SEU^2-Net: Multi-Scale U^2-Net with SE attention mechanism for liver occupying lesion CT image segmentation

Lizhuang Liu [1] , **Kun Wu** [1] , **Ke Wang** [2] , **Zhenqi Han** [Corresp., 1] , **Jianxin Qiu** [2] , **Qiao Zhan** [3] , **Tian Wu** [4] , **Jinghang Xu** [4] , **Zheng Zeng** [Corresp. 4]

[1] Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, No.99, Haike Road, Zhangjiang Hi-Tech Park, PuDong, Shanghai, China
[2] Radiology Department, Peking University First Hospital, Beijing, China
[3] Department of Infectious Diseases, The First Affiliated Hospital of Nanjing Medical University, Nanjing, No.300 GuangZhou Street, GuLou District, China
[4] Department of Infectious Diseases, Peking University First Hospital, Beijing, No.8 Xishiku Street, Xicheng District, China

Corresponding Authors: Zhenqi Han, Zheng Zeng
Email address: hanzq@sari.ac.cn, zeng@bjmu.edu.cn

Liver occupying lesion can have significant consequences for a person’s health and wellbeing. To assist physicians in the diagnosis and treatment planning of abnormal areas in the liver, we propose a novel network named SEU^2-Net by introducing the channel attention mechanism into U^2-Net for accurate and automatic liver occupying lesion segmentation. We design the SE-RSU block, which is to add the SE attention mechanism at the residual connections of the RSU (the component unit of U^2-Net). SEU^2-Net not only retains the advantages of U^2-Net in capturing context information at multiple scales, but also can adaptively recalibe channel feature responses to emphasize useful feature information according to the channel attention mechanism. In addition, we present a new abdominal CT dataset for liver occupying lesion segmentation from Peking University First Hospital's clinical data (PUFH dataset). We evaluate the proposed method and compare it with five deep learning networks on the PUFH and LiTS dataset. The experimental results show that SEU^2-Net has state-of-the-art performance and good robustness in liver occupying lesion segmentation.

SEU^2-Net: Multi-Scale U^2-Net with SE attention mechanism for liver occupying lesion CT image segmentation

Lizhuang Liu[1], Kun Wu[1], Ke Wang[2], Zhenqi Han[1], Jianxin Qiu[2], Qiao Zhan[3], Tian Wu[4], Jinghang Xu[4], and Zheng Zeng[4]

[1]**Shanghai Advanced Research Institute, Chinese Academy of Sciences, No.99, Haike Road, Zhangjiang Hi-Tech Park, PuDong, Shanghai, China**
[2]**Radiology Department, Peking University First Hospital, Beijing, China**
[3]**Department of Infectious Diseases, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China No.300 GuangZhou Street, GuLou District, Nanjing City, JiangSu Province, China**
[4]**Department of Infectious Diseases, Peking University First Hospital, Beijing, China**

Corresponding author:
Zhenqi Han, Zheng Zeng

Email address: hanzq@sari.ac.cn, zeng@bjmu.edu.cn

ABSTRACT

Liver occupying lesions can profoundly impact an individual's health and well-being. To assist physicians in the diagnosis and treatment planning of abnormal areas in the liver, we propose a novel network named SEU^2-Net by introducing the channel attention mechanism into U^2-Net for accurate and automatic liver occupying lesion segmentation. We design the SE-RSU block, which is to add the SE attention mechanism at the residual connections of the RSU (the component unit of U^2-Net). SEU^2-Net not only retains the advantages of U^2-Net in capturing context information at multiple scales, but also can adaptively recalibrate channel feature responses to emphasize useful feature information according to the channel attention mechanism. In addition, we present a new abdominal CT dataset for liver occupying lesion segmentation from Peking University First Hospital's clinical data (PUFH dataset). We evaluate the proposed method and compare it with five deep learning networks on the PUFH and LiTS dataset. The experimental results show that SEU^2-Net has state-of-the-art performance and good robustness in liver occupying lesion segmentation.

INTRODUCTION

Liver occupying lesion segmentation is a significant research focus in the field of medical image analysis (Xue et al., 2021). It not only assists doctors in achieving more accurate analysis, diagnosis, and effective treatment planning for liver tumors (Li et al., 2018; Peng et al., 2022) but also fosters scientific research and technological innovation in the field of medical imaging(Li and Ma, 2022). The liver occupying lesion refer to various abnormal structures, tissues, or diseases that appear within the liver tissues, including liver tumors, liver cysts, liver abscesses, and so on. Most research papers on liver occupying lesion segmentation focus primarily on liver tumor segmentation. Liver tumor segmentation is divided into three methods based on gray scale segmentation algorithms, machine learning and deep learning. Based on gray scale segmentation algorithm: Qi et al. (Qi et al., 2008) proposed a semi-automatic segmentation method of CT liver tumor based on Bayesian rule 3D seed region growing (SRG). In the iterative updating process of region growing, Bayesian decision rules and model matching metrics are used as the growth criteria. This method had a good segmentation effect when the intensity difference between tumor and normal tissue is large, but there were large errors in the top and bottom slices in complex. Wong et al. (Wong et al., 2008) proposed a 2D region growing semi-automatic liver tumor segmentation method based on knowledge constraints. The segmentation effect was poor when the occupying lesion tissue and normal liver tissue had low contrast.

Segmentation algorithm based on machine learning: Amita Das et al. (Das and Sabut, 2016) used adaptive threshold, morphological processing and kernel fuzzy C-mean (KFCM) clustering algorithm together with spatial information to segment liver tumor region. The segmented liver tumor image of this method had high peak signal-to-noise ratio and low uniform error value. Munipraveena Rela et al. (Rela et al., 2020) proposed superpixel-based fast fuzzy C-means clustering algorithm (SFFCM) for liver tumor image segmentation to achieve an accuracy of 99.5 %. The segmentation algorithms based on machine learning can only achieve good segmentation results on a single dataset but have no transfer ability.

The method based on deep learning realizes automatic liver tumor segmentation, which overcomes the shortcomings of poor segmentation results when the contrast between tumor and normal liver tissue is low in traditional segmentation algorithms. The method based on deep learning has the ability of transfer learning. There are difficulties in making datasets and a relatively small number of datasets in liver tumor image segmentation based on deep learning. However, the proposal of U-Net (Ronneberger et al., 2015) solves the problem of fast training in a small number of data sets. Many improved models based on U-Net have achieved good results in liver tumor segmentation. Liu et al. (Liu et al., 2019) improved U-Net by increasing the depth of U-Net and only copying pooling layer features during skip-connection. After that, use graph segmentation to optimize the segmentation results. The method smoothed the upper boundary of liver tumor segmentation. Xu et al. (Xu et al., 2020) improved UNet++ and added the residual structure in convolution blocks to avoid the problem of gradient disappearing. The method made the Dice coefficient 93.36%. Seo et al. (Seo et al., 2019) improved the skip connection part of U-Net by adding the residual path with deconvolution layer and activation operation. This method solved the repetition of low-resolution feature information and had a good effect on the segmentation of liver tumor edge and small structure. Li et al. (Li et al., 2020b) added an attention mechanism module to the convolution block of UNet++. This method achieved good segmentation results in the LiTS dataset. However, the complex structure led to slow training and prediction speed.

The deep learning methods discussed above have two main problems: on the one hand, some traditional deep learning models such as UNet and SegNet. (Badrinarayanan et al., 2017) only use single-scale feature maps for segmentation, which cannot effectively capture information at different scales. On the other hand, when dealing with complex tumor contours and small tumors, the effect is not perfect. Other models often have problems such as boundary blur and excessive smoothing, which make it difficult to accurately segment complex tumor contours and small tumors. Based on these considerations, we choose U^2-Net (Qin et al., 2020) with multi-scale feature fusion and SE blocks with channel attention mechanism.

We apply U^2-Net to liver occupying lesion segmentation, which is a new attempt of U^2-Net in the field of medical image processing. In order to better segment small occupying lesion and complex occupying lesion contours, the SE module is introduced into the U^2-Net architecture (Li et al., 2020a; Gong et al., 2022). The SE(Squeeze-and-Excitation) block (Hu et al., 2018) is a channel attention mechanism, which can selectively emphasize informative features and suppress less useful ones by explicitly modeling the interdependencies between their convolutional feature channels. Therefore, our model is termed SEU^2-Net. We propose an abdominal CT image dataset for liver occupying lesion segmentation from Peking University First Hospital's clinical data (PUFH dataset). Use PUFH dataset and LiTS dataset to train SEU^2-Net model. By testing on the PUFH and LiTS dataset, prove the superiority of SEU^2-Net model in liver occupying lesion segmentation by comparing with different models. SEU^2-Net obtains IoU 90.86%, Acc 99.72%, $Kappa$ coefficient 95.07%, $Dice$ coefficient 95.21% on PUFH dataset and IoU 80.81%, Acc 99.83%, $Kappa$ coefficient 89.30%, $Dice$ coefficient 89.39% on LiTS dataset.

MATERIALS AND METHODS

In this section, we design the architecture of SEU^2-Net for liver occupying lesion segmentation to test on the PUFH and LiTS dataset. The Biomedical Research Ethics Committee of Peking University First Hospital approved the study of the PUFH dataset (Ethical Review No. 2020 Scientific Research 101 Amendments). The PUFH dataset did not require consent from study participants. Figure 1 shows that SEU^2-Net which is an encoder-decoder U-shaped structure consists of 11 SE-RSU structures with different stages including six stages encoder and five stages decoder. The saliency map fusion module is after the decoder levels and the last encoder level. The structure of SE-RSU is still a U-shaped encoder-decoder structure. Compared to the architecture of U^2-Net, our innovation is SE-RSU that adds the channel attention mechanism SE block at the residual connection of RSU. In SEU^2-Net, a mix of CrossEntropyLoss() and DiceLoss() loss functions is used instead of the standard binary cross-entropy loss function used in the U^2-Net paper. The hybrid use of the above two loss functions can optimize the prediction accuracy, prediction accuracy and model robustness of liver occupying lesion segmentation.

SEU^2-Net has the attention mechanism and multi-scale feature fusion strategy. For the attention mechanism: SEU^2-Net employs an attention mechanism using SE blocks to adaptively weight the channel features, allowing the model to focus on important feature information.The SE block mainly includes the global pooling and sigmod activation functions. Then, the two operations of scale and residual are used to emphasize the occupying lesion information in liver occupying lesion segmentation, especially the small occupying lesion and complex occupying lesion contour.

For multi-scale feature fusion strategy: Firstly, the SEU^2-Net model adopts multi-scale images input in the encoder. Specifically, it scales the original image to different scales by a certain proportion, and inputs the scaled image into the model for feature extraction and segmentation prediction, whose advantage is that the image features at different scales can be effectively extracted, and the cross-layer fusion between the features at different scales can be performed to further improve the accuracy and robustness of the model. In the decoder, for each decoder layer, SEU^2-Net will obtain the feature map of the corresponding scale from the encoder, and fuse it into the feature map of the current layer through upsampling and deconvolution operations. The feature information of different scales can be used to improve the detection and segmentation ability of the model for objects.

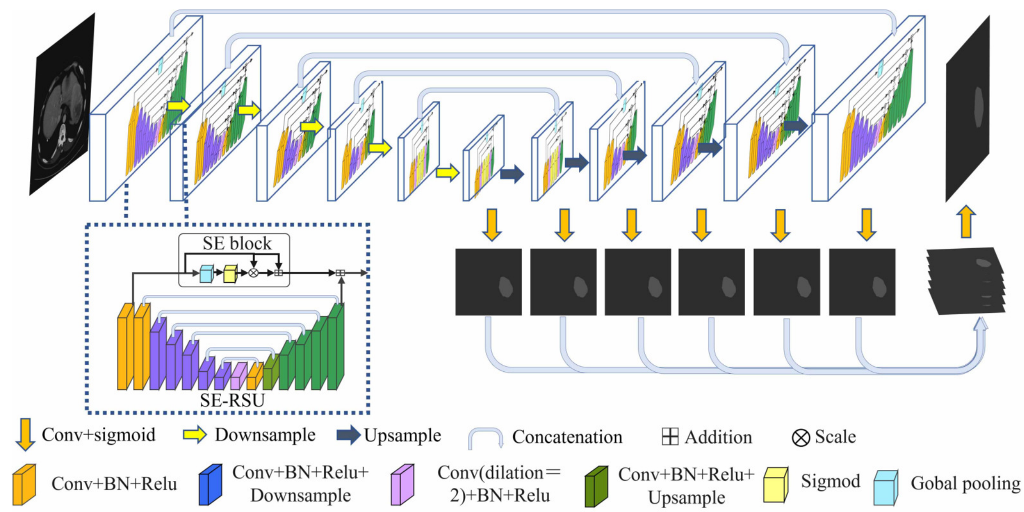

Figure 1. The architecture of SEU^2-Net. SEU^2-Net consists of 11 SE-RSU structures and multi-scale fusion operations. SE-RSU integrates SE attention mechanism into the residual connection of RSU. The input image is a 512×512 abdominal CT slice, and the output image is a 512×512 segmented liver occupying lesion image. (a) SEU^2-Net employs a multi-scale feature fusion strategy to enhance its detection and segmentation capabilities by utilizing feature information from different scales. (b) SEU^2-Net utilizes SE blocks as attention mechanism to adaptively weight channel features, allowing the model to focus on important feature information.

SE Residual U-blocks

SE-RSU (Residual U-block with Squeeze-and-Excitation) is a new structure introduced in U^2-Net, which incorporates the Squeeze-and-Excitation (SE) mechanism into the Residual U-block (RSU) of U^2-Net. The SE Residual U-blocks (SE-RSU) terms that the SE block is added to the position of the residual connection in the RSU. The SE block is an attention mechanism used to enhance the model's feature representation capability by adaptively learning the correlations between feature channels to better capture important feature information. The purpose of adding SE after the first convolutional block of SE-RSU is coarse-grained context detection in the initial layer network and fine-grained context detection in the deep network as SE-RSU goes deeper in the SEU^2-Net network. Figure 2 shows that the SE block is implemented by passing the output of the convolution operation through a Global Pooling layer, and then normalizing it using a Sigmoid function in SE-RSU. Firstly, global spatial information compression into channel descriptors is achieved by using global average pooling to generate channel statistics. i.e., the size $H \times W \times C$ feature is compressed to $1 \times 1 \times C$. Then, the sigmoid activation function is used to learn nonlinear interactions between channels to capture channel dependencies. The output of sigmod function is channel-multiplied with the original feature to highlight the useful feature information and ignore the less useful ones. Finally, residual connection in the SE block is used to avoid gradient vanishing.

Compared to other papers(Tian et al., 2021; Zhang et al., 2022; Fan et al., 2021; Zhang and Zhang, 2021), the SE block in our paper only includes global average pooling and sigmoid activation function, without fully connected layers. The role of fully connected layers is to map the feature vector after global pooling to a lower-dimensional vector. However, the SE block is connected to the first convolutional layer of the SE-RSU, where the shallow convolutional layer and feature extraction result in a smaller depth of the feature map. Using global average pooling and sigmoid activation function directly for feature fusion can meet the demand for adaptive weighting of channel features. On the other hand, not using fully connected layers reduces the number of model parameters and computational complexity, preventing overfitting of the model.

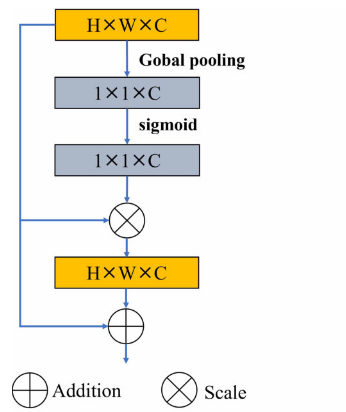

Figure 2. The SE block. The SE block includes global pooling and sigmod activation operations, which is added to the residual connection of the SE-RSU.

Loss functions

Compared to the U^2-Net paper which uses the standard binary cross-entropy loss function, we use a combination of CrossEntropyLoss() and DiceLoss() in our approach. The DiceLoss() function is primarily used to optimize the model's prediction accuracy and robustness, particularly for enhancing the prediction precision and robustness of the model. DiceLoss() is used to optimize the model's prediction precision and robustness. By using these two loss functions together, the model can find a balance between accuracy and precision and has better generalization ability. However, the binary cross entropy loss function is sensitive to pixel class distribution imbalance, which can cause the model's prediction results to be biased towards regions with more pixel classes, and can result in misclassification in regions with fewer pixel classes. Additionally, the binary cross entropy loss function cannot handle the problem of target boundaries well, which can lead to prediction results with breaks or blur at the boundaries, affecting the segmentation performance of the model. Therefore, the binary cross entropy loss function is not effective for small or complex occupying lesion contours in liver occupying lesion segmentation.

Choose to use a mix of CrossEntropyLoss() and DiceLoss() loss functions with corresponding weights of 0.7 and 0.3. Equation 1 is the loss function expression of CrossEntropyLoss(). N represents the number of samples, C represents the number of classes, y_{ij} represents the j-th label of i-th sample, and p_{ij} represents the predicted probability of the j-th class for i-th sample. The cross-entropy loss function measures how close the predicted value is to the ground truth.

$$CrossEntropyLoss() = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} * log(p_{ij}) \tag{1}$$

Equation 2 is the expression of the loss function for DiceLoss(), where Dice is given by Equation 6. DiceLoss() is used when the background area is much larger than the target area. Because the liver occupying lesion area is smaller than the liver background area, DiceLoss() will have a better effect.

$$DiceLoss() = 1 - \frac{|A \cap B|}{|A| + |B|} \tag{2}$$

Evaluation Metrics

Iou refers to the ratio between the intersection and union of the true value and the predicted value in a class. The formula is as shown in equation 3. In our paper, TP represents the intersection of the predicted liver occupying lesion and labeled occupying lesion. $TP+FP+FN$ represents the union of the predicted liver occupying lesion and labeled occupying lesion.

$$Iou = \frac{TP}{TP+FP+FN} \tag{3}$$

Acc(Accuracy) is the overall classification accuracy, which is the probability of predicting the correct number of samples over the total number of samples. Acc can be expressed in equation 4 using confusion matrices. $intersect\_area$ represents the intersection area of prediction and ground truth on all classes. All classes in the experiment refers to the liver space-occupying lesion and background. $pred\_area$ represents the prediction area on all classes.

$$Acc = \frac{intersect\_area}{pred\_area} \tag{4}$$

The $Kappa$ coefficient measures the effect of whether the predicted value is consistent with the actual classification value. The $Kappa$ coefficient compensates for the bias towards large categories in Acc caused by the gap in the number of categories. Equation 5 is the calculation method of $Kappa$ coefficient, where P_e represents the ratio of the product of the number of true categories and the number of corresponding predicted categories to the square of the total number of samples. A Kappa coefficient between 0.81 and 1 indicates almost perfect agreement.

$$Kappa = \frac{Acc - P_e}{1 - P_e} \tag{5}$$

The $Dice$ coefficient represents the proportion of duplicate parts between the predicted segmented image and the annotated image. $Dice$ has a value between 0 and 1. In equation 6, A represents the pixel value of the true image and B represents the pixel value of the predicted image.

$$Dice = \frac{|A \cap B|}{|A| + |B|} \tag{6}$$

Process of experiment

The datasets were divided into training set, test set and verification set in a ratio of 8:1:1. Mix CrossEntropyLoss() and DiceLoss() loss functions. The Momentum optimizer which contains the Newtonian momentum flag can better eliminate the wobble phenomenon in the process of updating the hyperparameter. Use Cosine annealing methods to dynamically adjust the learning rate shown in equation 7.

$$\eta_t = \eta^i_{min} + \frac{1}{2}(\eta^i_{max} - \eta^i_{min})(1 + \cos\frac{T_{cur}\pi}{T_i}) \tag{7}$$

Let i denote the number of learning rate changes at the $i_t h$ time. The initial learning rate η^i_{max} is 0.0015. The minimum learning rate η^i_{min} is 0. T_{cur} is the current epoch. T_i is the number of iterations in one learning rate cycle, which is set to 5400. The batch size is 4. The model is saved every 900 training runs, for a total of 9000 training runs. The experiments are trained on NVIDIA Tesla V100 GPU(32GB).

RESULTS

PUFH dataset

dataset setup

The PUFH dataset consisted of 200 3D abdominal CT images and corresponding labeled images, which labeled various types of liver occupying lesions such as hepatic cysts, liver abscess, and hepatocellular carcinoma. The number of slices that can be scanned at a time ranges from 39 to 107 depending on the information of each 3D image when the dataset is sliced horizontally. The resolution of slice images ranged from [0.5839844, 0.5839844, 5.0] to [0.88671875, 0.88671875, 5.0]mm, and the image intensity ranges from [-1024.0, 3071.0] to [-3024.0, 1210.0]. The slice size is 512×512. Figure 3 shows the fusion slice image of a random abdominal CT image and its corresponding annotated image in the PUFH dataset.

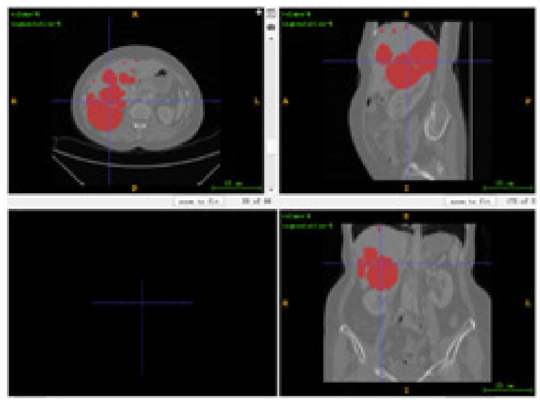

Figure 3. A random fusion image from PUFH dataset.

dataset processing

The dataset is abdominal CT data containing multiple organs. In order to ensure the accuracy of model training, find the slice data of the original image from the beginning to the end of the liver according to the labeled image data. The Nii format data containing only the liver region is converted into 2D data in PNG format, and the size of the converted 2D data is 512×512.

Experimental results

The training results of U-Net, AttentionU-Net(Oktay et al., 2018), UNet3Plus(Huang et al., 2020) ,UNetPlusPlus(Zhou et al., 2020) and U^2-Net are compared with SEU^2-Net in the PUFH dataset. Table 1 shows the evaluation metrics of PUFH dataset in five models. SEU^2-Net outperforms for liver CT image segmentation in five methods, achieving the IoU ratio increased 0.75%, the Acc increased 0.04%, the $Kappa$ coefficient increased 0.44%, the $Dice$ coefficient increased by 0.41% over U^2-Net. Compared with AttentionU-Net, which is the combination of attention mechanism and UNet, SEU^2-Net had the highest improvement, among which the Iou ratio increased by 8.8%, the Acc increased by 0.36%, the $Kappa$ coefficient increased by 5.26% and the $Dice$ coefficient increased by 5.07%. Figure 4 shows that Six 2D slices are randomly selected in the dataset to observe the liver occupying lesion segmentation effect of the five models. The effect of liver occupying lesion segmentation depends on the location of segmentation, the overall contour and the recognition of small liver occupying lesion. SEU^2-Net is most similar to label in terms of the location and size of liver occupying lesion. For example, the liver occupying lesion in the fifth slice has complex contours and fine regions. Comparing the segmentation effects of the six models in the fifth slice, SEU^2-Net can identify very small liver occupying lesion and complex edges (the mark of the red box), which greatly illustrates the importance of SEU^2-Net's attention mechanism and its ability to capture feature information from multiple scales perspectives.

LiTS dataset

dataset setup and processing

As liver tumors are among the different types of liver occupying lesions, the LiTS dataset consists of 131 contrast-enhanced 3D abdominal CT scans, including 70 scans for training and testing with corresponding

	IoU(%)	Acc(%)	Kappa(%)	Dice(%)
U-Net	86.85	99.54	92.72	92.96
AttentionU-Net	82.06	99.36	89.81	90.14
UNet3Plus	65.10	98.53	78.10	78.86
UNetPlusPlus	81.92	99.42	89.76	90.06
U^2-Net	90.11	99.68	94.63	94.80
SEU^2-Net	**90.86**	**99.72**	**95.07**	**95.21**

Table 1. Evaluation metrics table of PUFH dataset

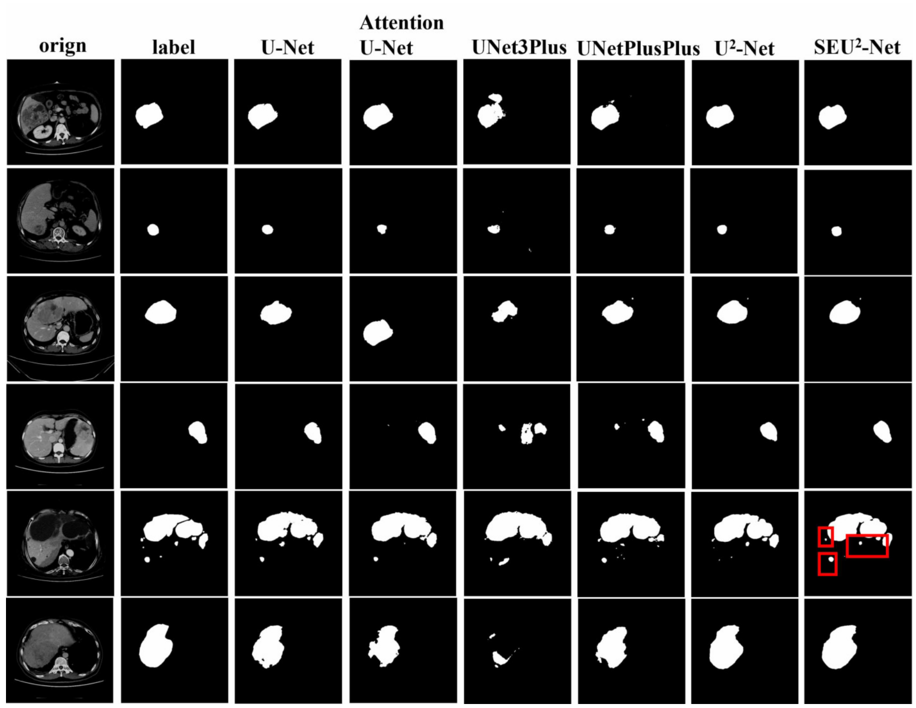

Figure 4. Liver occupying lesion segmentation of PUFH dataset. Column 1: Origin. Column 2: Label. Column 3: U-Net. Column 4: AttentionU-Net. Column 5: UNet3Plus. Column 6: UNetPlusPlus. Column 7: U^2-Net. Column 8: SEU^2-Net.

annotated images specifically focusing on liver occupying lesion segmentation. The number of tumors varied from 0 to 75, and the size varied from $38mm^3$ to $349mm^3$. The layer spacing (section thickness) is $[0.45,6]mm$. The number of slices that can be scanned at a time ranges from 42 to 1026 according to the different information of each 3D image. The size of 2D slices is 512×512, where the planar resolution is $[0.6 \times 0.6, 1.0 \times 1.0]mm$.

For the processing of the LiTS dataset, only the tumor label is retained during liver tumor segmentation because the dataset contained both liver and tumor labels. Then the processing of the dataset is the same as that of the PUFH dataset. Figure 5 shows the fusion slice image of a random abdominal CT image and its corresponding annotated image in the LiTS dataset.

Experimental results

The six models trained on the PUFH dataset are used for training and testing in the Lits dataset. The evaluation metrics of LiTS dataset are shown in Table 2. Compared with U^2-Net, SEU^2-Net has the Iou increased by 0.14%, $Kappa$ coefficient increased by 0.09%, and $Dice$ coefficient increased by 0.09% in the four evaluation metrics. UNet3Plus performs the worst among the four evaluation metrics. Compared with UNet3Plus, the Iou of SEU^2-Net is increased by 17.55%, Acc is increased by 0.2%, $Kappa$ coefficient is increased by 11.99%, and $Dice$ coefficient is increased by 11.90%. At the same time, the evaluation metrics of SEU^2-Net on LITS dataset are still higher than those of AttentionU-Net. Figure 6 shows that

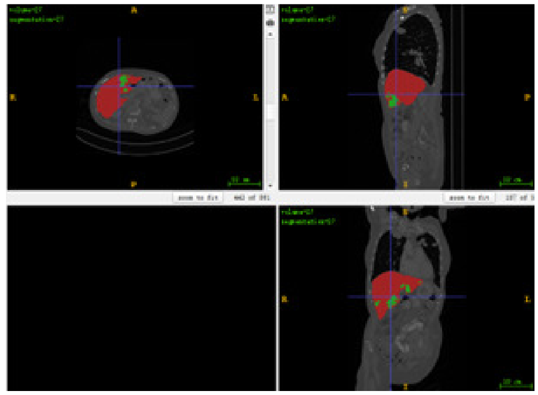

Figure 5. A random fusion image from LiTS dataset.

randomly select six 2D slices in the LiTS dataset to observe the liver tumor segmentation effect of the five models. Most of liver tumors in these 6 pictures are small, and SEU^2-Net is most similar to label in terms of the location and size of liver tumors, which proves that SEU^2-Net has advantages in the segmentation of small tumors.

	IoU(%)	Acc(%)	Kappa(%)	Dice(%)
U-Net	75.57	99.78	85.98	86.08
AttentionU-Net	75.98	99.78	86.24	86.35
UNet3Plus	63.26	99.63	77.31	77.49
UNetPlusPlus	73.46	99.75	84.57	84.70
U^2-Net	80.67	99.83	89.21	89.30
SEU^2-Net	**80.81**	**99.83**	**89.30**	**89.39**

Table 2. Evaluation metrics table of LiTS dataset

DISCUSSION

Automatic segmentation of liver occupying lesion plays an important role in the clinical diagnosis and treatment of liver diseases. Accurate location and size of liver occupying lesion it provides greatly improves the efficiency of diagnosis. In this paper, we propose the SEU^2-Net deep learning model, which adds the SE block with channel attention mechanism to the U^2-Net model that can capture context information from multiple scales. In order to verify the robustness of SEU^2-Net and its superior performance in liver occupying lesion segmentation, we use SEU^2-Net model to train and test on the PUFH and LiTS dataset. Compared to U-Net, AttentionU-Net, UNet3Plus, UNetPlusPlus and U^2-Net, table 1 and table 2 show that SEU^2-Net achieved superior results in terms of IoU, Acc, $Kappa$ coefficient and $Dice$ coefficient, indicating that SEU^2-Net can more accurately locate and segment liver lesions in the liver occupying lesion segmentation task, while maintaining low error and misclassification rates. Specifically, the improvement of SEU^2-Net compared to U^2-Net is mainly concentrated in terms of accuracy and stability, which suggests that the success of SEU^2-Net is closely related to the performance of the SE block. Figure 4 and figure 6 show that SEU^2-Net performs very well in segmenting small and complex liver occupying lesion, which can be attributed to the powerful characteristics of the SE block. The SE block can adaptively adjust the weights of each channel in the feature map to better capture features at different levels. When segmenting small and complex liver lesions, the model requires higher resolution and better feature expression ability, and the introduction of the SE block can effectively improve the feature expression ability of the model, thereby capturing these tiny yet important details. Therefore, SEU^2-Net performs better in the segmentation of small and complex liver occupying lesion. At the same time, by introducing the SE block, the model can better avoid overfitting, improve its generalization ability, and make its performance more stable and reliable on different datasets.

SEU^2-Net performs differently on two datasets. It shows excellent performance on the PUFH dataset, achieving high IoU, Acc,$Kappa$ coefficient, and $Dice$ coefficient. However, its performance on the LiTS

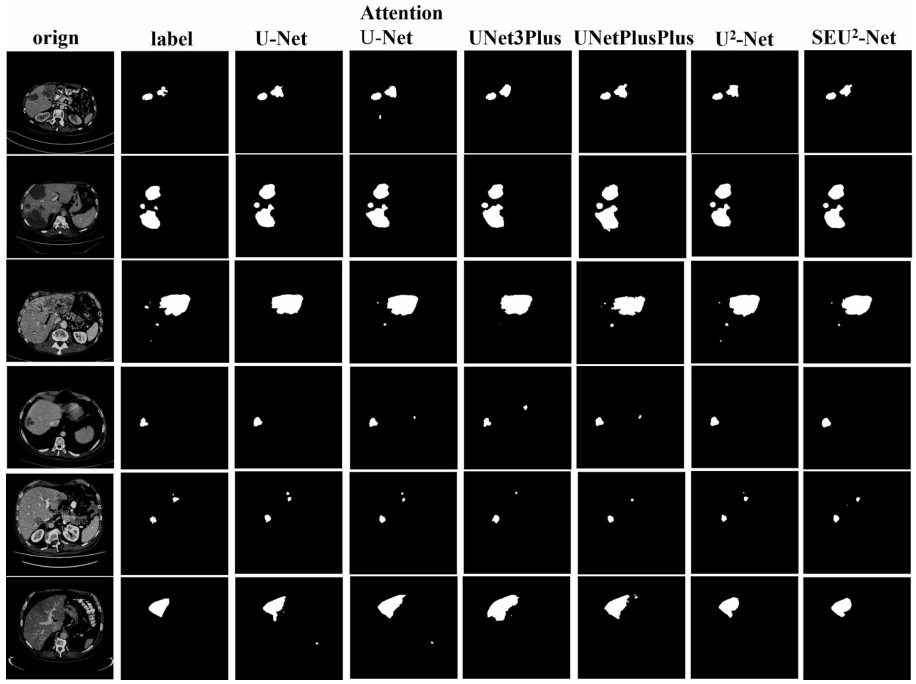

Figure 6. Liver occupying lesion segmentation of LiTS dataset. Column 1: Origin. Column 2: Label. Column 3: U-Net. Column 4: AttentionU-Net. Column 5: UNet3Plus. Column 6: UNetPlusPlus. Column 7: U^2-Net. Column 8: SEU^2-Net.

dataset is relatively lower. This may be due to the differences in characteristics and backgrounds between the two datasets. The PUFH dataset includes various liver occupying lesion such as liver tumors, liver cysts, and liver abscesses, which have relatively obvious shape and texture features, making them easier to be captured and distinguished by the model. On the other hand, the liver lesions in the LiTS dataset only include liver tumors, some of which have very small volumes and may be affected by randomness and noise interference, leading to a slight decrease in the model's performance on the LiTS dataset. Overall, SEU^2-Net's performance on the LiTS dataset is relatively poor, indicating that there is still room for improvement in SEU2Net's generalization ability when handling complex datasets.

Although SEU^2-Net performs well, there are still some limitations in this study. One limitation is that the dataset used only focuses on liver occupying lesions. Therefore, it is necessary to evaluate the performance of SEU^2-Net on larger and more diverse datasets in future research. Another limitation is that we have not conducted a thorough analysis to assess the contribution of the SE block to the performance of SEU2Net. In the future, we plan to study the effects of different SE block configurations and compare them with other attention mechanisms such as CBAM(Woo et al., 2018) and BAM(Park et al., 2018).

CONCLUSIONS

Our paper propose a network architecture called SEU^2-Net based on improved U^2-Net for liver occupying lesion segmentation. SEU^2-Net combines the SE block and U^2-Net, which retains the ability of U^2-Net to capture context information at multiple scales, and emphasizes useful information and ignores useless information through the channel attention mechanism. In addition, we present a new liver occupying lesion CT dataset from Peking University First Hospital's clinical data (PUFH dataset). SEU^2-Net is compared with U-Net, AttentionU-Net, UNet3Plus ,UNetPlusPlus and U^2-Net for liver occupying lesion segmentation on LiTS and PUFH datasets. The Iou, Acc, $Kappa$ coefficient and $Dice$ coefficient of SEU^2-Net are 90.86%, 99.72%, 95.07%, 95.21% and 80.81%, 99.83%, 89.30%, 89.39% on PUFH and LiTS datasets. The experiment shows that SEU^2-Net with attention mechanism outperforms AttentionU-Net with attention mechanism. SEU^2-Net can predict small liver occupying lesion and complex contours from the liver occupying lesion segmentation images of the two datasets. Compared with the other five U-Net series networks, the accuracy, repeatability and consistency between the liver occupying lesion segmentation images predicted by SEU^2-Net and the label images are the best. Our proposed method is an attempt of U^2-Net in the field of medical image processing. The superior performance of SEU^2-Net with attention mechanism in liver occupying lesion segmentation indicates that it has good development potential in the field of medical image processing.

ACKNOWLEDGMENTS

This research was funded by the National Natural Science Foundation of China (Grant No. 61972007) and Science and technology service network plan of Dongwan Chinese Academy of Sciences(STS) (Grant No.20211600200102).

REFERENCES

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis Machine Intelligence*, pages 1–1.

Das, A. and Sabut, S. K. (2016). Kernelized fuzzy c-means clustering with adaptive thresholding for segmenting liver tumors. *Procedia Computer Science*, 92:389–395.

Fan, Y., Yao, Y., and Joe-Wong, C. (2021). Gcn-se: Attention as explainability for node classification in dynamic graphs. In *2021 IEEE International Conference on Data Mining (ICDM)*, pages 1060–1065.

Gong, M., Zhao, B., Soraghan, J., Di Caterina, G., and Grose, D. (2022). Hybrid attention mechanism for liver tumor segmentation in ct images. In *2022 10th European Workshop on Visual Information Processing (EUVIP)*, pages 1–6.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., and Wu, J. (2020). Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE.

Li, C., Tan, Y., Chen, W., Luo, X., Gao, Y., Jia, X., and Wang, Z. (2020a). Attention unet++: A nested attention-aware u-net for liver ct image segmentation. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 345–349.

Li, C., Tan, Y., Chen, W., Luo, X., and Wang, Z. (2020b). Attention unet++: A nested attention-aware u-net for liver ct image segmentation. In *2020 IEEE International Conference on Image Processing (ICIP)*.

Li, L. and Ma, H. (2022). Rdctrans u-net: A hybrid variable architecture for liver ct image segmentation. *SENSORS*, 22(7).

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., and Heng, P.-A. (2018). H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Transactions on Medical Imaging*, 37(12):2663–2674.

Liu, Z., Song, Y. Q., Sheng, V. S., Wang, L., Jiang, R., Zhang, X., and Yuan, D. (2019). Liver ct sequence segmentation based with improved u-net and graph cut. *Expert Systems with Application*, 126(JUL.):54–63.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Park, J., Woo, S., Lee, J., and Kweon, I. S. (2018). BAM: bottleneck attention module. In *British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018*, page 147. BMVA Press.

Peng, Q., Yan, Y., Qian, L., Suo, S., Guo, Y., Xu, J., and Wang, Y. (2022). Liver tumor segmentation and classification using flas-unet plus plus and an improved densenet. *TECHNOLOGY AND HEALTH CARE*, 30(6):1475–1487.

Qi, Y., Wei, X., Leow, W. K., Qi, T., and Wang, S. C. (2008). Semi-automatic segmentation of liver tumors from ct scans using bayesian rule-based 3d region growing. *Region Growing the Midas Journal Grand Challange Liver Tumor Segmentation Miccai Workshop*.

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., and Jagersand, M. (2020). U2-net: Going deeper with nested u-structure for salient object detection. *Pattern recognition*, 106:107404.

Rela, M., Nagaraja, S., and Ramana, P. (2020). Liver tumor segmentation using superpixel based fast fuzzy c means clustering. *International Journal of Advanced Computer Science and Applications*, 11(11).

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Seo, H., Huang, C., Bassenne, M., Xiao, R., and Xing, L. (2019). Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images. *IEEE Transactions on Medical Imaging*.

Tian, H., Ji, B., Quan, W., and Qin, J. (2021). Mpa-net: Multi-scale pyramid attention network for liver tumor segmentation. In *2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*, pages 658–661.

Wong, D., Liu, J., and Fengshou, Y. (2008). A semi-automated method for liver tumor segmentation based on 2d region growing with knowledge-based constraints.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., editors, *Computer Vision – ECCV 2018*, pages 3–19, Cham. Springer International Publishing.

Xu, P., Chen, C., Wang, X., Li, W., and Sun, J. (2020). Roi-based intraoperative mr-ct registration for image-guided multimode tumor ablation therapy in hepatic malignant tumors. *IEEE Access*, PP(99):1–1.

Xue, Z., Li, P., Zhang, L., Lu, X., Zhu, G., Shen, P., Ali Shah, S. A., and Bennamoun, M. (2021). Multi-modal co-learning for liver lesion segmentation on pet-ct images. *IEEE Transactions on Medical Imaging*, 40(12):3531–3542.

Zhang, H. and Zhang, S. (2021). A yolov5s-se model for object detection in x-ray security images. In *2021 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pages 626–631.

Zhang, X., Li, J., and Hua, Z. (2022). Mrse-net: Multiscale residuals and se-attention network for water body segmentation from satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:5049–5064.

Zhou, Z., Siddiquee, M., Tajbakhsh, N., and Liang, J. (2020). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867.