

# ELTS-Net: An Enhanced Liver Tumor Segmentation Network with Augmented Receptive Field and Global Contextual Information

Xiaoyue Guo, Zidong Wang, Peishu Wu, Yurong Li, Fuad E. Alsaadi, and Nianyin Zeng\*

**Abstract**—The liver is one of the organs with the highest incidence rate in the human body, and late-stage liver cancer is basically incurable. Therefore, early diagnosis and lesion location of liver cancer are of important clinical value. This study proposes an enhanced network architecture ELTS-Net based on the 3D U-Net model, to address the limitations of conventional image segmentation methods and the underutilization of image spatial features by the 2D U-Net network structure. ELTS-Net expands upon the original network by incorporating dilated convolutions to increase the receptive field of the convolutional kernel. Additionally, an attention residual module, comprising an attention mechanism and residual connections, replaces the original convolutional module, serving as the primary components of the encoder and decoder. This design enables the network to capture contextual information globally in both channel and spatial dimensions. Furthermore, deep supervision modules are integrated between different levels of the decoder network, providing additional feedback from deeper intermediate layers. This constrains the network weights to the target regions and optimizing segmentation results. Evaluation on the LiTS2017 dataset shows improvements in evaluation metrics for liver and tumor segmentation tasks compared to the baseline 3D U-Net model, achieving 95.2% liver segmentation accuracy and 71.9% tumor segmentation accuracy, with accuracy improvements of 0.9% and 3.1% respectively. The experimental results validate the superior segmentation performance of ELTS-Net compared to other comparison models, offering valuable guidance for clinical diagnosis and treatment.

**Index terms**— 3D convolutional neural network, attention mechanism, deep supervision, residual connection, liver tumor segmentation.

This work was supported in part by the Natural Science Foundation of China under Grant 62073271, the Fundamental Research Funds for the Central Universities of China under Grant 20720220076, the Natural Science Foundation for Distinguished Young Scholars of the Fujian Province of China under Grant 2023J06010, and the National Science and Technology Major Project of China under Grant J2019-I-0013-0013.

X. Guo is with the College of Engineering, Peking University, Beijing 100871, China, and also with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China.

Z. Wang is with the Department of Computer Science, Brunel University, London, Uxbridge UB8 3PH, U.K. Email: [zidong.wang@brunel.ac.uk](mailto:zidong.wang@brunel.ac.uk)

P. Wu and N. Zeng are with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China. Email: [zny@xmu.edu.cn](mailto:zny@xmu.edu.cn)

Y. Li is with the College of Electrical Engineering and Automation, Fuzhou University, Fujian 350116, China, and also with the Fujian Key Lab of Medical Instrumentation & Pharmaceutical Technology, Fujian 350116, China.

F. E. Alsaadi is with the Communication Systems and Networks Research Group, Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

\*Corresponding author.

## I. INTRODUCTION

The liver is a vital metabolic organ in the human abdomen, but liver cancer, a pathological condition affecting the liver, has become one of the most prevalent cancers worldwide [2]. Early stage liver cancer usually has no obvious symptoms, so early diagnosis and treatment of liver tumors become very important, of which resection is the more effective means. Physicians often rely on Computed Tomography (CT) scan results to segment the liver and its tumors. Correct segmentation of liver tumors facilitates the development of treatment plans by physicians, thereby dramatically improving diagnostic accuracy and treatment outcomes. However, manual annotation and segmentation of tumors pose significant challenges in clinical practice due to the diverse morphology of the liver and tumors, similarity in grayscale with adjacent tissues, image noise. Firstly, the liver of each individual has different shapes and shows different morphology in CT images; tumors are of various shapes, sizes, and numbers, with significant differences between different patients, and even between different tumors in the same patient. Secondly, the liver and adjacent tissues in CT images have similar grayscales, and there are large connecting areas, which are difficult to be accurately segmented. In addition, the complex textural differences and dense noise interference between the liver and the tumor in the images make it difficult to locate and identify the boundaries. Finally, the instability of CT techniques also affects the accuracy of segmentation results, such as localized body effects, noise, artifacts, and tissue motion. These challenges make the current manual segmentation of liver tumors very limited in both accuracy and efficiency. There is also the problem of high subjectivity of manual segmentation and poor reproducibility of segmentation.

To address these issues, both domestic and international researchers have conducted extensive studies in the field of liver tumor segmentation. Research methods can be categorized into three types: traditional image processing techniques, machine learning-based approaches, and deep learning-based image segmentation. Traditional medical image segmentation techniques have reached a certain level of maturity, but they have limitations in handling complex scenarios, multi-modal images, and images with high levels of noise [22]. Machine learning-based segmentation methods are still constrained by feature selection and model generalization. In contrast, deep learning techniques can adaptively extract features, and through end-to-end learning and continuous parameter tuning,

excellent feature representations can be learned from a large amount of data, thus effectively improving the performance of various segmentation tasks. In addition, deep learning has good generalization ability, which can well solve the low nonlinear performance problem of shallow neural networks. Therefore, automatic segmentation methods based on deep learning have a wide range of application prospects and important clinical significance, and are now widely used in the field of image segmentation [1].

According to the characteristics of medical image data, deep learning-based liver tumor segmentation networks are usually divided into 2D, 2.5D and 3D networks. Among them, 2D network has lower requirements on equipment performance, but the segmentation accuracy is not high; 2.5D network can take into account the information of multiple slices at the same time, but its processing needs to stitch multiple slices into a 3D volume, the integration is not mature enough, and at the same time, it increases the computational complexity of the network and the storage requirements; since most of the medical images are three-dimensional, compared with 2D networks, 3D networks are able to take into account both the spatial information of the image and the relevant characteristics of the cuts, which can capture the target morphology and features more completely and perform more accurate segmentation. 3D network does not need to slice the image of the 3-dimensional image operation, and it can make better use of the image's interlayer information, and the segmentation effect is overall better compared to 2D network, but it still faces the problems that the data volume of medical images is too small, the features can not be extracted sufficiently, and the segmentation accuracy can not be balanced with the network computational cost. Due to the characteristics of abdominal CT images such as noise interference, small size of liver tumors, strong heterogeneity, and fuzzy structural boundaries, the current segmentation accuracy of liver tumor segmentation algorithms based on deep learning is still low, and it is not suitable for direct use in clinical diagnosis and treatment. Therefore, this study focuses on improving the expressiveness and segmentation accuracy of deep learning-based liver tumor segmentation algorithms. Based on the above discussion, an enhanced liver tumor segmentation network (ELTS-Net) based on 3D U-Net is developed in this paper. In particular, by designing modules to enable the network with multi-scale feature extraction and multimodal information fusion, the proposed ELTS-Net can effectively learn multi-dimensional features and irregular features of complex structures. Meanwhile, the developed ELTS-Net network also has certain robustness and generalization performance, which can be better applied in the field of image segmentation. The major contributions of this article can be summarized as follows:

- 1) The ELTS-Net based on 3D U-Net is developed for CT images, which can simultaneously segment the liver and its tumors.
- 2) In the proposed ELTS-Net, dilated convolution is employed for expanding the receptive field, residual connection is adopted to learn more complex and abstract representations, and spatial-channel attention mechanism

is developed for capturing relevant features and suppressing unimportant information.

- 3) Additional supervisory signals are incorporated in ELTS-Net, where by adding extra depth feedback on shallower intermediate layers, the deep supervision of the feature map in the multi-layer CNN is realized. Furthermore, the experimental results have substantiated the effectiveness of the proposed enhancements in liver and tumor segmentation tasks.

The rest of this article is organized as follows. Section II introduces the literature review and research status of related content. Section III elaborates on the details of the proposed ELTS-Net framework and its components. Section IV presents substantive experimental validation and comprehensive discussions. Finally, section V draws conclusions with future prospects.

## II. RELATED WORKS

In this section, relevant deep learning-based methods for medical image segmentation are reviewed. Since it is important to employ attentional mechanisms in image classification tasks to change the level of attention of deep learning models to different input parts and thus reduce segmentation errors, a brief overview of representative attentional mechanisms is also provided.

### A. Medical image segmentation

In the research of medical image segmentation using deep learning algorithms, Convolutional Neural Networks (CNNs) have been widely researched and applied due to their local sensing and end-to-end feature extraction capabilities [20]. Hinton et al. [16] proposed the AlexNet network, which exhibited outstanding performance in image classification tasks. Long et al. [27] replaced fully connected layers with convolutional layers to develop the fully convolutional neural network (FCN) for semantic segmentation of images. FCN has been extensively used for liver and tumor segmentation from volumetric images. Sun et al. [38] proposed the multi-channel fully convolutional network (MC-FCN), which trained images at different stages through three channels in the network and fused features at different stages, successfully achieving accurate segmentation of liver tumors. Ronneberger et al. [36] introduced the U-Net architecture, which incorporates both encoding and decoding functionalities in a symmetrical structure. U-Net has been widely adopted for liver and liver tumor segmentation. Lin et al. [25] evaluated the C-Means algorithm in U-Net and achieved good results. Zhou et al. [51] proposed a new network architecture U-Net++ based on U-Net based on nested and dense skip connections. U-Net++ uses stacked skip connections and thick skip connections to combine the encoder and the feature maps of the decoder are combined to solve the problem of semantic gap between the encoder and decoder feature maps. Xu et al. [47] added a residual structure to the U-Net++ network structure and used this network to segment the liver.

Since most medical images are 3D images, in order to make full use of the inter-layer information of the image, domestic

and foreign scholars have conducted in-depth research on the 3D network structure. The 3D network does not need to perform slicing operations on 3D images and can directly use 3D convolution. The overall segmentation effect is better than that of the 2D network. Arnab et al. proposed the V-Net network structure. This network uses three-dimensional convolution and pooling operations, which can retain the information between CT image slices and improve the segmentation effect. Reza et al. [35] used U-Net, V-Net, and FPN (Feature Pyramid Network) to segment the liver respectively. As a result, the segmentation effect of the FPN network was better than that of U-Net and V-Net. Milletari et al. [29] applied the V-Net network to liver segmentation, where the encoder extracts global features of the liver from CT images, and the decoder produces full-resolution outputs. Lei et al. [19] proposed the lightweight LV-Net network as a solution to the high computational complexity and memory consumption of the V-Net model. LV-Net reduces memory usage while maintaining segmentation accuracy. Chen and his team [5] proposed a Feature-fusion Encoder-Decoder Network (FED-Net), which achieves the fusion of high-resolution and low-resolution features of images by using an attention mechanism, while using residual structure and dense upsampling to reduce information loss during the upsampling process. Jin et al. [15] proposed a residual attention-aware segmentation method (RA-Unet). This network uses residual blocks to replace traditional convolution blocks, and applies an attention residual mechanism in the skip connection part, thereby integrating low-level It is combined with high-level feature maps to extract contextual information and then used for liver tumor segmentation. Jeong [14] and others proposed the Deep 3D attention U-Net network. The attention mechanism module used in this network can learn liver structures of different shapes and sizes to achieve more refined liver segmentation. However, the structural design of these models is not efficient enough, the robustness and reliability are lacking, and the segmentation accuracy is not high enough.

Based on the above analysis, we propose an ELTS-Net architecture that further optimizes segmentation results and enhances the accuracy of semantic segmentation.

### B. Attention mechanism

The attention mechanism, widely used in deep learning models. By adjusting the weights of each feature in the model, attention mechanisms enable the model to concentrate on the most relevant information for the current task, thereby enhancing the feature extraction capability [39] [50]. Oktay [30] incorporated attention gates into the skip connections of U-Net to control the importance of different features. Hu et al. [12] introduced Squeeze-and-Excitation Networks (SENet), which calibrates the importance of different channel features, enhancing effective feature channels. Roy [37] proposed a concurrent Spatial and Channel Squeeze and Excitation block (scSE-block) based on SENet, which has shown promising results in image segmentation. Woo et al. [43] introduced a Convolutional Block Attention Module (CBAM), which combines spatial and channel attention mechanisms to weight

the feature maps adaptively, allowing the network to focus on important feature information. Park et al. [32] proposed a bottleneck attention module, which can be integrated with any feedforward convolutional neural network and can infer attention maps along two different paths, channel and space. Wang [42] proposed a channel attention module, which is inserted into the skip connection between the encoder and the decoder, which can improve the performance of medical image segmentation. Zhang et al. [49] introduced scale and axis attention mechanisms and verified that they can effectively capture basic information in global pooling. Wang et al. [41] proposed the Non-Local attention mechanism, which expands the receptive field of the network by stacking convolutional layers, introduces global information, and improves accuracy. However, the required parameters are large and require high computer memory and performance. Huang et al. [13] proposed the Cross-Cross Attention module (CCNet), which allows each pixel to capture the long-term dependence of all pixels on it. Compared with Non-Local attention, it takes up less memory and has better performance. High computational efficiency.

In this work, we incorporated SE and CBAM attention modules into the downsampling and upsampling stages of the 3D U-Net network. This allows the network better focus on the spatial information of the target region and the relative importance of feature channels [45] and achieve more precise image segmentation.

## III. METHODOLOGY

In this section, we will provide a detailed description of the improved network architecture. The developed ELTS-Net is based on the 3D U-Net structure and incorporates dilated convolutions, residual connections, and various attention mechanisms. Additionally, a deep supervision mechanism is employed in the decoder to facilitate the training of features at each scale, thereby enhancing the segmentation performance. To start with, the overall framework of ELTS-Net is illustrated in Fig. 1.

Module A represents the SE residual connection module, module B represents the CBAM residual connection module, module C represents the regular residual module, and module D includes convolutional and softmax operations. During the model training process, there are four output layers: Map1, Map2, Map3, and Map4.

The specific structures of the attention mechanism and residual connection modules are shown in Fig. 2. Fig. 2(a) depicts the residual connection module incorporating the SE attention mechanism, while Fig. 2(b) illustrates the residual connection module incorporating the CBAM attention mechanism.

### A. Dilated convolution

Dilated Convolution (also known as Atrous Convolution) is a special type of convolutional operation that, compared to traditional convolutions, can increase the receptive field of a convolutional layer while keeping the feature map size constant. Additionally, it achieves this without increasing the number of parameters, thereby enhancing the network's

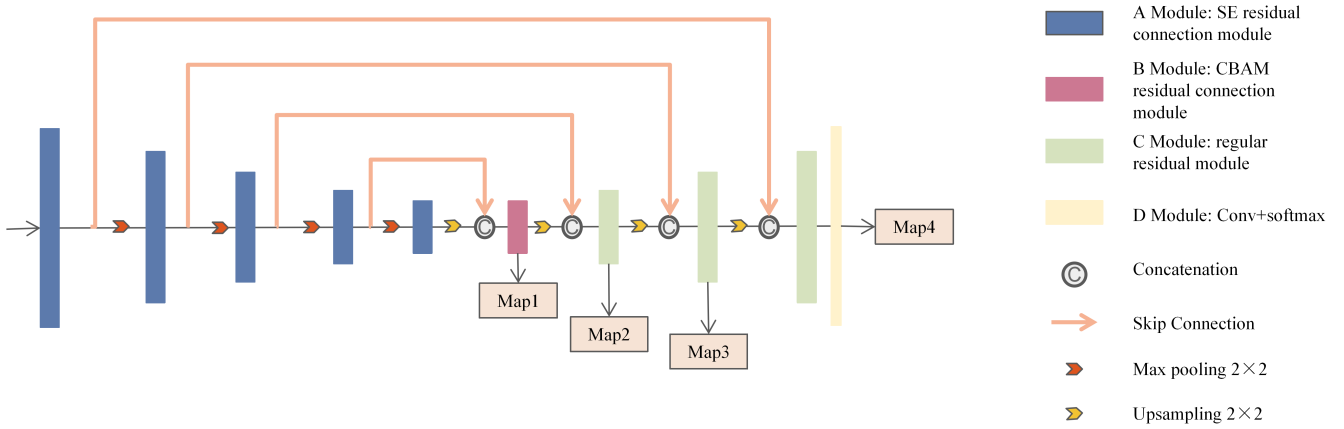


Fig. 1. The framework of enhanced liver tumor segmentation network (ELTS-Net).

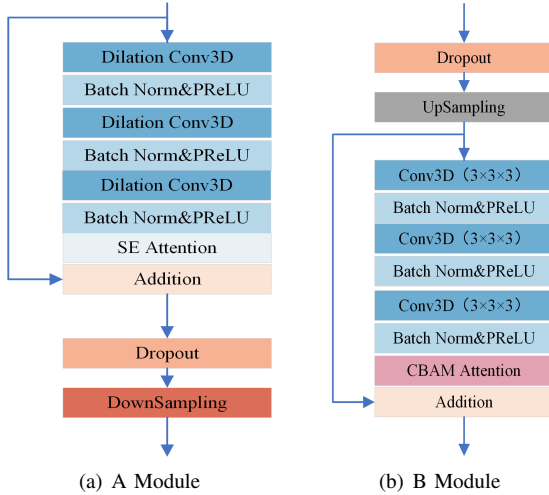


Fig. 2. The developed residual modules with attention mechanism.

representational capacity and precision while reducing the risk of overfitting.

In traditional convolutional operations, the convolution kernel applies convolution to every pixel of the input. However, in dilated convolution, the convolution kernel can sample the input pixels with intervals during the convolution process, and this interval is referred to as the dilation rate. By increasing the dilation rate, the effective receptive field of the convolution kernel expands, enabling the network to better capture contextual information in the input image. At the same time, since this expansion of the receptive field is realized by increasing the internal span of the convolutional kernel, it does not increase the number of parameters. This allows the null convolution to improve the performance of convolutional neural networks without increasing the computational effort and the number of parameters. The calculation method for the receptive field of dilated convolution is given by the following

equation:

$$K + (K - 1)(r - 1) \quad (1)$$

$K$  represents the size of the convolution kernel, and  $r$  denotes the dilation rate. Fig. 3 illustrates the receptive field of dilated convolution at different dilation rates.

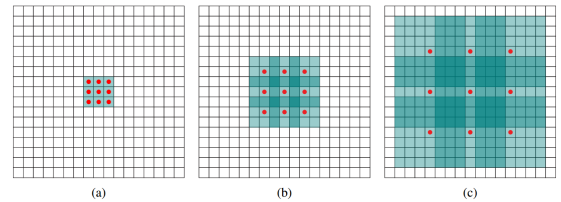


Fig. 3. The sketch of dilated convolution operations with different dilation rates.

In the ELTS-Net, the dilated convolution method is used instead of the ordinary convolution. Under the condition that the image resolution is constant, the receptive field and coverage range of a single convolution kernel can be expanded, which can more effectively deal with the situation of different scale sizes and tumor boundaries, and improve the accuracy and accuracy of image segmentation.

### B. Residual connection

Increasing the depth of a network structure enhances its feature extraction and representation capabilities. However, as neural networks become deeper and more complex, deep networks are prone to the problems of vanishing or exploding gradients, leading to biased model outputs and a decline in training performance, thereby limiting the model's overall performance. To address these issues, He et al. [11] proposed the ResNet in 2016. The ResNet employs residual learning by introducing residual connections across layers, where the

outputs of preceding and succeeding layers are added together through these connections to obtain the final output. The structure of residual connections is illustrated in Fig. 4.

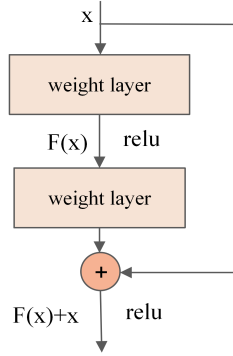


Fig. 4. The structure of residual connection operation.

In the ResNet architecture, the function  $H(x)$  represents the part that a traditional network needs to learn. Within the residual connection,  $H(x)$  can be decomposed into two components: the residual function  $F(x)$  and the identity mapping function  $x$ . That is,  $H(x) = F(x) + x$ . The residual function  $F(x)$  represents the bias or residual of the intermediate layer operation, while the identity mapping function  $x$  represents the original input signal that bypasses the intermediate layer processing [21]. In this context, the part that the network needs to learn is no longer  $H(x)$  but rather  $F(x) = H(x) - x$ .

At training time of the network, it is easier to adjust the network weight parameters to fine-tune the network's constant mapping to the input data than it is to retrain a complete mapping. Residual connectivity allows certain layers in the network to use the constant mapping of the input data directly, thus mitigating the problems of information loss and gradient vanishing in the network. As a result, the network can learn the residual parts more easily, allowing the network to be trained faster and optimized better. When the network reaches the optimal state, if the depth of the network continues to increase, the weight of the residual function will gradually converge to zero, and only the constant mapping part is retained. Therefore, using the residual connection structure ensures that the network is in the optimal state, thus avoiding the network performance degradation with increasing depth.

The ELTS-Net replaces each convolutional module in the 3D U-Net network with a module that includes residual connections. By adopting cross-layer residual connections, more low-level features can be preserved, aiding the network in learning low-level features more effectively while mitigating the issue of gradient vanishing.

### C. Multi-domain visual attention mechanism

The Squeeze-and-Excitation (SE) attention mechanism [12] is an attention mechanism used in convolutional neural networks that learns the importance of each channel. It guides the network to better focus on features that are useful for classification or regression tasks by weighting the input feature maps.

The core of the SE attention mechanism is the ‘‘Squeeze-and-Excitation’’ operation. In the squeeze phase, each channel’s feature map is transformed into a single value through global pooling, capturing the channel’s global characteristics [40]. In the excitation phase, the importance of each channel is reassigned using learned weights, enabling adaptive channel-wise weighting. The algorithmic flow of the SE module is illustrated in Fig. 5.

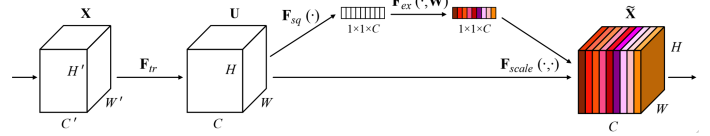


Fig. 5. The flowchart of SE attention.

The SE attention mechanism does not alter the size of the feature maps, allowing it to be embedded into different layers of a convolutional neural network to weight features at different levels. This enables the network to more effectively extract crucial information from the data, thereby enhancing the model’s performance.

In the ELTS-Net, the SE module is embedded before each downsampling step, specifically before each max pooling operation. This ensures that the importance of each channel is weighted both before the size reduction and after the feature extraction.

The Convolutional Block Attention Module (CBAM) attention mechanism [43] consists of two sub-modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). These modules focus on channel and spatial information, respectively. The CBAM structure is illustrated in Fig. 6.

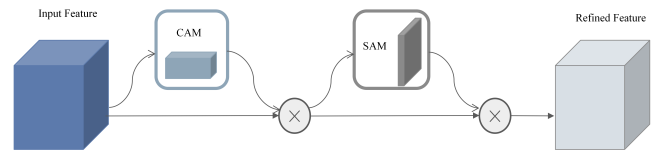


Fig. 6. The diagram of convolutional block attention mechanism (CBAM).

After the filtering process of CAM and SAM, the input features are recalibrated to emphasize important features and compress unimportant ones. The structure diagram of CAM is shown in Fig. 7.

The Channel Attention Module dynamically learns the interdependencies between channels by utilizing global information and generates weight coefficients for each channel. This module performs global average pooling on the feature map along the spatial dimension, followed by a series of fully connected layers and activation functions, to obtain a weight vector for each channel. This weight vector adjusts the importance of channel features, allowing the network to focus

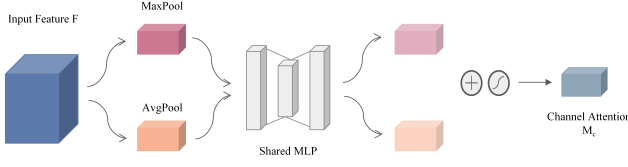


Fig. 7. The diagram of channel attention module (CAM).

more on significant channel features. The calculation formula for channel attention is expressed as follows:

$$\begin{aligned} W_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \end{aligned} \quad (2)$$

The structure of the SAM module is illustrated in Fig. 8.

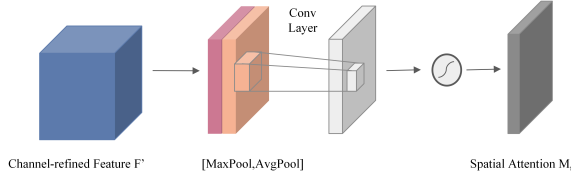


Fig. 8. The diagram of spatial attention module (SAM).

The output feature map from the CAM serves as the input to the SAM. Firstly, global max pooling and global average pooling operations are performed on the feature map to extract the maximum and average values for each pixel, resulting in two  $H \times W \times 1$  feature maps that preserve spatial information. The pooled features are then concatenated to obtain a  $H \times W \times 2$  image. Subsequently, a convolutional operation is applied to compress the channel dimension to 1. Finally, after passing through the sigmoid function, the spatial attention weights are obtained and multiplied with the input feature map to generate the final features. The calculation formula for SAM is as follows:

$$\begin{aligned} W_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7 \times 7}(F_{avg}^c; F_{max}^c)) \end{aligned} \quad (3)$$

The resulting spatial attention weights obtained from SAM incorporate channel attention. Therefore, after performing the aforementioned two operations, the original feature map is multiplied element-wise with the features obtained from SAM.

The CBAM attention module is inserted after the convolution operation in the first decoding stage, which addresses the issue of missing cross-scale information between the encoder and decoder and improves the accuracy of the feature reconstruction process.

We employ CBAM and SE attention mechanisms during upsampling and downsampling, respectively, creating a cohesive synergy. This innovative approach of integrating different

components and attention mechanisms allows our model to outperform using a single component in certain crucial aspects. The SE attention mechanism adaptively learns inter-channel relationships, weighting different channels to extract key features, which helps the network focus better on morphological characteristics of the liver and tumor regions and enhances recognition accuracy. The CBAM attention mechanism combines spatial and channel attention to effectively capture correlations between spatial and channel dimensions.

By simultaneously incorporating SE and CBAM attention mechanisms, we can better utilize the complementary nature of the components, fully exploit their advantages, and improve the encoding of contextual information in abdominal CT images of the liver and its tumors, enabling the network to be able to capture global and local information more flexibly in the downsampling and upsampling phases, and to better understand the location information of these structures, thus improving the network's perceptual ability and segmentation performance.

#### D. Deep supervision mechanism

In order to address the challenges of training deep neural networks, such as the difficulty of convergence and the occurrence of gradient vanishing or exploding due to their complex structure and numerous parameters, Lee et al. [17] proposed the Deep Supervision mechanism to accelerate the convergence speed and improve the segmentation performance of the network.

Traditional multilayer convolutional neural networks usually use only the output of the last layer to backpropagate progressively during training for parameter updating and error computation to reduce the loss of model prediction and labeling. Deep supervision, on the other hand, adds additional objective functions to the hidden layers at different depths to judge the feature maps. That is, each stage in a multilayer neural network model produces a separate output, called an intermediate output, so that lower-level information can be introduced in time to help train that network during training. As the network is trained, the intermediate outputs provide more feature representations that are associated with the input data at lower levels, and in this way the model can better learn complex abstract representations of the data. In this study, additional supervision signals were incorporated into the hidden layers of the convolutional neural network to assess the quality of the feature maps. This enables the timely introduction of lower-level information during training. The deep supervision is illustrated in Fig. 9.

In the decoder stage, in addition to the final output, an additional output layer is added after each decoder stage, resulting in four output layers: Map1, Map2, Map3, and Map4. During the training process, the model returns all four intermediate output results as inputs for deep supervision and performs corresponding evaluation and error calculation to obtain the final loss. Since the loss values of outputs at different scales may vary in magnitude and importance, it is necessary to combine them with weights. Since the loss function computation results corresponding to the outputs of

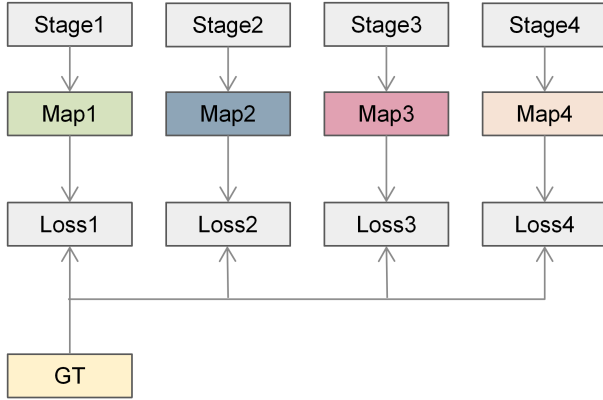


Fig. 9. The scheme of deep supervision.

different scales may be of different sizes and have different importance, it is necessary to combine them in a weighted manner and adjust the weights dynamically during the training process to avoid any one scale having too large or too small an impact on the training results in order to compute the gradient more correctly and update the model parameters to achieve the best segmentation results. The calculation of the final loss is as follows:

$$L(y_i, \hat{y}_i) = \alpha (f_1(y_1, \hat{y}_1) + f_2(y_2, \hat{y}_2) + f_3(y_3, \hat{y}_3) + f_4(y_4, \hat{y}_4)) \quad (4)$$

$y_i$  and  $\hat{y}_i$  represent the probability values, ground truth annotations, and the difference between the predicted results and the true segmentation standards at different scales, respectively.  $\alpha$  represents the deep supervision coefficient, which is used to balance the weight ratio between the loss function and the final loss function.

The first three output layers typically correspond to different depths or resolutions within the network, capturing features at various levels of abstraction. During training, applying  $\alpha$  balance to these layers helps prevent the model from becoming overly dependent on specific depths, ensuring a more stable and effective learning process. The fourth output layer often corresponds to the final overall prediction of the network.

In segmentation tasks, the depth of the hidden layers can have an impact on the segmentation results. Adding supervisory information only to the raw resolution output may overlook to the impact that the hidden layer feature maps may have on the results. For the task of liver and tumor segmentation, the liver and tumor sizes in each CT map are uncertain. Adding additional depth feedback on the shallower hidden layers can constrain the network weights to the corresponding target regions, thus improving the effect of the deeper feature maps and optimizing the final segmentation results. In addition, supervised learning only through the output layer at the original resolution of the image causes the network to focus too much on details and texture information, whereas

supervising the deeper hidden layer allows the network to directly learn the semantic features of the liver and the tumor, improves the robustness of the segmentation task, and enables the network to perform well in dealing with a variety of tumor and liver segmentation tasks. By deeply supervising the multi-layer neural network feature maps during the network decoding process, it can help the network to be trained and improve the segmentation accuracy.

#### IV. EXPERIMENTS AND RESULTS

In this paper, we validate the effectiveness of the proposed improved network using the publicly available LiTS2017 dataset. Firstly, we introduce the dataset and experimental setup used in our study. Then, we present a series of evaluation metrics employed to assess the network performance. Finally, we validate the effectiveness of the network through testing results and comparisons with different algorithms.

##### A. Experimental dataset

Despite its excellence in soft tissue contrast and detailed information, MRI suffers from limitations such as difficulty in acquisition and high cost. CT images, on the other hand, possess higher spatial resolution, which is crucial for liver structure and tumor segmentation tasks, and are more efficient in terms of computational cost and resource utilization. Therefore, in this study, we chose to use abdominal CT images, and the dataset used was the LiTS2017 dataset [10]. It consists of CT images from 131 subjects contributed by six different hospitals with various types of liver tumor diseases. The size of each CT image is  $512 \times 512 \times A$ , where  $A$  represents the number of slices in the three-dimensional data. The axial slice count varies between 42 and 1026. The slice thickness is 3mm, and the pixel spacing ranges from 0.55mm to 1mm. The slice interval is between 0.55mm and 6.0mm [18]. The number of tumors included in each sample ranged from 0 to 75, with sizes ranging from  $38 \text{ mm}^3$  to  $349 \text{ mm}^3$ . After one case was randomly discarded, the image data was randomly divided into training set, verification set and test set according to the ratio of 7:1:2. There were 91 cases in the final training set, 13 cases in the validation set and 26 cases in the test set. A series of image preprocessing techniques are then adopted to process all the data: firstly, due to the limitation of GPU video memory resources, the data and its annotations need to be segmented according to the z-axis to constitute segmented data. Then window enhancement technique is used to enhance the image contrast by adjusting the range of gray values of the image to highlight the tissue structures or lesion regions of interest. Next, for the noise points present in the image, local histogram equalization is performed for each pixel point in the medical image data based on the gray level distribution of its neighboring pixels, which further improves the visual quality and diagnostic accuracy of the medical image on the basis of windowing. After that, data enhancement methods such as rotation, reduction, and scaling are used to enhance the dataset to improve the accuracy and generalizability of the model [23]. It is foreseeable that augmenting the dataset not only reduces the probability of overfitting, but also improves the robustness

of the system [31]. When performing data enhancement, the same transformation needs to be performed on the labels at the same time to ensure the consistency in position and size of the labels and images [44].

### B. Experimental environment

The experimental process was based on the PyTorch learning framework, utilizing the Adam optimizer. Due to GPU memory constraints, a batch size of 2 was employed in this experiment. The learning rate was set to  $10^{-4}$ , and the number of epochs was set to 200. The Tversky loss function was utilized to address the class imbalance issue, and a dropout strategy was applied to prevent model overfitting.

### C. Evaluation metrics

The segmentation task of the liver and its tumors aims to achieve detection and segmentation of the target region. To comprehensively evaluate the segmentation performance of the model, multiple metrics were employed in this study. Among them, the Dice Similarity Coefficient (DSC) was the primary hybrid metric used in this study, while the Dice coefficient and Intersection over Union (IoU) were used to assess the model's performance in liver and tumor segmentation. The recall rate (Recall) was used to evaluate the model's performance in liver and tumor detection, while the Volume Overlap Error (VOE) was used to measure the degree of segmentation errors in the segmentation process, assessing its segmentation performance. The formulas for these metrics are as follows:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$VOE = 1 - \frac{2 \times TP}{TP + FP + FN} \quad (8)$$

In this context, TP represents True Positives, which are the number of samples that are predicted as positive and are actually positive. FP represents False Positives, which are the number of samples that are predicted as positive but are actually negative. FN represents False Negatives, which are the number of samples that are predicted as negative but are actually positive.

The Dice coefficient is used to evaluate the similarity between the predicted values and the labels. A value closer to 1 indicates a higher similarity between the predicted values and the labels, indicating better segmentation performance. IoU is used to measure the degree of overlap between the model's predictions and the ground truth annotations. A larger value implies that the predicted results are more consistent with the ground truth, indicating higher accuracy. Recall is a metric used to measure the model's ability to correctly identify positive samples, i.e., how many true positive samples the model can correctly detect. A higher value indicates a stronger ability of the model to detect positive samples. VOE

describes the volume overlap between the segmentation result and the ground truth. A smaller VOE value indicates a larger overlap between the segmentation result and the ground truth, indicating better segmentation performance.

### D. Quantitative results and visualizations

In order to demonstrate the segmentation performance of the ELTS-Net on liver and tumor segmentation tasks, this study compared the Dice index of 26 sample CT images in the test dataset. The results are shown in Fig. 10 and Fig. 11.

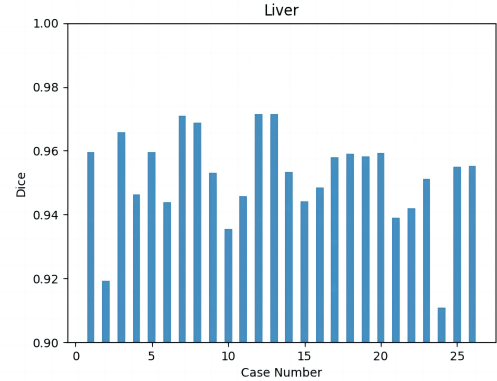


Fig. 10. Dice coefficient for liver segmentation by ELTS-Net.

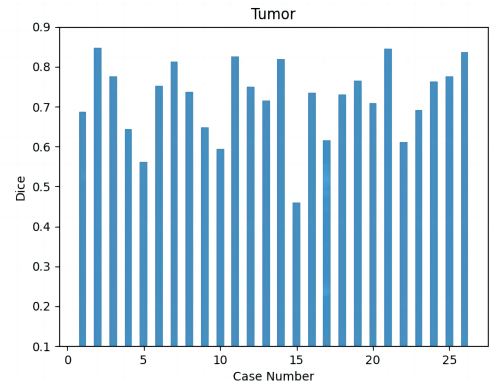


Fig. 11. Dice coefficient for tumor segmentation by ELTS-Net.

Based on the Dice coefficients shown in the figures, the liver segmentation results indicate values predominantly distributed between 92% and 98%. This demonstrates that the proposed network model performs well in accurately localizing the liver region. Furthermore, the Dice coefficients for liver tumor segmentation exhibit values mostly ranging from 60% to 90%. This indicates that the ELTS-Net effectively segments liver tumor regions, with the segmentation results aligning well with the ground truth.

To verify the structural advantages of ELTS-Net, we used 3D U-Net network as a baseline model for comparison. The results of the evaluation of the two networks on the effect of liver and tumor segmentation are shown in Table I. From the

table, it can be seen that ELTS-Net improves all the indicators to different degrees and performs better.

TABLE I  
LIVER AND TUMOR SEGMENTATION RESULTS OF 3D U-NET AND ELTS-NET

Models	Metrics			
	<i>Dice</i>	<i>IoU</i>	<i>Recall</i>	<i>VOE</i>
3D U-Net(Liver) [7]	94.3%	89.1%	95.7%	10.6%
<b>ELTS-Net(Liver)(ours)</b>	<b>95.2%</b>	<b>90.8%</b>	<b>97.3%</b>	<b>9.2%</b>
3D U-Net(Tumor) [7]	68.8%	52.4%	88.2%	47.6%
<b>ELTS-Net(Tumor)(ours)</b>	<b>71.9%</b>	<b>56.1%</b>	<b>89.4%</b>	<b>42.9%</b>

Specifically, when using the ELTS-Net structure for liver segmentation, the Dice coefficient, IoU index, and Recall evaluation metrics improved by 0.9%, 1.6%, and 1.6%, respectively, while the VOE metric decreased by 1.4%. In the case of liver tumor segmentation, the improvement in these metrics was more significant, with a 3.1% increase in the Dice coefficient, a 3.7% increase in the IoU index, a 1.2% increase in Recall, and a 4.7% decrease in the VOE metric.

These changes in the metrics indicate an increase in the consistency and overlap between the predicted segmentation regions and the ground truth labels. The ELTS-Net more accurately captures the true liver and tumor pixels, reducing the occurrence of falsely labeling liver and tumor regions as background or missing them.

In addition, the comparison reveals that the improved network structure improves both liver and tumor segmentation, but the improvement in tumor segmentation is more obvious. This may be due to the fact that the addition of structures such as residual connection, CBAM with SE attention mechanism, and null convolution can help the network better capture and learn the multi-scale features of tumors, complex textures, locations with high uncertainty, boundary information, and complex growth patterns, thus improving the accuracy of segmenting tumors. In contrast, the improvement of segmentation accuracy of the liver by these structures is more limited due to the relatively simple and fixed features of the liver's shape, texture, and location.

To better observe the segmentation performance of the ELTS-Net, three randomly selected examples from the set of 26 test images are displayed in Fig. 12, illustrating the segmentation results of the network.

Fig. 12(a) shows the original abdominal CT image of the patient, Fig. 12(b) depicts the ground truth mask for liver and tumor, and Fig. 12(c) presents the liver and tumor segmentation results obtained using the proposed ELTS-Net. From the segmentation results in the figures, it can be observed that the algorithm performs well in terms of liver and liver tumor localization and identification, achieving more accurate liver and tumor segmentation.

However, this algorithm still has limitations for edge segmentation of target tissues, and there are still some problems that need to be solved in our research: acquiring a large number of medical images is expensive and difficult to annotate [34]; the model parameters are large, and most medical The

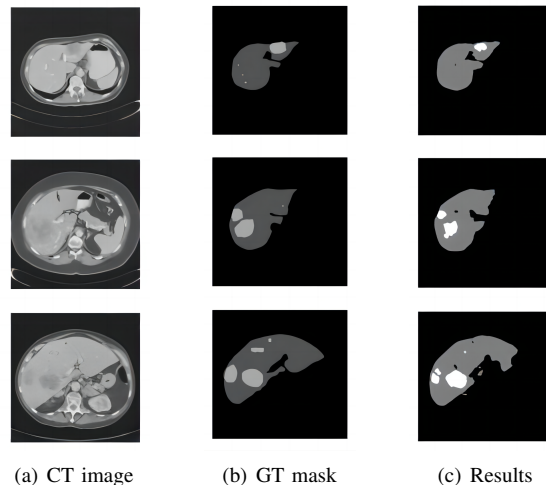


Fig. 12. Visualization results on randomly selected test set.

computing equipment of the institution may not be able to meet the high requirements of the network; the segmentation results show a certain degree of non-smoothness, and it is difficult for the algorithm to effectively identify and segment smaller tumor locations, etc.

### E. Comparative analysis

In this study, the ELTS-Net structure was compared with other mainstream algorithms used for liver tumor segmentation, including the baseline model U-Net [36], Attention U-Net (AttUNet) [30] in 2D networks, ResUNet-a [8], and the 3D network structure V-Net [29], three-dimensional dual path multiscale convolutional neural network (TDP-CNN) [28]. The comparison results are shown in Table 2.

TABLE II  
COMPARISON OF SEGMENTATION PERFORMANCE AMONG DIFFERENT ALGORITHMS

Models	Metrics	
	<i>Dice(Liver)</i>	<i>Dice(Tumor)</i>
U-Net [36]	92.2%	61.1%
AttUNet [30]	91.2%	60.4%
ResUNet-a [8]	93.8%	62.0%
V-Net [29]	94.7%	67.2%
TDP-CNN [28]	96.5%	68.9%
<b>ELTS-Net (ours)</b>	<b>95.2%</b>	<b>71.9%</b>

By observing the Dice coefficients of different networks, we found that the ELTS-Net outperformed all the comparative algorithms in tumor segmentation tasks. In liver segmentation tasks, it outperformed all the comparative algorithms except for the TDP-CNN structure. The overall segmentation accuracy of the 3D network was superior to that of the 2D network, which could be attributed to the 3D network's better utilization of spatial information from the input images.

In summary, the ELTS-Net achieves relatively accurate liver segmentation, and the learning ability of multi-dimensional features and irregular features of complex structures is significantly enhanced.

## V. CONCLUSION

This paper proposes ELTS-Net, an enhanced 3D network architecture for liver tumor segmentation. ELTS-Net incorporates several key design elements to enhance segmentation accuracy and network convergence. Dilated convolution is utilized to expand the receptive field of convolutional layers, enabling improved feature learning with fewer layers. Skip connections are introduced to prevent overfitting and facilitate convergence by connecting layers. The SE and CBAM dual-attention mechanisms are used to enhance the interaction of semantic information across spatial and channel dimensions. This facilitates adaptive fusion of local and global information for enhanced feature representation. Furthermore, the network employs a deep supervision mechanism to enhance segmentation performance and accelerate convergence. Experimental results on the LiTS2017 dataset show that the proposed ELTS-Net achieves a liver segmentation accuracy of 95.2% and a tumor segmentation accuracy of 71.9%.

Although the network performs well in the liver and its tumor segmentation task, there are still problems such as large number of model parameters and difficulty in obtaining medical data. Therefore, our goals in the future are to 1) Consider adjusting network parameters to make the network performance optimal. 2) Explore lightweight network architecture, balance model complexity and recognition accuracy, and effectively improve network performance [46], [48]. 3) Explore deep learning methods that use unlabeled data for feature learning and image segmentation, extract more effective features in the data, and apply them to image segmentation and recognition to improve segmentation accuracy.

## REFERENCES

- [1] G. Bao, L. Ma and X. Yi, "Recent advances on cooperative control of heterogeneous multi-agent systems subject to constraints: A survey", *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 539-551, 2022.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries", *CA: A Cancer Journal for Clinicians*, vol. 68, pp. 394-424, 2018.
- [3] Q. Cao, H. Yu, P. Charisse, S. Qiao and B. Stevens, "Is high-fidelity important for human-like virtual avatars in human computer interactions?" *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 1, pp. 15-23, 2023.
- [4] J. Chen, S. Bai, G. Wan and Y. Li, "Research on YOLOv7-based defect detection method for automotive running lights", *Systems Science & Control Engineering*, vol. 11, no. 1, article no. 2185916, 2023.
- [5] X. Chen, R. Zhang, and P. Yan, "Feature fusion encoder decoder network for automatic liver lesion segmentation", *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 430-433, 2019.
- [6] Y. Chen and T. Deng, "Leader-Follower UAV formation flight control based on feature modelling", *Systems Science & Control Engineering*, vol. 11, no. 1, article no. 2268153, 2023.
- [7] Özgün. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation", *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016: 19th International Conference (MICCAI)*, pp. 424-432, 2016.
- [8] F.I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94-114, 2020.
- [9] Y. Ding, M. Fu, P. Luo and F. X. Wu, "Network learning for biomarker discovery", *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 1, pp. 51-65, 2023.
- [10] S. Gul, M.S. Khan, A. Bibi, A. Khandakar, M.A. Ayari, and M.E. Chowdhury, "Deep learning techniques for liver and liver tumor segmentation: A review", *Computers in Biology and Medicine*, vol. 147, article no. 105620, 2022.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132-7141, 2018.
- [13] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CcNet: Criss-cross attention for semantic segmentation", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 603-612, 2019.
- [14] J.G. Jeong, Y.J. Kim, K.G. Kim, and W.S. Lee, "Deep 3D attention U-Net based whole liver segmentation for anatomical volume analysis in abdominal CT images", *International Forum on Medical Imaging in Asia 2021*, pp. 13-18, 2021.
- [15] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans", *Frontiers in Bioengineering and Biotechnology*, vol. 8, article no. 605132, 2020.
- [16] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks", *Communications of the ACM*, vol. 60, pp. 84-90, 2017.
- [17] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets", *Artificial Intelligence and Statistics*, pp. 562-570, 2015.
- [18] T. Lei, D. Zhang, X. Du, X. Wang, Y. Wan, and A.K. Nandi, "Semi supervised medical image segmentation using adversarial consistency learning and dynamic convolution network", *IEEE Transactions on Medical Imaging*, vol. 42, pp. 1265-1277, 2023.
- [19] T. Lei, W. Zhou, Y. Zhang, R. Wang, H. Meng, and A.K. Nandi, "Lightweight v-net for liver segmentation", *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1379-1383, 2020.
- [20] H. Li, P. Wu, N. Zeng, Y. Liu and F.E. Alsaadi, "A survey on parameter identification, state estimation and data analytics for lateral flow immunoassay: from systems science perspective", *International Journal of Systems Science*, vol. 53, pp. 3556-3576, 2022.
- [21] H. Li, N. Zeng, P. Wu, and K. Clawson, "Cov-Net: A computer-aided diagnosis method for recognizing covid-19 from chest x-ray images via machine vision", *Expert Systems with Applications*, vol. 207, article no. 118029, 2022.
- [22] J. Li, Z.L. Yu, Z. Gu, H. Liu, and Y. Li, "Dilated-inception net: multi scale feature aggregation for cardiac right ventricle segmentation", *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 3499-3508, 2019.
- [23] W. Li, Y. Niu and Z. Cao, "Event-triggered sliding mode control for multi-agent systems subject to channel fading", *International Journal of Systems Science*, vol. 53, no. 6, pp. 1233-1244, 2022.
- [24] X. Li, Q. Song, Z. Zhao, Y. Liu and F. E. Alsaadi, "Optimal control and zero-sum differential game for Hurwicz model considering singular systems with multifactor and uncertainty", *International Journal of Systems Science*, vol. 53, no. 7, pp. 1416-1435, 2022.
- [25] B.S. Lin, K. Michael, S. Kalra, and H.R. Tizhoosh, "Skin lesion segmentation: U-nets versus clustering", *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-7, 2017.
- [26] M. Liu, Z. Wang, H. Li, P. Wu, F.E. Alsaadi, and N. Zeng, "AA-WGAN: Attention augmented wasserstein generative adversarial network with application to fundus retinal vessel segmentation", *Computers in Biology and Medicine*, vol. 158, article no. 106874, 2023.
- [27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015.
- [28] L. Meng, Y. Tian, and S. Bu, "Liver tumor segmentation based on 3D convolutional neural network with dual scale", *Journal of Applied Clinical Medical Physics*, vol. 21, pp. 144-157, 2020.
- [29] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation", *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565-571, 2016.
- [30] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert "Attention U-Net: Learning where

- to look for the pancreas”, *arXiv:1804.03999*, [online] Available: <https://arxiv.org/abs/1804.03999>, 2018.
- [31] F. Pang and X. Chen, “MS-YOLOv5: a lightweight algorithm for strawberry ripeness detection based on deep learning”, *Systems Science & Control Engineering*, vol. 11, no. 1, article no. 2285292, 2023.
- [32] J. Park, S. Woo, J.-Y. Lee, and I.S. Kweon, “BAM: Bottleneck attention module”, *arXiv:1807.06514*, [online] Available: <https://doi.org/10.48550/arXiv.1807.06514>, 2018.
- [33] F. Qu, X. Zhao, X. Wang and E. Tian, “Probabilistic-constrained distributed fusion filtering for a class of time-varying systems over sensor networks: a torus-event-triggering mechanism”, *International Journal of Systems Science*, vol. 53, no. 6, pp. 1288–1297, 2022.
- [34] B. Rahi, M. Li and M. Qi, “A review of techniques on gait-based person re-identification”, *International Journal of Network Dynamics and Intelligence*, vol. 2, pp. 66-92, 2023.
- [35] S.M. Reza, D. Bradley, N. Aiosa, M. Castro, J.H. Lee, B.-Y. Lee, R.S. Bennett, L.E. Hensley, Y. Cong, R. Johnson et al., “Deep learning for automated liver segmentation to aid in the study of infectious diseases in nonhuman primates”, *Academic Radiology*, vol. 28, pp. S37–S44, 2021.
- [36] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation”, *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, pp. 234–241, 2015.
- [37] A.G. Roy, N. Navab, and C. Wachinger, “Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks”, *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference*, pp. 421–429, 2018.
- [38] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, “Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on fcns”, *Artificial Intelligence in Medicine*, vol. 83, pp. 58–66, 2017.
- [39] J. Wang, Y. Zhuang, and Y. Liu, “FSS-Net: A Fast Search Structure for 3D Point Clouds in Deep Learning”, *International Journal of Network Dynamics and Intelligence*, vol. 2, article no. 100005, 2023.
- [40] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient channel attention for deep convolutional neural networks”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11534–11542, 2020.
- [41] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, 2018.
- [42] Z. Wang, Y. Zou, and P.X. Liu, “Hybrid dilation and attention residual u-net for medical image segmentation”, *Computers in Biology and Medicine*, vol. 134, article no. 104449, 2021.
- [43] S. Woo, J. Park, J.-Y. Lee, and I.S. Kweon, “CBAM: Convolutional block attention module”, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.
- [44] P. Wu, Z. Wang, H. Li, and N. Zeng, “KD-PAR: A knowledge distillation based pedestrian attribute recognition model with multi-label mixed feature learning network”, *Expert Systems with Applications*, vol. 237, article no. 121305, 2024.
- [45] P. Wu, Z. Wang, B. Zheng, H. Li, F.E. Alsaadi, and N. Zeng, “AGGN: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion”, *Computers in Biology and Medicine*, vol. 152, article no. 106457, 2023.
- [46] T. Xie, Z. Wang, H. Li, P. Wu, H. Huang, H. Zhang, F.E. Alsaadi, and N. Zeng, “Progressive attention integration-based multi-scale efficient network for medical imaging analysis with application to COVID-19 diagnosis”, *Computers in Biology and Medicine*, vol. 159, article no. 106947, 2023.
- [47] P. Xu, C. Chen, X. Wang, W. Li, and J. Sun, “RoI-based intraoperative mr-ct registration for image-guided multimode tumor ablation therapy in hepatic malignant tumors”, *IEEE Access*, vol. 8, pp. 13613–13619, 2020.
- [48] N. Zeng, X. Li, P. Wu, H. Li and X. Luo, “A novel tensor decomposition-based efficient detector for low-altitude aerial objects with knowledge distillation scheme”, *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 2, 2024.
- [49] C. Zhang, J. Lu, Q. Hua, C. Li, and P. Wang, “SAA-Net: U-shaped network with scale-axis-attention for liver tumor segmentation”, *Biomedical Signal Processing and Control*, vol. 73, article no. 103460, 2022.
- [50] K. Zhou, W. Qi, Z. Gui, and Q. Zeng, “PSU-Net: Paired spatial U-Net for hand segmentation with complex backgrounds”, *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 549–563, 2022.
- [51] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A nested u-net architecture for medical image segmentation”, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018*, pp. 3–11, 2018.