 3146–3154
12. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
13. 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, Cham, pp 421–429
14. Sinha A, Dolz J (2020) Multi-scale self-guided attention for medical image segmentation. *IEEE J Biomed Health Inform*
15. Li C, Tan Y, Chen W, Luo X, He Y, Gao Y, Li F (2020) ANU-Net: Attention-based Nested U-Net to exploit full resolution features for medical image segmentation. *Comput Graph* 90:11–20
16. Sun J, Darbehani F, Zaidi M, Wang B (2020) SAUNet: shape attentive U-Net for interpretable medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp 797–806
17. Estrada S, Lu R, Conjeti S, Orozco-Ruiz X, Panos-Willuhn J, Breteler MM, Reuter M (2020) Fatsegnet: A fully automated deep learning pipeline for adipose tissue segmentation on abdominal dixon MRI. *Magnet Resonance Med* 83(4):1471–1483
18. Dabiri S, Popuri K, Ma C, Chow V, Feliciano EMC, Caan BJ, ..., Beg MF (2020) Deep learning method for localization and segmentation of abdominal CT. *Comput Med Imaging Graph* 85:101776
19. Pan S, Hou X, Li H, Sheng B, Fang R, Xue Y, ..., Qin J (2019) Abdominal Adipose Tissue Segmentation in MRI with Double Loss Function Collaborative Learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp 41–49
20. Sadananthan SA, Prakash B, Leow MKS, Khoo CM, Chou H, Venkataraman K, ..., Velan SS (2015) Automated segmentation of visceral and subcutaneous (deep and superficial) adipose tissues in normal and overweight men. *J Magnet Reson Imag* 41(4):924–934
21. Choi YJ, Seo YK, Lee EJ, Chung YS (2015) Quantification of visceral fat using dual-energy x-ray absorptiometry and its reliability according to the amount of visceral fat in Korean adults. *J Clin Densitom* 18(2):192–197
22. Feng Z, Nie D, Wang L, Shen D (2018) Semi-supervised learning for pelvic MR image segmentation based on multi-task residual fully convolutional networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp 885–888
23. Bai W, Oktay O, Sinclair M, Suzuki H, Rajchl M, Tarroni G, ..., Rueckert D (2017) Semi-supervised learning for network-based cardiac MR image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp 253–260
24. Nie D, Gao Y, Wang L, Shen D (2018) ASDNet: attention based semi-supervised deep networks for medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 370–378
25. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, ..., Bengio Y (2014) Generative adversarial networks. arXiv:1406.2661
26. Luc P, Couprise C, Chintala S, Verbeek J (2016) Semantic Segmentation using Adversarial Networks. In: NIPS Workshop on Adversarial Training
27. Zhang Y, Yang L, Chen J, Fredericksen M, Hughes DP, Chen DZ (2017) Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp 408–416

28. Han L, Huang Y, Dou H, Wang S, Ahamad S, Luo H, ..., Zhang J (2020) Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network. *Comput Methods Programs Biomed* 189:105275
29. Hung WC, Tsai YH, Liou YT, Lin YY, Yang MH (2019) Adversarial learning for semi-supervised semantic segmentation. In: 29th British Machine Vision Conference BMVC, p 2018
30. Li G, Wan J, He S, Liu Q, Ma B (2020) Semi-supervised semantic segmentation using adversarial learning for pavement crack detection. *IEEE Access* 8:51446–51459
31. Decourt C, Duong L (2020) Semi-supervised generative adversarial networks for the segmentation of the left ventricle in pediatric MRI. *Comput Biol Med* 123:103884
32. Roy AG, Siddiqui S, Pölsterl S, Navab N, Wachinger C (2020) Squeeze, excite guided few-shot segmentation of volumetric images. *Med Image Anal* 59:101587
33. Shaban A, Bansal S, Liu Z, Essa I, Boots B (2017) One-shot learning for semantic segmentation. arXiv:[1709.03410](https://arxiv.org/abs/1709.03410)
34. Fei-Fei L, Fergus R, Perona P (2006) One-shot learning of object categories. *IEEE Trans Pattern Anal Machine Intell* 28(4):594–611
35. Miller EG, Matsakis NE, Viola PA (2000) Learning from one example through shared densities on transforms. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662), vol 1. IEEE, pp 464–471
36. Fei-Fei L (2006) Knowledge transfer in learning to recognize visual objects classes. In: Proceedings of the International Conference on Development and Learning (ICDL), p 11
37. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
38. Lee CY, Xie S, Gallagher P, Zhang Z, Tu Z (2015) Deeply-supervised nets. In: Artificial intelligence and statistics. PMLR, pp 562–570
39. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S (2017) Least squares generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2794–2802
40. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, ..., Zheng X (2016) Tensorflow: A system for large-scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI), vol 16, pp 265–283
41. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv:[1412.6980](https://arxiv.org/abs/1412.6980)
42. Dice LR (1945) Measures of the amount of ecologic association between species. *Ecol* 26(3):297–302
43. Holger R, Lu L, Seff A, Cherry KM, Hoffman J, Wang Sn, ..., Summers RM (2015) A new 2.5 D representation for lymph node detection in CT. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2015.AQHDCNM>
44. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, ..., Prior F (2013) The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imag* 26(6):1045–1057

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”). Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com