OPEN ACCESS
Journal of Innovative Optical Health Sciences
2250018 (11 pages)
© The Author(s)
DOI: 10.1142/S1793545822500183

ICA-Unet: An improved U-net network for brown adipose tissue segmentation

Haolin Wang*, Zhonghao Wang*, Jingle Wang‡, Kang Li*, Guohua Geng*,
Fei Kang†,§ and Xin Cao*,¶

*School of Information Sciences and Technology
Northwest University, Xi'an, Shaanxi 710127, P. R. China

†Department of Nuclear Medicine, Xijing Hospital
Fourth Military Medical University
Xi' an, Shaanxi 710127, P. R. China

‡University of Wisconsin-Madison
Madison, Wisconsin 53715, USA
§fmmukf@qq.com
¶xin_cao@163.com

Received 2 January 2022
Accepted 20 February 2022
Published 30 March 2022

Brown adipose tissue (BAT) is a kind of adipose tissue engaging in thermoregulatory thermogenesis, metaboloregulatory thermogenesis, and secretory. Current studies have revealed that BAT activity is negatively correlated with adult body weight and is considered a target tissue for the treatment of obesity and other metabolic-related diseases. Additionally, the activity of BAT presents certain differences between different ages and genders. Clinically, BAT segmentation based on PET/CT data is a reliable method for brown fat research. However, most of the current BAT segmentation methods rely on the experience of doctors. In this paper, an improved U-net network, ICA-Unet, is proposed to achieve automatic and precise segmentation of BAT. First, the traditional 2D convolution layer in the encoder is replaced with a depth-wise over-parameterized convolutional (Do-Conv) layer. Second, the channel attention block is introduced between the double-layer convolution. Finally, the image information entropy (IIE) block is added in the skip connections to strengthen the edge features. Furthermore, the performance of this method is evaluated on the dataset of PET/CT images from 368 patients. The results demonstrate a strong agreement between the automatic segmentation of BAT and manual annotation by experts. The average DICE coefficient (DSC) is 0.9057, and the average Hausdorff distance is 7.2810. Experimental results suggest that the method proposed in this paper can achieve efficient and accurate automatic BAT segmentation and satisfy the clinical requirements of BAT.

Keywords: PET/CT; segmentation of brown adipose tissue; U-net; medical image processing; deep learning.

¶Corresponding author.
Haolin Wang and Zhonghao Wang contributed equally to this work and are considered co-first authors.

1. Introduction

Brown adipose tissue (BAT) is a special type of adipose tissue. From one perspective, it is similar to white adipose tissue (WAT) and is an energy storage tissue in the human body. From another perspective, BAT is undergoing extremely active metabolic activities owing to a considerable number of mitochondria in BAT cells. Studies have revealed that three main physiological purposes of BAT can be identified, namely, thermoregulatory thermogenesis, metaboloregulatory thermogenesis, and secretory.[1,2] Therefore, BAT has essential physiological significance for human body temperature regulation, resistance to cold, prevention of obesity, regulation of energy balance, and resistance to infection.[3–5] The activity of BAT expresses significant differences at different ages and genders. Its activity in infants and women is higher than that in adults and men. The latest research has reported that BAT activity is negatively correlated with body weight, and BAT may be adopted as a target tissue for the treatment of obesity.[6] Therefore, BAT-related research is associated with a wide range of clinical applications. With the vigorous development of modern medicine, detection methods for BAT are constantly improved. Positron emission tomography/computer tomography (PET/CT) technology has become the mainstream detection technology for BAT. Besides, 18F-fluoro-2-deoxyglucose (18F-FDG) is broadly used as the PET/CT tracer.[7] The mechanism is that glucose is metabolized vigorously and accumulated in BAT due to the different metabolic states of different tissues of the human body.[8,9] These characteristics can be reflected through PET images for detection and analysis, as shown in Fig. 1.

Regarding the segmentation of BAT, existing studies are primarily based on the experience of radiologists and nuclear medicine physicians. Researchers have demonstrated that thresholding and clustering are very suitable for the segmentation of BAT since BAT exists in specific anatomical locations and PET images have high contrast.[10–13] Generally, simple thresholding was used for segmenting BAT. First, the individual's standard uptake value (SUV) is calculated based on PET data.[14,15] Second, it is manually divided with medical imaging software. BAT is present if the diameter of the tissue area is greater than 5 mm and the CT density is restricted to $-190 - -30\,\mathrm{HU}$,[16] and the SUV is more than $2\,\mathrm{g/ml}$ or $3\,\mathrm{g/ml}$ in the corresponding 18F-FDG PET image.[17,18] Finally, BAT is distinguished from lymph nodes, blood vessels, bones, thyroid, and other tissues according to anatomical knowledge.[19] However, this type of method has the following shortcomings: (1) The traditional BAT segmentation step requires excessive imaging and clinical knowledge, and the conclusions are mostly the subjective judgment of clinicians; (2) the traditional BAT segmentation method depends on the standard division of thresholds, while it is difficult for standard thresholds segmentation in complex differentiation problems because of the specificity of the tissues and organs of different patients. Therefore, it is urgent to develop an automatic BAT segmentation method for tackling the above problems.

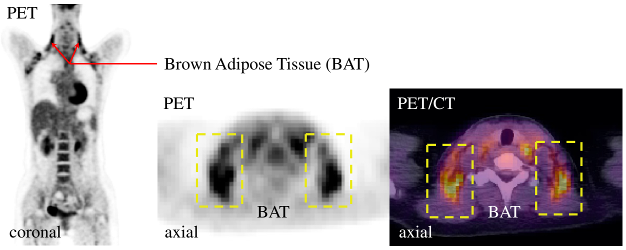

Fig. 1. PET/CT image and anatomical location of BAT.

Recent success in deep learning, especially the use of a Deep Convolutional Neural Network,[20] has accelerated the development of automatic image segmentation. In deep learning algorithms, the U-net network structure[21] is widely used for medical image segmentation.[22–26] The U-net network consists of an encoder and a decoder, as well as skip connections between the encoder and the decoder. The image is first convolutionally down-sampled by the encoder several times to obtain feature maps of different scales, indicating that the features of different scales are learned. Then, the bottom feature map is up-sampled in the decoder, and the encoder is correspondingly connected by skipping scale feature maps. As a result, feature maps of different scales are merged. Finally, the network combines low-resolution information and high-resolution information. The low-resolution information provides the location and category information of the segmentation target, while high-resolution information is required in the edge segmentation. Under the combination of both, the medical image segmentation task can be well completed by U-net.

Accordingly, a BAT segmentation method is designed in this paper based on improved U-net, ICA-Unet. The U-net network is improved based on channel attention block, a depth-wise over-parameterized convolutional (Do-Conv) layer, and image information entropy (IIE) block, which will be described in detail in Sec. 2.2.

2. Materials and Methods

2.1. *Materials*

This retrospective analysis was approved by the Ethics Committee of Xijing Hospital (Approval No. KY201730081), and the informed consent was waived. The PET/CT images involved in this paper are obtained from the Department of Nuclear Medicine, Xijing Hospital, Fourth Military Medical University. The dataset includes 368 sets of DICOM image sequences, containing CT images with a size of 512 × 512 and a corresponding PET image with a size of 274 × 274.

The images need to be preprocessed before the dataset is segmented. First, the SUV value is calculated based on the original PET data according to formula (1) (hereinafter referred to as the SUV calculated based on the original PET data as the PET data). During the process of calculating the SUV value, the calculation rule based on the DICOM tag is adopted.[27] Besides, X_{PET} is set as a three-dimensional matrix of PET data and Y_{SUV} is set as a three-dimensional matrix of SUV data. Other variables are acquisition time (T_A), patient weight (W_P), radiopharmaceutical start time (T_{RS}), radionuclide total dose (D_{RT}), radionuclide half-life (L_{RH}), rescale intercept (I_R), and rescale slope (S_R). The calculation formula is:

$$Y_{\mathrm{suv}} = \frac{(I_R + S_R X_{\mathrm{PET}}) \exp\left(\frac{\ln 2(T_A - T_{\mathrm{RS}})}{L_{\mathrm{RH}}}\right)}{D_{\mathrm{RT}} \div W_P}. \quad (1)$$

Second, the original CT data have been optimized with the median filtering method, which is widely used in medical images noise processing, to reduce the noise in the CT data. Finally, the PET data are up-sampled by the nearest neighbor interpolation resampling method and enlarged to be the same as the CT data. Besides, the related spatial parameters of the PET data are matched with the corresponding CT parameters to manage the problem of the inconsistent size and image parameters of the PET and CT.

Since BAT is broadly distributed in the neck region of the human body, it is necessary to obtain the neck ROI region before BAT segmentation for avoiding interference from other nonBATs with high SUV values in the human body. Additionally, the PET/CT image cannot be guaranteed to be in a fixed proportion of the image in the process of acquiring the images. Thus, Faster RCNN[28,29] has been employed to obtain the ROI area of the neck to overcome this complication.

The ROI area is obtained through the network. On this basis, the corresponding upper and lower boundary slice interval $(x, x + h)$ of the axial direction of the PET/CT data are acquired. This is taken as the interval to obtain several PET/CT images and construct a bimodal dataset.

After the data processing described above, about 4500 sets of bimodal PET/CT images can be obtained. Then, annotations for each PET/CT image are created. Consequently, the annotated data are randomly divided into 70% training data and 30% test data for training and testing.

2.2. *Methods*

2.2.1. *Network architecture*

The architecture of the proposed ICA-Unet network is illustrated in Fig. 2. The network is based on the classic U-net structure,[21] with the introduction of image information entropy,[30] channel attention,[31] and Do-Conv layer.[32] Specifically, our study follows the classic structure of the U-net. In the encoder module, the PET/CT bimodal image is first input into the network as two channels. Second, the 2D convolution module in the U-net structure is replaced with the Do-Conv module. After a layer of 3 × 3 × 3 convolution with the stride of 1 and zero padding, the rectified linear unit (ReLU) activations and batch normalization (BN) are calculated. Then, the channel weight is calculated by the channel attention module and multiplied by the feature map. Finally, it passes through a block of Do-Conv, ReLU, and BN. Additionally, successive 2 × 2 × 2 max pooling with the stride of 2 is performed to enlarge receptive fields after the double-layer convolution.

Symmetrically with the encoder, the feature maps of the subsequent decoder are up-sampled four times with de-convolutions to restore spatial details. Specifically, a 2 × 2 de-convolutions module

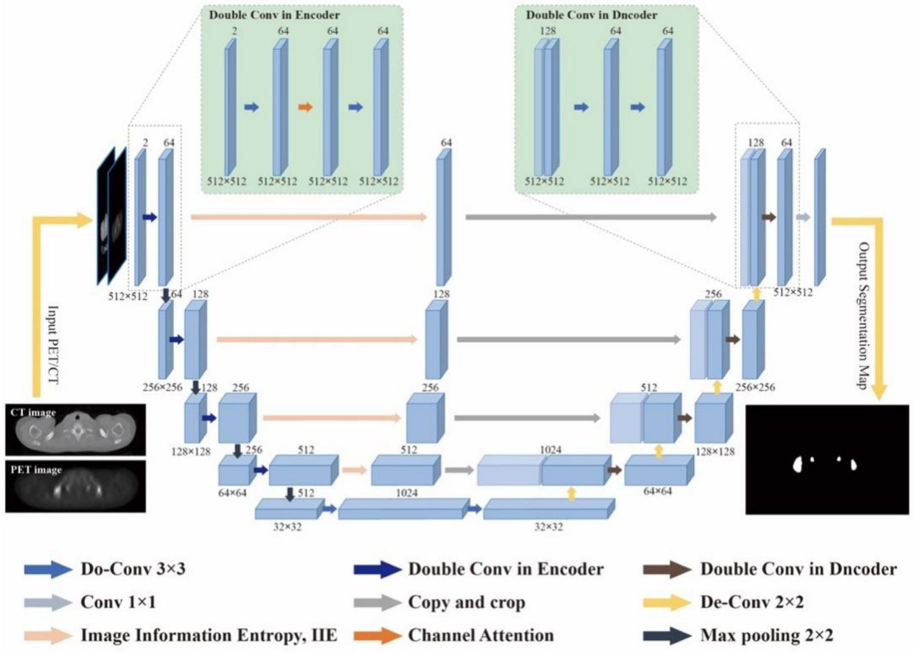

Fig. 2. ICA-Unet network structure diagram.

with a step size of 2 is performed, followed by the same convolution operation as the encoding module. Furthermore, the skip connections are deployed, and the IIE module is introduced to calculate the IIE of the down-sampled feature map, so as to strengthen the edge information. Then, this feature map is fused with the feature map of the same level obtained from the decoding module. The global context information is complementary to the spatial details. Finally, the segmentation result is output after 1×1 single-layer convolution.

2.2.2. *Image information entropy*

In 1948, C. E. Shannon, the father of information theory, published a paper "A Mathematical Theory of Communication", pointing out that any information has redundancy, and its size is related to the probability or degree of confusion of each symbol in the information.[33] Shannon cited the concept of thermodynamics and called the average amount of information after eliminating redundancy as IIE.

In the information theory, information entropy is defined as the expectation of a random variable $I(X)$ in the set $(X, q(X))$,

$$H(x) = \sum_{x \in X} q(x) I(x) = \sum_{x \in X} q(x) \log q(x), \qquad (2)$$

where $H(X)$ denotes the information entropy of X, which describes the degree of confusion and uncertainty of the elements in X.

Pixel is the basic unit of a digital image. Image data are essentially a matrix of pixels in a computer. Essentially, the difference in images is that pixels of different gray levels are distributed in different spatial regions with different probabilities. Therefore, the value of k is set to 255 for the image of k-level grayscale, and the $i(i \in 1, \ldots, k)$ level grayscale is represented by pi. Then, the entropy is as follows:

$$H(p_i) = p_i \log_2 \frac{1}{p_i}, \qquad (3)$$

where $0 \leq i \leq k$, the accumulation of information entropy of different grayscale levels is defined as image information entropy, then the IIE of the entire image is as follows:

$$H = \sum_{i=0}^{k} H(p_i) = -\sum_{i=0}^{k} p_i \log_2(p_i), \quad k = 255, \quad (4)$$

where p_i indicates the probability of each level of grayscale pixels in the entire image. When $p_i = 0$, $p_i \log(p_i) = 0$. p_i is calculated by the grayscale histogram, that is, the quotient of the number of pixels of grayscale i and the total number of pixels in the image.

In the U-net network, the decoder involves a combination of four times double convolutional layers and pooling layers. There are multiple feature maps of different levels, scales, and aspects in the output of each pooling layer. By calculating the IIE based on the feature map, the edge of the object and the rapidly changing pixel information in the image can be captured through the retention of the detailed texture structure of the original image. Meanwhile, the edge feature of the object can be enhanced to make the generated image feature more expressive. Furthermore, the enhancement of edge information can contribute to better contour integrity and coherence of the final segmentation results of the network.

2.2.3. *Channel attention*

In 2017, Senet[31,34] won the championship in the image classification task of the ImageNet competition. It performed an attention mechanism in the channel dimension to significantly improve the network performance.

The channel attention mechanism consists of three operations: Squeeze, Excitation, and Scale. The Squeeze operation compresses the two-dimensional features of each channel into a real number through global pooling, equivalent to having a global receptive field. Assuming that there are C channels in total, a $1 \times 1 \times C$ feature will eventually be obtained. The purpose of the Excitation operation is to generate a weight for each channel. Specifically, the dimension of the feature map is first reduced to $1/r$ of the original through a fully connected layer. Then, the ReLU is calculated, and the original dimension C of the feature map is obtained through a fully connected layer. Finally, the sigmoid function is performed for normalization. The Scale operation is to multiply the normalized weight coefficient with the feature map of each channel.

In this paper, bimodal data of PET/CT are introduced. The channel attention module can assist the network in judging the importance of different channels. In other words, the network can better judge the information importance of CT and PET data after convolution. Therefore, the introduction of the channel attention module is conducive to learning crucial information in the bimodal data and better extracting image features.

2.2.4. *Depth-wise over-parameterized convolution*

Li *et al.* proposed depth-wise over-parameterized convolution (Do-Conv),[32] which can replace the traditional convolution to accelerate the network convergence and improve the performance of the network. Do-Conv is a combination of traditional convolution and depth-wise convolution.

Assume that the number of channels of the input feature map is C_{in}, the size of the convolution kernel is $M \times N$, and the output channel of the feature map is C_{out}. The convolution kernel $\boldsymbol{W}$ can be expressed as $\boldsymbol{W} \in R^{C_{\text{out}} \times (M \times N) \times C_{\text{in}}}$. With $*$ representing the traditional convolution operation, $\boldsymbol{O} = \boldsymbol{W} * \boldsymbol{P}$.

$$\boldsymbol{O}_{C_{\text{out}}} = \sum_{i}^{(M \times N) \times C_{\text{in}}} \boldsymbol{W}_{C_{\text{out}} i} \boldsymbol{P}_i. \tag{5}$$

Different from the traditional convolution operation, a channel of the output feature in depth-wise convolution is only related to a specific channel of the input feature rather than other channels of the input feature. Assuming that there are D_{mul} convolution kernels with a size of $M \times N$ and the input feature channel is C_{in}, the convolution kernel $\boldsymbol{D}$ can be expressed as $\boldsymbol{D} \in R^{(M \times N) \times D_{\text{mul}} \times C_{\text{in}}}$, and the output channel can be expressed as $D_{\text{mul}} \times C_{\text{in}}$. With $\circ$ indicating depth-wise convolution, $\boldsymbol{O} = \boldsymbol{D} \circ \boldsymbol{P}$.

$$\boldsymbol{O}_{D_{\text{mul}} C_{\text{in}}} = \sum_{i}^{M \times N} \boldsymbol{W}_{i D_{\text{mul}} C_{\text{in}}} \boldsymbol{P}_{i C_{\text{in}}}. \tag{6}$$

Do-Conv is to perform depth-wise convolution on the input feature vector and then calculates traditional convolution. It can be written as follows:

$$\boldsymbol{O} = \boldsymbol{W} * (\boldsymbol{D} \circ \boldsymbol{P}) = \left(\boldsymbol{D}^{\mathrm{T}} \circ \boldsymbol{W}\right) * \boldsymbol{P}. \tag{7}$$

Since the network performs IIE modules and channel attention modules, the convergence speed of the network will be affected. Therefore, Do-Conv is conducted in the encoder and decoder to replace the traditional 2D convolution, so as to accelerate the network convergence speed and improve the network performance.

2.2.5. *Implementation*

The proposed method was implemented using Python language and Pytorch package[35] on the workstation with single graphics processing unit

(NVIDIA GeForce GTX TITAN V). The Loss function of the network was composed of Sigmoid and BECLoss. Assuming there are N batches and each batch predicts n labels, the Loss function can be defined as follows:

$$\begin{aligned}\text{loss} &= \{l_1, \ldots, l_N\}, l_n \\ &= -[y_n \log_2(\sigma(x_n)) + (1 - y_n)\log_2(1 - \sigma(x_n))],\end{aligned} \tag{8}$$

where $\sigma(x_n)$ indicates the Sigmoid function, which can map x to the interval (0, 1):

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \tag{9}$$

The network was trained by an RMSProp optimizer[36] with rho of 0.9 and e of 0.0001. The initial learning rate was set to 0.00003. The training contained 100 epochs, and the batch size was set to 4. In the training and testing, 512×512 PET and CT images were combined to form a double-channel $2 \times 512 \times 512$ matrix and input into the network. After inference, the segmentation of each image was formed. Moreover, the proposed ICA-Unet was performed three times and the average value was taken as the final result to alleviate the impact of random initialization in training. The network had been fully trained, as shown in Fig. 3. The loss of the network has converged.

2.2.6. *Evaluation metrics*

With expert manual annotations as ground truth, the segmentation performance of ICA-Unet was quantitatively evaluated with the following six metrics[37]: (1) mIoU, (2) Sensitivity (SEN), (3) Specificity (SPE), (4) Dice Similarity Coefficient (DSC), (5) accuracy (ACC), and (6) Hausdorff Distance (HD)[38]:

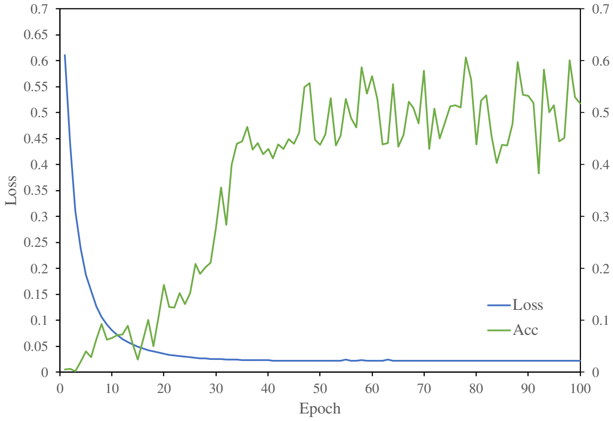

Fig. 3. The loss curve in the process of training.

$$\text{mIoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}, \tag{10}$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{11}$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{12}$$

$$\text{DSC} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}}, \tag{13}$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}, \tag{14}$$

$$\text{HD} = \max\{d_{\text{HD}}(A, B), d_{\text{HD}}(B, A)\}, \tag{15}$$

$$d_{\text{HD}}(A, B) = \max_{x \in A}\left\{\min_{y \in B}\{d(x, y)\}\right\}, \tag{16}$$

where TP and FP denote the numbers of true positives and false positives, respectively; TN and FN refer to the numbers of true negatives and false negatives, respectively; HD indicates the maximum distance between two pixels sets; $d_{\text{HD}}(A, B)$ designates the directed Hausdorff distance between the ground truth and the predicted value; $d(x, y)$ stands for the Euclidean distance between two pixels.

3. Results and Discussion

3.1. *Comparison to state-of-the-art methods*

The segmentation results on the PET/CT dataset built previously were obtained by our method with the ICA-Unet network. The other six state-of-the-art methods, CT threshold, PET threshold, CT and PET Threshold intersection, U-net,[21] U-net++,[39] and SegNet,[40] are presented in Fig. 4 and Table 1. The visualization results imply that in the examples, the automatic segmentation results obtained by our method are more anxious and consistent with the ground truth. Particularly, the segmentation boundary is clearer, complete, and coherent.

It can be observed from Fig. 4 that the segmentation results obtained from the CT threshold method (Fig. 4(b)), in which the areas with a CT density of greater than −190 HU and less than −30 HU are regarded as BAT, contain the area of BAT. However, there are significant errors of

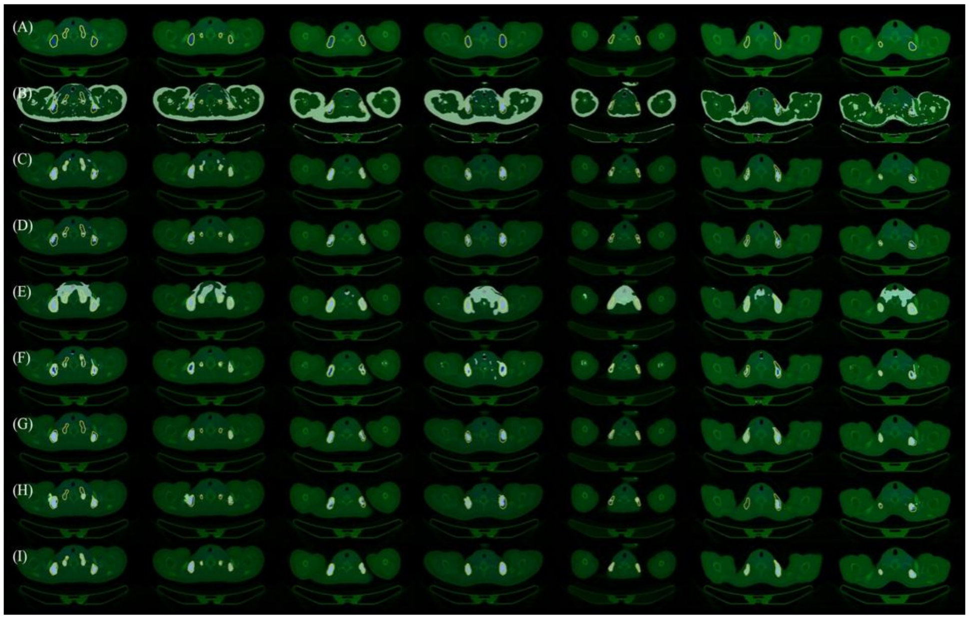

Fig. 4. Comparison of experimental results. (a) Comparison diagram, actual results; (b) CT threshold; (c) pet threshold; (d) intersection of CT and pet threshold; (e) K means; (f) U-net; (g) U-net++; (h) Segnet; (i) Proposed.

segmentation and a considerable number of WATs treated as a BAT area mistakenly. The results obtained from the CT threshold method are unreasonable. The segmentation results from the PET threshold method (Fig. 4(c)), in which the areas with the SUV values of greater than 2 g/ml or 3 g/ml are regarded as the BAT area, exclude a large number of WATs, and contain a more accurate area of BAT. However, the results are under-segmented and over-segmented, as well as discontinuous and noisy. Moreover, the edges are not smooth, reflecting that the poor segmentation results from the PET threshold. Segmentation results from CT and PET threshold intersection methods (Fig. 4(d)) demonstrate that the accuracy of the segmentation area is significantly improved compared with the PET threshold method and the CT threshold method. Nevertheless, the results are also under-segmented, and the region is discontinuous, resulting in the average segmentation results. The segmentation results obtained from the K means method (Fig. 4(e)) contain the area of BAT. However, there is significant over-segmentation, and a considerable number of surrounding tissues are mistakenly treated as BAT regions. The results obtained from the K means method are unreasonable. Segmentation results from U-net (Fig. 4(f)) contain most BAT areas, while the area is hollow, discontinuous, and noisy, and the edges are not smooth. This is contrary to the ground truth. Compared with the results from U-net, the accuracy of segmentation results from U-net++ (Fig. 4(g)) has been enhanced. Nonetheless, there is still a certain amount of under-segmentation.

Table 1. Evaluation index of BAT segmentation results by different methods.

Methods	mIoU	SEN	SPC	DSC	ACC	HD
CT_range	0.0567	0.6075	0.8911	0.1026	0.8876	112.8274
PET_range	0.4163	0.6256	0.9981	0.6628	0.9934	22.2598
CT_and_PET_range	0.4650	0.4399	0.9995	0.5486	0.9925	23.4087
K means	0.1941	0.4348	0.9205	0.2595	0.9152	76.7140
Unet	0.4862	0.6190	0.9970	0.6390	0.9929	70.9661
Unet++	0.6557	0.6789	0.9975	0.7066	0.9930	18.8122
Segnet	0.5774	0.5898	0.9970	0.6848	0.9922	30.6596
Proposed	**0.8322**	**0.8400**	**0.9999**	**0.9057**	**0.9982**	**7.2810**

Segmentation results from SegNet (Fig. 4(h)) are similar to those from U-net++, while there are more under-segmentation and over-segmentation compared to U-net++. The method proposed (Fig. 4(i)) in this paper can obtain semblable segmentation results with the ground truth compared with the previous methods, with the more delicate and smoother boundary of results. It grapples with the problems of under-segmentation and discontinuity in other methods.

Table 1 suggests that the ICA-Unet we proposed had the highest Dice coefficient (DSC) and mIoU compared with the traditional threshold methods. Besides, the ICA-Unet we proposed also had the highest DSC and mIoU and the lowest Hausdorff distance compared with mainstream medical image segmentation networks.

The data in Table 1 reveal that the method we proposed had significant advantages in performance and accuracy and avoids the problems of under-segmentation, discontinuity, and difference in individual thresholds caused by static threshold segmentation methods. Objectively, the ICA-Unet can realize the automatic segmentation of BAT.

3.2. *Evaluation of network architectures*

In this study, the six possible combinations of the Unet, IIE, Do-Conv, and CAT modules were conducted to assess the effectiveness of our network (ICA-Unet) architecture. Moreover, its accuracy was evaluated with mIoU and loss convergence speed. The quantification results are provided in Table 2 and Fig. 5. Thus, the following conclusions can be drawn. (1) Our architecture (U-net + IIE+ Do + CAT) had the highest accuracy compared with other architectures; (2) the convergence rate of Loss is relatively fast when our architecture is trained.

As suggested from Table 2 and Fig. 5, further analysis was conducted on the IIE module. To sum up, the segmentation accuracy of the network had been improved with the introduction of the image IIE module. It exhibited an improvement of 0.07

Table 2. Ablation experiment results.

Methods	mIoU	SEN	SPC	ACC	HD	Loss convergence epoch
Unet	0.4862	0.6190	0.9970	0.9929	70.9661	12
Unet + IIE	0.5523	0.6726	0.9970	0.9931	68.2250	32
Unet + Do	0.4483	0.6134	0.9930	0.9895	76.7479	**10**
Unet + CAT	0.4840	0.4037	0.9670	0.9583	97.4203	28
Unet + IIE + Do	0.3977	0.3447	0.9943	0.9845	96.0314	37
Unet + IIE + CAT	0.6068	0.6641	0.9998	0.9939	16.2010	50
Unet + Do + CAT	0.5391	0.5750	0.9982	0.9929	23.0305	16
Proposed	**0.8322**	**0.8400**	**0.9999**	**0.9982**	**7.2810**	28

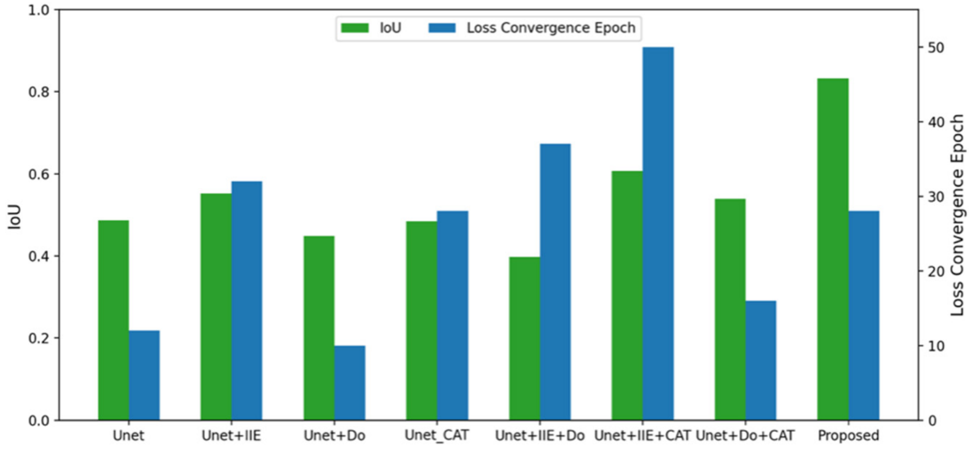

Fig. 5. Histogram of ablation experiment results.

compared to the mIoU of U-net, U-net + IIE. However, the loss convergence speed of the network has significantly slowed down (convergence is reached at epoch = 32).

The analysis of Do-Conv implied that the loss convergence speed can be significantly improved by replacing the 2D convolutional layer with the Do-Conv convolution layer in the U-net. Besides, it can also effectively manage the problem of the decrease of network convergence rate caused by the introduction of the IIE module and channel attention module (CAT).

The CAT module demonstrated that the introduction of the CAT module can effectively improve the segmentation accuracy of the network. Additionally, it can dramatically improve the segmentation accuracy when used together with IIE. Simultaneously, the loss convergence speed of the network decreased (loss convergence is reached at epoch = 50).

In summary, the IIE and CAT modules can effectively strengthen the segmentation accuracy of the network, and the Do-Conv layer can highly accelerate the loss convergence speed of the network. Therefore, the three are combined in the U-net network architecture. The experimental results verified that the best segmentation result can be obtained on the basis of retaining a certain loss convergence speed of the network.

4. Conclusions

In this study, an improved U-net network (ICA-Unet) for BAT segmentation has been proposed based on the classic U-net architecture, the image information entropy, channel attention, and Do-Conv modules. The network can learn the PET/CT channel weights from the channel attention modules and enhance the edge features of the maps through IIE modules. After the introduction of these modules, the decrease in loss convergence speed of the network can be mitigated by Do-Conv layers. The network architecture and method we proposed present the mIoU score of 0.832. Compared with other methods, ICA-Unet has significant advantages and avoids the problems of under-segmentation, discontinuity, and threshold difference caused by static threshold segmentation methods, realizing automatic BAT segmentation. The proposed method can assist radiologists and nuclear medicine physicians in efficiently segmenting BAT and significantly facilitate clinicians and researchers to conduct related research on BAT.

Conficts of Interest

The authors declare that there are no conficts of interest relevant to this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61701403, 82122033, 81871379); National Key Research and Development Program of China (2016YFC0103804, 2019YFC1521103, 2020YFC1523301, 2019YFC-1521102); Key R&D Projects in Shaanxi Province (2019ZDLSF07-02, 2019ZDLGY10-01); Key R&D Projects in Qinghai Province (2020-SF-143); China Post-doctoral Science Foundation (2018M643719); Young Talent Support Program of the Shaanxi Association for Science and Technology (20190107).

References

1. B. Cannon, J. Nedergaard, "Brown adipose tissue: Function and physiological significance," *Physiol. Rev.* **84**(1), 277 (2004).
2. N. J. Rothwell, M. J. Stock, "A role for brown adipose tissue in diet-induced thermogenesis," *Obesity* **5**(6), 31–35 (1997).
3. M. Van *et al.*, "Cold-activated brown adipose tissue in healthy men," *N. Engl. J. Med.* **360**(15), 1500–1508 (2009).
4. V. Salem *et al.*, "Glucagon increases energy expenditure independently of brown adipose tissue activation in humans," *Diabetes Obesity Metabol.* **18**, 72–81 (2016).
5. S. Madsbad, A. V. Astrup, "Obesity, the metabolic syndrome and cardiovascular disease," *Ugeskr Laeger.* **166**(17), 1561–1564 (2004).
6. A. M. Cypess *et al.*, "Identification and importance of brown adipose tissue in adult humans," *New Engl. J. Med.* **360**(15), 1509 (2009).
7. R. Boellaard, "Standards for PET image acquisition and quantitative data analysis," *J. Nucl. Med. Official Publication Society of Nuclear Medicine* **50**(1), 11S (2009).
8. R. Hao, L. Yuan, N. Zhang, C. Li, J. Yang, "Brown adipose tissue: Distribution and influencing factors on FDG PET/CT scan," *J. Pediatric Endocrinol. Metabol.* **25**(3–4), 233–237 (2012).

9. J. Nedergaard, T. Bengtsson, B. Cannon, “Unexpected evidence for active brown adipose tissue in adult humans,” *Amer. J. Physiol. Endocrinol. Metabol.* **293**(2), 444–452 (2007).
10. H. Zbib, S. Mouysset, S. Stute, J. M. Girault, C. Tauber, “Unsupervised spectral clustering for segmentation of dynamic PET images,” *IEEE Trans. Nucl. Sci.* **62**(3), 840–850 (2015).
11. K. P. Wong, D. Feng, S. R. Meikle, M. J. Fulham, “Segmentation of dynamic PET images using cluster analysis,” *IEEE Trans. Nucl. Sci.* **3**(1), 200–207 (2002).
12. Y. Venel, H. Garhi, J. L. Baulieu, C. Prunieraesch, A. D. Muret, I. Barillot, “Comparison of six methods of segmentation of tumor volume on the 18F-F. D.G. PET scan with reference histological volume in non small cell bronchopulmonary cancers,” *Med. Nucleaire* **32**(6), 339–353 (2008).
13. J. Ashburner, J. Haslam, C. Taylor, V. J. Cunningham, T. Jones, “A cluster analysis approach for the characterization of dynamic PET data,” *Quantif. Brain Func. PET* **160**(3), 301–306 (1996).
14. S. C. Huang, “Anatomy of SUV,” *Nucl. Med. Biol.* **27**(7), 643–646 (2000).
15. J. Keyes, “SUV: Standard uptake or silly useless value?,” *J. Nucl. Med.* **36**(10), 1836–1839 (1995).
16. K. Chen *et al.*, “Brown adipose reporting criteria in imaging studies (BARCIST 1.0): Recommendations for standardized FDG-PET/CT experiments in humans,” *Cell Metabol.* **24**(2), 210–222 (2016).
17. O. Muzik, T. J. Mangner, W. R. Leonard, A. Kumar, J. Janisse, J. G. Granneman, “15O PET measurement of blood flow and oxygen consumption in cold-activated human brown fat,” *J. Nucl. Med.* **54**(4), 523–531 (2013).
18. S. Baba, H. A. Jacene, J. M. Engles, H. Honda, R. L. Wahl, “CT Hounsfield units of Brown adipose tissue increase with activation: Preclinical and clinical studies,” *J. Nucl. Med.* **51**(2), 246–250 (2010).
19. V. Gilsanz, S. A. Chung, H. Jackson, F. J. Dorey, H. H. Hu, “Functional brown adipose tissue is related to muscle volume in children and adolescents,” *J. Pediatrics* **158**(5), 722–726 (2011).
20. H. C. Shin *et al.*, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *IEEE Trans. Med. Imag.* **35**(5), 1285–1298 (2016).
21. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Springer, Cham, 2015), pp. 234–241.
22. M. Noori, A. Bahri, K. Mohammadi, Attention-guided version of 2D UNet for automatic brain tumor segmentation, *2019 9th Int. Conf. Computer and Knowledge Engineering (ICCKE)* (IEEE, 2019), pp. 269–275.
23. S. M. K. Hasan and C. A. Linte, CondenseUNet: A memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation, *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling. International Society for Optics and Photonics*, 2020, 113151J.
24. S. Moradi *et al.*, “MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography,” *Physica Medica* **67**, 58–69 (2019).
25. R. E. Jurdi, C. Petitjean, P. Honeine, F. Abdallah, “BB-UNet: U-Net with bounding box prior,” *IEEE J. Sel. Top. Signal Process.* **14**(6), 1189–1198 (2020).
26. A. Lou, S. Guan, M. Loew, “DC-UNet: Rethinking the U-Net architecture with dual channel efficient CNN for medical images segmentation,” *Medical Imaging 2021: Image Processing. International Society for Optics and Photonics*, 2021, 115962T.
27. H. B. Shin, H. Sheen, H. Y. Lee, J. Kang, D. K. Yoon, T. S. Suh, “Digital Imaging and Communications in Medicine (DICOM) information conversion procedure for SUV calculation of PET scanners with different DICOM header information,” *Phys. Med.* **44**, 243–248 (2017).
28. S. Ren, K. He, R. Girshick, J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017).
29. A. Yang, X. Jin, L. Li, “CT Images Recognition of Pulmonary Tuberculosis Based on Improved Faster RCNN and U-Net,” *2019 10th Int. Conf. Information Technology in Medicine and Education (ITME)* (IEEE, 2019), pp. 93–97.
30. Q. Li, H. Wang, B. Y. Li, Y. Tang, J. Li, “IIE-SegNet: Deep semantic segmentation network with enhanced boundary based on image information entropy,” *IEEE Access*, **9**, 40612–40622 (2021).
31. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, “CBAM: Convolutional block attention module,” *Eur. Conf. Computer Vision (ECCV)* (2018), pp. 3–19.
32. J. Cao *et al.*, “DO-Conv: Depthwise over-parameterized convolutional layer,” arXiv preprint arXiv:2006.12030, 2020.
33. N. J. A. Sloane and D. W. Aaron A. Wyner, *A Mathematical Theory of Communication* (1993), pp. 5–83.
34. H. Jie, S. Li, S. Gang, S. Albanie, “Squeeze-and-excitation networks,” *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(8), 2011–2023 (2019).
35. A. Paszke, S. Gross, F. Massa, *et al.* “Pytorch: An imperative style, high-performance deep learning

library," *Advances in Neural Information Processing Systems* 32 (2019).

36. S. Ruder, "An overview of gradient descent optimization algorithms," arXiv preprint arXiv:1609.04747, 2016.
37. Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognit.* **29**(8) 1335–1346 (1996).
38. P. D. Huttenlocher, A. G. Klanderman, J. W. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(9) 850–863 (1993).
39. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh *et al.*, UNet++: A Nested U-Net Architecture for Medical Image Segmentation, *4th International Workshop on Deep Learning in Medical Image Analysis, DLMIA 2018 and 8th International Workshop on Multimodal Learning for Clinical Decision Support, ML-CDS 2018 Held in Conjunction with MICCAI 2018* (Springer Verlag, 2018), pp. 3–11.
40. V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).