IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 January 2024, accepted 25 January 2024, date of publication 13 February 2024, date of current version 29 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365577

RESEARCH ARTICLE

Lung Parenchyma Segmentation Based on U-Net Fused With Shape Stream

LUN ZHU[1], YINGHUI CAI[1], JIAHAO LIAO[1], AND FAN WU[2]
[1]School of Computers and Artificial Intelligence, Changzhou University, Changzhou 213000, China
[2]School of Information Technology, Zhejiang Shuren University, Hangzhou 310000, China

Corresponding author: Fan Wu (wufan@zjsru.edu.cn)

This work was supported in part by the Leading Talents of Science and Technology Innovation in Zhejiang Province under Grant 2020R52042, in part by the Zhejiang-Netherlands Joint Laboratory for Digital Diagnosis and Treatment of Oral Diseases, in part by the Key Research and Development Program of Zhejiang under Grant 2021C01189, in part by the National Natural Science Foundation of China under Grant 82011530399, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LGG20F020015, in part by the National Key Research and Development Plan of China under Grant 2021YFE0191200, in part by the Hangzhou Major Artificial Intelligence Science and Technology Innovation Project under Grant 2022AIZD0128, and in part by the National College Student Innovation and Entrepreneurship Training Plan of China under Grant 202211842025.

ABSTRACT Accurate lung parenchyma segmentation is vital for computer-aided lung cancer diagnosis. Existing lung parenchyma segmentation networks excel at segmenting large and clear lung parenchyma regions, but struggle with small and blurry regions. This study proposes an improved network structure to improve the segmentation performance of small and blurry lung parenchyma regions while maintaining accuracy for large and clear lung parenchyma regions. The proposed network is an improved network based on U-Net. A shape stream branch and multi-scale convolutional blocks are introduced into the network. The proposed network takes computed tomography (CT) images as inputs and generates corresponding binary masks as outputs. In this study, the original CT images are derived from the Open Source Imaging Consortium (OSIC) Pulmonary Fibrosis Progression dataset. In the experimental results, the overall mean Dice Similarity Coefficient (DSC) of the proposed network reaches 0.932907, which is 0.64% higher than that of the second-best network. The segment test results show that the proposed network has significantly better segmentation performance than other networks when the lung parenchyma is relatively small. On a patient-by-patient basis, the mean DSC of the proposed network is 0.63% higher than that of the second-best network. The proposed network improves the segmentation performance of small and blurry lung parenchyma while maintaining accuracy for large and clear lung parenchyma.

INDEX TERMS Deep learning, lung parenchyma segmentation, shape stream, multi-scale feature.

I. INTRODUCTION

Lung malignancy is one of the leading causes of cancer morbidity and mortality worldwide. Chest computed tomography (CT) scan is a common way to detect lung abnormalities, which can help in the early detection and treatment of lung cancer. However, chest CT scans usually consist of hundreds of slices. Relying solely on the subjective judgment of radiologists may be prone to bias and visual fatigue, leading to diagnostic error. To improve diagnostic accuracy, many scholars have studied computer-aided diagnosis (CAD) systems, which assist physicians in making accurate judgments. CAD systems provide a secondary diagnosis result beyond human interpretation and assist radiologists in relevant medical imaging diagnoses. In general, accurate lung parenchyma segmentation is a primary requirement for a CAD system used to diagnose lung cancer. Hence, designing a precise method for lung parenchyma segmentation is of paramount importance.

However, designing an effective lung parenchyma segmentation method is a challenging problem. Firstly, the heart and bronchi often present challenges in differentiating them from lung parenchyma in CT images, leading to over-segmentation in the segmentation results. Secondly, the lung parenchyma regions exhibit significant morphological variations across chest CT scans, as shown in Fig. 1. Additionally, the intricate internal details of lung lesions further complicate the segmentation task. Various methods for lung parenchyma

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

segmentation exist, including traditional thresholding, region growing, and prevalent deep learning approaches. However, almost none of them specifically address the effective handling of small, elongated, and blurry lung parenchyma in CT images. The segmentation of these images cannot be overlooked. On the one hand, CT images with a small proportion of lung parenchyma may still contain certain pathological information. On the other hand, 3D reconstruction of the lungs has become an important tool in the treatment of lung diseases. Accurately segmenting both large and small lung parenchyma regions is beneficial for more precise 3D reconstruction of the patient's lungs. Therefore, it is necessary to improve the segmentation accuracy of small and blurry lung parenchyma regions.

This paper proposes enhancements to the U-Net network to significantly improve the segmentation performance of small and blurry lung parenchyma. The main contributions of this paper are as follows:

(1) A two-stream lung parenchyma segmentation network: The paper proposes a two-stream lung parenchyma segmentation network. A shape stream branch is introduced into the network, which is supervised by edge detection loss. The edge detection results from the shape stream are fused with deep-layer features from the backbone network. This method enables the clear edge information generated by the shape stream to be provided to the backbone network, assisting in the accurate localization of the target segmentation regions. Thus improving the segmentation performance of the network.

(2) A multi-scale convolutional feature extraction method: To enable the network to simultaneously capture global features and detailed features, multi-scale convolutional blocks are used in the feature extraction path of the backbone network. These blocks utilize convolutional kernels of different sizes to extract features from the image, resulting in features of different scales.

II. RELATED WORK

A. THRESHOLD-BASED SEGMENTATION METHODS

Traditional lung parenchyma segmentation methods primarily rely on signal thresholding techniques. These methods determine the region to which each pixel belongs by selecting an appropriate threshold. They are suitable for scenarios with a significant contrast between the target and background regions. Ng et al. [1] improved segmentation accuracy by combining watershed segmentation with texture-based region merging. Kotropoulos and Pitas [2] achieved precise medical image segmentation by fusing support vector machines with threshold segmentation.

Due to the lung region having lower density compared to other regions in the body (such as the aorta and cavities), it appears darker [3]. This enables threshold-based algorithms to effectively handle atypical and pathological lungs. Wang et al. [4] improved the effectiveness of lung tissue segmentation by using an enhanced Freeman chain code and Bezier curves. However, when applied to CT images with numerous densely packed lesions, a noticeable decrease in edge smoothness is observed in the segmentation results. Yang et al. [5] employed a threshold-based pre-segmentation method to automatically select seed points and used morphological post-processing to refine lung parenchyma segmentation. Nevertheless, this method overlooks lesion areas and may result in inadequate segmentation when applied to complex CT lesion images. Walczak et al. [6] enhanced an automated lung segmentation method [7] by incorporating false-positive elimination processing and refining the procedure for separating the two lungs. However, this algorithm does not explicitly segment structures such as pulmonary nodules, small airways, and bronchi.

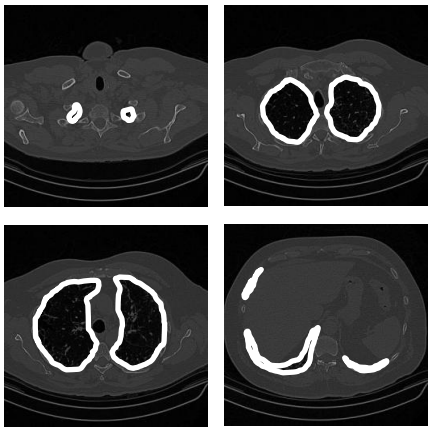


FIGURE 1. Different forms of lung parenchyma in CT scans. The white coils are lung parenchyma regions.

Although threshold-based segmentation methods are known for their simplicity and effectiveness, they frequently encounter challenges in striking a balance between sensitivity and specificity. Moreover, they are susceptible to noise interference, heavily reliant on manual parameter adjustments, and cannot handle morphological variations adeptly.

B. REGION GROWING METHODS

During the process of region growing, the ROI (Region of Interest) is obtained based on predefined conditions. A seed point is manually defined initially, and pixels with similar or identical properties around this point are grouped into the same set. This process continues until no more similar pixels can be found, concluding the operation. Region growing has achieved favorable results in the segmentation of lesions such as brain abnormalities [8], brain tumors [9], and masses [10].

As a simple and easily implementable image segmentation method, region growing has advantages in its applicability to homogeneous regions, independence from pre-training, and ease of handling irregular shapes. But it also comes with drawbacks, such as difficulties in handling boundaries, challenges with uneven grayscale, and connectivity issues. One drawback of region growing in lung parenchyma segmentation is its inability to accurately segment nodules that extend along the edges of the lungs [11].

C. EDGE-DETECTION METHODS

Edge-detection methods address image segmentation issues by detecting edges between different regions. There are primarily two types of methods within this approach: those based on grayscale histograms and those based on gradients. This method is well-established in medical image segmentation. Campadelli et al. [12] achieved an edge detection accuracy of 94.37% from chest X-ray images. Mendonca et al. [13] utilized spatial detectors to identify image edges for the segmentation of lung tissue in X-ray images. They employed 47 X-ray images and obtained a sensitivity of 0.9225 and a positive predictive value of 0.968.

The drawback of edge-detection methods lies in their sensitivity to noise, leading to suboptimal segmentation results when applied to medical images heavily influenced by noise.

D. DEEP LEARNING METHODS

Compared to traditional segmentation methods, deep learning methods possess more complex computational logic and can adapt to more challenging segmentation tasks. Deep learning methods leverage large amounts of annotated data and deep neural networks to iteratively train and refine the segmentation task, leading to highly effective segmentation outcomes. In 2023, Ma and Wang [14] proposed the Segment Anything in Medical Images Model (MedSAM), an enhanced version of the Segment Anything Model (SAM) [15]. In terms of 2D medical image segmentation, MedSAM achieved a Dice Similarity Coefficient (DSC) improvement of 0.2251 over SAM. However, there is still an obvious gap in segmentation performance between MedSAM and models trained specifically for the target task. Therefore, there is still a need for study on targeted segmentation models for specific medical images.

U-Net has achieved significant success in image segmentation tasks with its unique architecture, becoming one of the most commonly used segmentation networks in deep learning [16]. Currently, U-Net has been widely applied in the field of biomedical image segmentation [17], [18], [19], [20]. Numerous U-Net-based modified networks have been developed, such as Residual U-Net (ResUNet) [21], Densely Connected U-Net (DenseUNet) [22], Neural Architecture Search U-Net (NAS-UNet) [23], Full-scale Connected U-Net (Unet-3+) [24], no-new-net (nnU-Net) [25], Fully Dense U-Net (FD-UNet) [26], Transformers U-Net (TransUNet) [27], and Swin Transformer-based U-Net (Swin-UNet) [28].

Currently, many scholars have proposed deep learning methods based on U-Net to segment lung parenchyma. Skourt et al. [3] utilized the U-Net for lung parenchyma segmentation tasks. This network consists of an encoding path for feature extraction and a symmetric decoding path, achieving good performance on lung parenchyma segmentation tasks with its simple network structure. However, this network only performs well in cases where the target is clear. Dutta et al. [29] made advancements to the U-Net network by incorporating dense recursive residual convolutional blocks, facilitating maximum information sharing between layers and achieving a 0.005 increase in the DSC for lung parenchyma segmentation compared to the baseline network. Hu et al. [30] employed scale-aware modules for feature extraction and incorporated attention mechanisms to better learn multi-scale features, resulting in more accurate lung parenchyma segmentation. Jalali et al. [31] utilized bidirectional convolutional long short-term memory networks as advanced integrator modules in their network, enabling more intricate feature map integration compared to simple concatenation. This approach enhanced the learning capabilities for lung parenchyma segmentation tasks.

The aforementioned studies did not focus on small and blurry lung parenchyma regions, and their test results only demonstrated the applicability of the proposed networks for segmenting large and clear lung parenchyma regions. Improving the segmentation accuracy of small and blurry lung parenchyma regions is advantageous for subsequent processing, such as lung 3D reconstruction. Accurate segmentation of lung parenchyma in chest CT sequences is a prerequisite for efficient lung 3D reconstruction tasks. Therefore, it is important to consider the segmentation accuracy of lung parenchyma regions with different sizes.

Inspired by the success of multi-stream network architectures in medical image processing tasks [32], [33], [34] [32], [33], [34] and the widespread application of multi-scale feature extraction methods in the field of computer vision [35], [36], [37] [35], [36], [37], this paper proposes a two-stream convolutional neural network (CNN) for the lung parenchyma segmentation task. The proposed network aims to enhance the segmentation performance of small and blurry lung parenchyma regions while maintaining the segmentation performance of large and clear lung parenchyma regions. This network consists of a backbone network and a shape stream branch. The backbone network is a modified U-Net that utilizes multi-scale feature extraction modules.

III. METHODS

A. NETWORK ARCHITECTURE

The proposed network takes a modified U-Net as the backbone network (referred to as the regular stream), and it integrates a shape stream and multi-scale convolutional blocks (MSC). The network structure is illustrated in Fig. 2. By adding a shape stream branch on the encoding path of the backbone network, shape-related information processing is performed to extract the edge features of the lung parenchyma. In the fusion module, the edge features are combined with the semantic features extracted by the backbone network to assist in more accurate localization of the target segmentation region. The original single-structure encoding convolutional block is modified to a module composed of multiple-scale convolutional kernels, enabling multi-scale feature extraction and enriching the network's acquisition of global and detailed features. Next, this paper will describe the

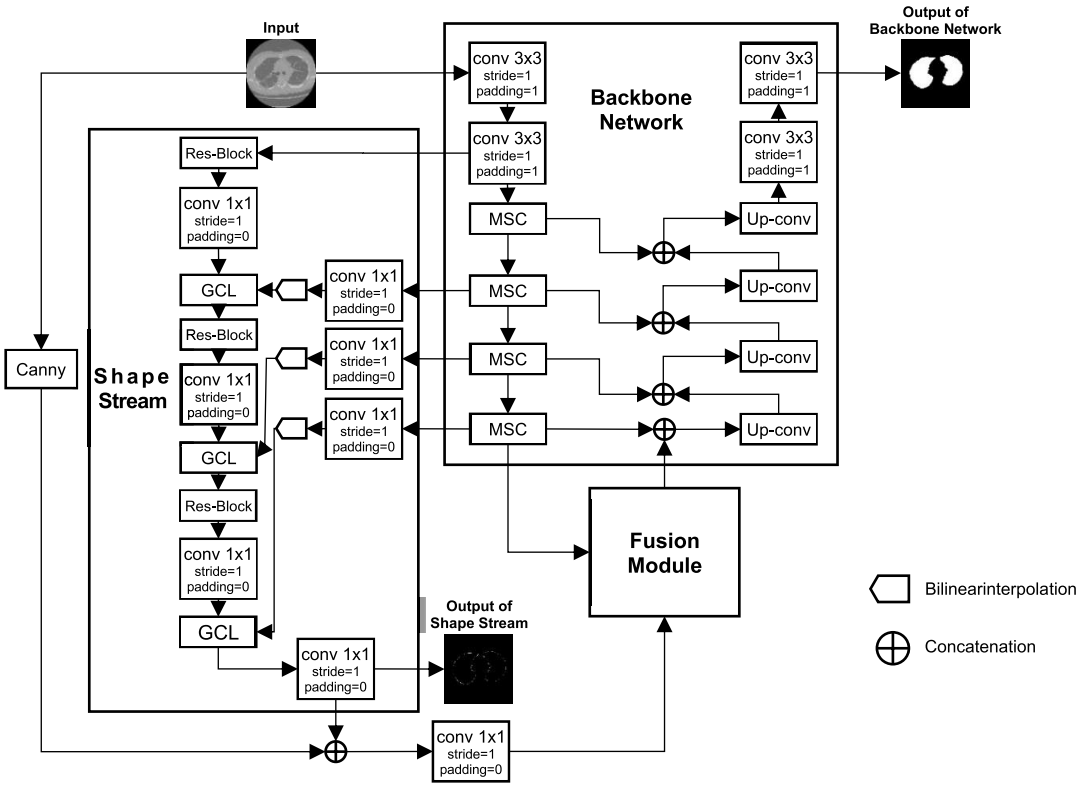


FIGURE 2. The general network structure of the proposed network. The proposed network consists of a backbone network, shape stream, and fusion module. The backbone network takes the CT image as input and outputs the final segmentation result. It contains an encoding path and a decoding path; The shape stream is composed of a few residual blocks interleaved with gated convolution layers (GCL). It obtains feature information from the backbone network, and finally outputs the result of the edge detection; The fusion module fuses the edge-related features extracted from the shape stream with the high-level features of the backbone network.

constituent structures of the network: the backbone network, the shape stream, and the fusion module.

The source code is available:
https://github.com/824779231/lungSeg.

1) BACKBONE NETWORK

This paper proposes an enhanced U-Net as the backbone network. The original U-Net follows an encoder-decoder structure. The encoding path performs feature extraction and the decoding path performs upsampling to map features to the output. Both the encoding and decoding paths consist of convolutional blocks, each block comprising two 3×3 convolutional layers. Each convolutional block reduces the feature map size by half and doubles the number of filters in the encoding path, while each convolutional block doubles the feature map size and halves the number of filters in the decoding path. The feature maps in the decoding path are connected through skip connections to the feature maps of the same size in the encoding path. U-Net combines shallow semantic features and high-level semantic features through skip connections, en-riching the feature hierarchy in the output layer. U-Net achieves image compression and noise removal by performing four downsampling operations, thereby saving hardware resources. Moreover, the original U-Net consists of only 23 convolutional layers, making it computationally efficient, hardware-friendly, and fast to train.

To achieve multi-scale feature extraction, this paper adapts the convolutional blocks in the encoding path into MSC blocks, as shown in Fig. 3. The first layer in the encoding path consists of two 3×3 convolutional layers. The

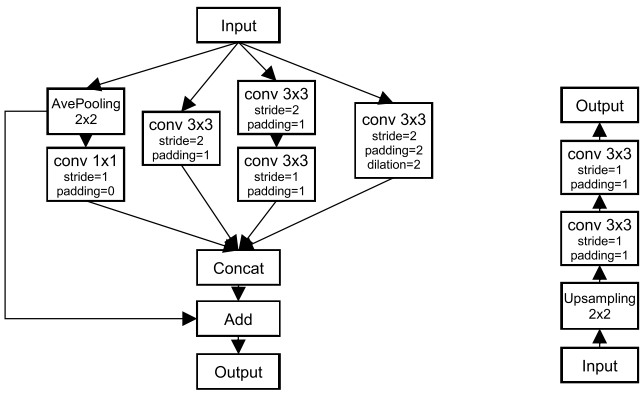


FIGURE 3. MSC block (left) and upsampling convolutional block (right). MSC block uses multiple convolution branches to process the feature map. The upsampling convolutional block restores the spatial resolution of the feature map layer by layer to generate the segmentation result.

other layers are MSC blocks, which include four branches: a 1 × 1 convolutional layer, a 3 × 3 convolutional layer, two 3 × 3 convolutional layers, and a 3 × 3 convolutional layer with a dilation rate of 2. The number of filters for each branch adds up to the number of filters in that layer. The ratio of the number of filters for each branch from left to right to the total number of filters is 0.125,0.125, 0.625, and 0.125, respectively. The structure of the MSC block is similar to the Inception module [38], with the difference that the 1 × 1 convolutional layer before the 3 × 3 convolutional layer is removed to prevent overfitting. Additionally, a residual branch is added to the block to sum the input with the concatenated result of four branches. The decoding path remains the same as the decoding path of the original U-Net. Except for the last layer, the other layers are upsampling convolutional blocks, as shown in Fig. 3. In the proposed backbone network, the number of filters of each layer of the encoding path is 32, 64, 128, 256, and 512

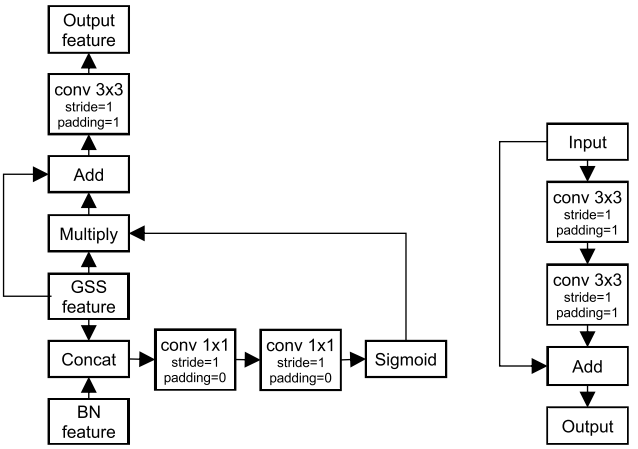


FIGURE 4. GCL (left) and residual block (right). The BN feature represents the features of the backbone network. The BN feature is concatenated with the GSS feature and then processed through two 1 × 1 convolutional layers. The resulting output is activated using the Sigmoid function, producing a one-channel attention map. This map is applied to GSS by GCL to obtain highly correlated shape feature information; By connecting input data directly to output data and using identity mapping, the residual block enables the network to learn the residual information.

2) SHAPE STREAM

Takikawa et al. [39] proposed a network architecture that combines the gated shape stream (GSS) with a CNN, referred to as gated shape CNNs (GSCNN). The GSS is composed of several residual blocks and gated convolutional layers (GCL). GSCNN can use any conventional CNN as the backbone network, integrating the GSS with different layers of the backbone network. Takikawa et al. ensured the consistency between edge detection and semantic segmentation by jointly supervising the edge detection task of the shape stream with semantic segmentation, leveraging the duality between these two tasks. Experimental results showed that the inclusion of GSS significantly improves the segmentation performance of thin and small objects.

In the proposed network, the GSS takes the output of the first layer of the backbone network as input and connects with the third, fourth, and fifth layers of the backbone network. Finally, the features from both streams are fused at multiple scales using the fusion module (see Fig. 2). The output of the first layer of the backbone network $m1 \in R^{N\times C1\times H\times W}$ is passed to the GSS, where it undergoes residual convolution and 1 × 1 convolution to form a feature map $s1 \in R^{N\times(C1/2)\times H\times W}$. The output $m3 \in R^{N\times C3\times H3\times W3}$ of the third layer of the backbone network is passed to the GSS, where it is processed through bilinear interpolation and 1 × 1 convolution to form a single-channel feature map $d3 \in R^{N\times 1\times H\times W}$. d3 is then processed together with s1 through the GCL, and the process continues similarly.

The core component of the shape stream is the GCL, as illustrated in Fig. 4. The GCL possesses the capability to effectively filter out irrelevant edge information, thereby assisting the shape stream in processing only relevant information. In the interaction with the deep layers of the backbone network, the shape stream captures the relevance of edge information from the high-level semantic features of the backbone network. By using the GCL to deactivate activations, irrelevant edge information is excluded. Since the backbone network performs advanced semantic analysis of the image and passes high-level features to the shape stream, the shape stream can focus only on the relevant regions. As a result, the shape stream can effectively process the image using a shallow architecture.

Let n denote the number of locations where the GCL is applied, and let $t \in 0, 1, \ldots\ldots, n$. Let $dt \in R^{N\times 1\times H\times W}$ and $st \in R^{N\times C\times H\times W}$ represent the intermediate states of the backbone network and the shape stream, respectively. To apply the GCL, dt and st are first concatenated. The concatenated features then go through convolution and normalization. The sigmoid function is applied to activate and generate an attention map $\alpha t \in R^{N\times 1\times H\times W}$, which are used to determine the relevance of edge information. By performing an element-wise multiplication between αt and st and adding the result to st, the output of the GCL is obtained after convolution.

Three GCLs are used in the proposed network, connecting to the third, fourth, and fifth layers of the backbone network. The output of the last GCL goes through convolution and sigmoid activation to generate the edge detection result. This result is then combined with the binary edge image obtained through canny edge detection of the input image. A feature map $\beta \in R^{N\times 1\times H\times W}$ is formed after convolution and sigmoid function, which is fused with the output of the fifth layer of the backbone network.

3) FUSION MODULE

The Atrous Spatial Pyramid Pooling module (ASPP) is used as the fusion module for integrating the backbone network and shape stream [40]. By applying convolution, pooling, and feature map resizing operations to the output $m5 \in R^{N\times C5\times H5\times W5}$ of the fifth layer in the encoding path of the

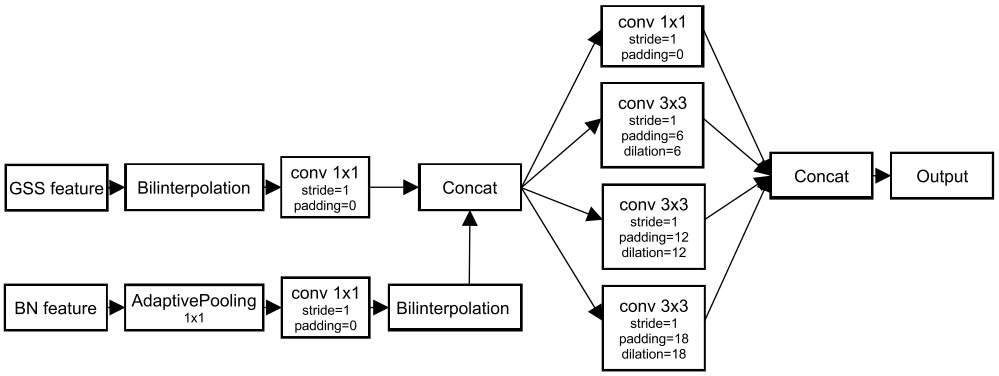


FIGURE 5. ASPP module. The BN feature represents the features of the backbone network. The ASPP module uses dilated convolutions of 4 sizes to process the concatenated results of the BN feature and GSS feature so that the shape feature information and the feature information of the backbone network are fully fused at multiple scales.

backbone network and the output β of the shape stream, they are combined to form $\gamma \in R^{N\times C\gamma\times H5\times W5}$. Four different sizes of convolutional kernels are utilized to perform multi-scale fusion of edge features and semantic features on γ. In the proposed network, the four convolutional kernels consist of a 1×1 kernel, a 3×3 kernel with a dilation rate of 6, a 3×3 kernel with a dilation rate of 12 and a 3×3 kernel with a dilation rate of 18, as illustrated in Fig. 5. Both the GSS feature and backbone network feature require 1×1 convolution to change the number of filters to 16 before concatenation. The number of filters for each branch is 24. Finally, the concatenated output of the four paths is passed to the decoding path for upsampling.

B. LOSS FUNCTION

To achieve a more accurate segmentation of lung parenchyma, a multi-task learning approach is employed. Both the segmentation task and edge detection task are jointly supervised.

The loss function of the segmentation task is defined as L_{area}. It utilizes a joint loss on predicted segmentation, which includes binary cross-entropy (BCE) loss L_{BCE_area} and DSC loss L_{DSC_area}, as shown in (1).

$$L_{area} = L_{BCE\_area} + L_{DSC\_area} \tag{1}$$

The loss function of the edge detection task is defined as L_{edg}, which utilizes DSC loss L_{DSC_edg} on predicted edge maps, as shown in (2).

$$L_{edg} = L_{DSC\_edg} \tag{2}$$

The overall loss function of this multi-task learning approach is defined as L, as shown in (3). Due to the edge detection loss can be relatively high, a weight 0.1 is assigned to L_{edg}.

$$L = L_{area} + 0.1 \times L_{edg} \tag{3}$$

C. DATASET

The original CT images used in the dataset are derived from the Open Source Imaging Consortium (OSIC) Pulmonary Fibrosis Progression dataset. Kónya et al. created ground truths for lung parenchyma segmentation on 111 patient samples (over 10,000 images) in this dataset. This subset is used as the dataset in the experiment. The voxels of the data in the dataset have a maximum of 7.5mm, a minimum of 0.5mm, and a median of 1.25mm. Kónya et al. created ground truths for these image sets in nrrd format (https://www.kaggle.com/sandorkonya/ct-lung-heart-trachea-segmentation).

In this study, the nrrd files are converted and saved in PNG format. We apply window level and window width processing (window level: 600, window width: 1200) to the CT images and scaled the pixel values to the range of 0-255, thereby highlighting the lung parenchyma region. The edge of the lung parenchyma region in the masks is extracted using the Canny edge detection algorithm, and these extracted edges are used as the ground truths for edge detection tasks. Due to the limitation of the processing power of the hardware equipment, the experiment does not use all samples but intercepts 6,000 samples as the training-validation dataset. The split ratio between the training dataset and the validation dataset is 3:1. In order to verify the performance of the network, CT scans of 10 patients (1468 images in total) are used as the test set.

D. NETWORK PARAMETERS AND TRAINING DETAILS

The input image size is set to 512×512. If the original size of the image is not 512×512, the following method will be used to adjust the image: firstly, the image is scaled equally using bilinear interpolation until the longer side of the image is equal to 512. Then the image will be padded with 0 pixels on the left and right or top and bottom sides centered around the image until the image size reaches 512×512. In the proposed network, batch normalization is applied before each convolution and ReLU function activation is used after each convolution.

The network has a parameter count of 7.9M. It is implemented by using the Python language and the PyTorch 1.6 framework, and the training process is accelerated by using

TABLE 1. Overall results of the ablation experiment. The scores in the table represent the mean of each metric. P-values are reported for the comparison of the proposed network with other networks. The significance level was set at 0.05.

Network	DSC	IOU	ACC	F1
U-Net	0.923899 (p=0.018)	0.847280 (p=0.002)	0.987554 (p<0.001)	0.963181 (p<0.001)
U-Net+GSS	0.930858 (p=0.050)	0.861050 (p=0.047)	0.988190 (p=0.022)	0.978672 (p<0.001)
U-Net+MSC	0.930495 (p=0.049)	0.862229 (p=0.048)	0.988363 (p=0.043)	0.977105 (p=0.037)
U-Net+MSC+GSS(Proposed network)	**0.932907**	**0.865887**	**0.988562**	**0.980520**

the NVIDIA Quadro P5000 GPU. The batch size for training is set to 6. The Adam optimizer is used for network optimization, with an initial learning rate of 1e-4. The learning rate decay schedule is set as (4).

$$\hat{\mathrm{lr}} = \mathrm{lr} \times (1 - \mathrm{bn}/\,(\mathrm{ep} \times \mathrm{len}))^{0.9} \tag{4}$$

where $\hat{\mathrm{lr}}$ represents the learning rate used in this epoch, lr represents the learning rate of the previous epoch, bn represents the number of decay steps, ep represents the total number of epochs and len represents the size of the training dataset.

The predefined number of training epochs is set to 200. If the loss does not decrease for 20 consecutive epochs, training is stopped. Each training epoch takes approximately 14 minutes.

IV. RESULTS

A. EVALUATION METRICS

This paper evaluates the segmentation performance using several metrics, including the DSC, Intersection over Union (IoU), Accuracy (ACC), F1-Score (F1), and 3D Dice Similarity Coefficient (3DDSC). DSC is a commonly used evaluation metric in medical image segmentation tasks, which measures the similarity between the predicted results and the ground truths. IoU represents the overlap ratio of the target regions between the predicted results and the ground truths. ACC calculates the proportion of correctly identified pixels in the predicted results out of the total number of pixels. F1 evaluates model performance by calculating the harmonic mean of precision and recall. 3DDSC is used to evaluate the similarity of two 3D image segmentation results. P-values are attached to the results. The range of these five metrics is between 0 and 1, where a value closer to 1 indicates better segmentation performance. On the basis of metric by metric, P-values are calculated to measure the statistically significant differences between the proposed network and other networks. P-values are calculated using the ttest_ind function in Python's scipy library.

To assess the performance of networks in segmenting complete lungs, we calculate the DSC and 3DDSC on a patient-by-patient basis. The DSC on a patient-by-patient basis is calculated as (5):

$$\overline{\mathrm{D}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{d_i}}{\mathrm{n}} \tag{5}$$

where $\overline{\mathrm{D}}$ represents the mean DSC for lung segmentation across all CT images of one patient, $\mathrm{d_i}$ represents the DSC for lung segmentation in each CT image per patient, n represents the number of CT images each patient possesses. Segmentation results from 2D slices for each patient are stacked to generate 3D images for calculating the 3DDSC.

B. ABLATION RESULTS

To investigate the impact of the network modules introduced in this study on network performance, ablation experiments are conducted on GSS and MSC. Results are presented in Table 1, Fig. 6, Fig. 7, and Table 2.

Table 1 shows the mean segmentation performance of each network over the entire test set. According to Table 1, it can be seen that the separate addition of MSC and GSS both improved the network's segmentation performance for the entire set. The combined utilization of GSS and MSC significantly enhances the network performance. The mean DSC is 0.91% higher than the baseline U-Net network, while the mean IoU is 1.9% higher. As all P-values are below the significance level, the statistically differences between the proposed network and other networks are quite significant.

To better verify the segmentation performance of each network on lung parenchyma of different sizes, this paper provides segmented statistical results of the ablation experiment, as shown in Fig. 6. It can be seen that the addition of GSS and MSC has improved the segmentation performance of lung parenchyma of different sizes from all the metrics, especially for relatively small lung parenchyma. In the range of 0-0.02 lung parenchyma proportion, the DSC, IoU, and F1 metrics significantly improve compared to the baseline network U-Net. The combined effect of MSC and GSS significantly enhances the segmentation performance of the network for small lung parenchyma, while further optimizing the segmentation results for larger lung parenchyma.

Fig. 7 presents visualized examples of the ablation results. From images (c) and (d), it can be observed that due to the presence of considerable noise within the lung parenchyma and strong interference from the surroundings, U-Net exhibits a noticeable under-segmentation issue. The incorporation of GSS and MSC has to some extent alleviated the under-segmentation issue of U-Net. In comparison, the effect of GSS is more pronounced. In summary,

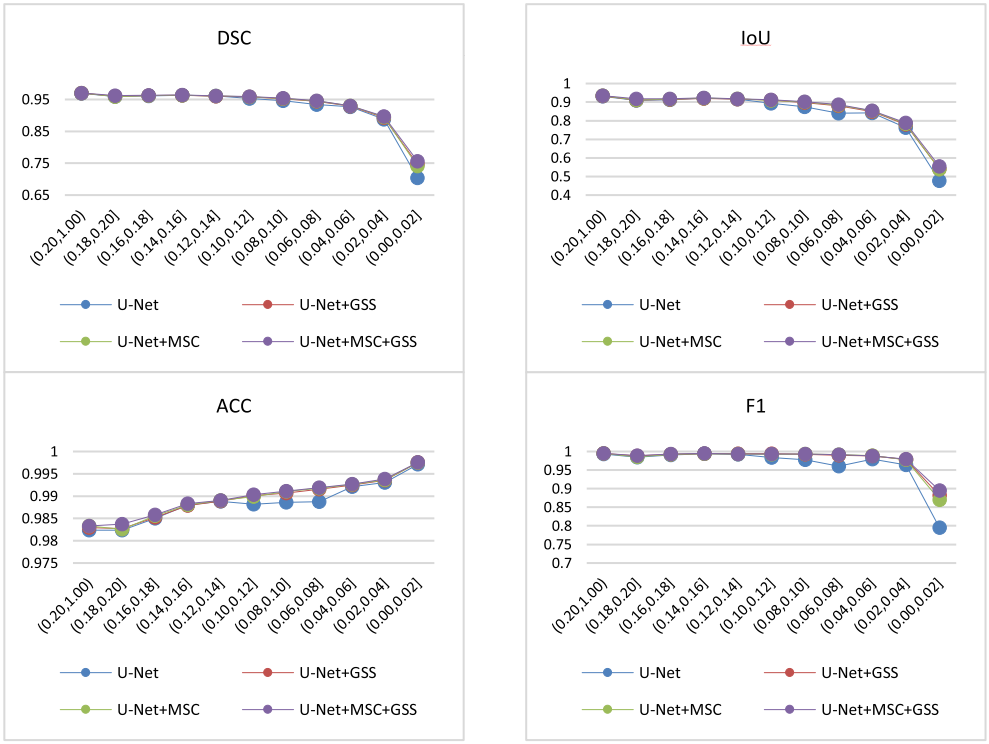


FIGURE 6. Segmented statistical results of the ablation experiment. It is found that the pixel proportion of relatively large lung parenchyma in CT images is typically above 0.2 but does not exceed 0.3. In light of this, this study employs a dynamic upper limit set at values greater than 0.2, with a step size of 0.02, to systematically assess the average segmentation performance of various networks on different-sized lung parenchyma segmentations.

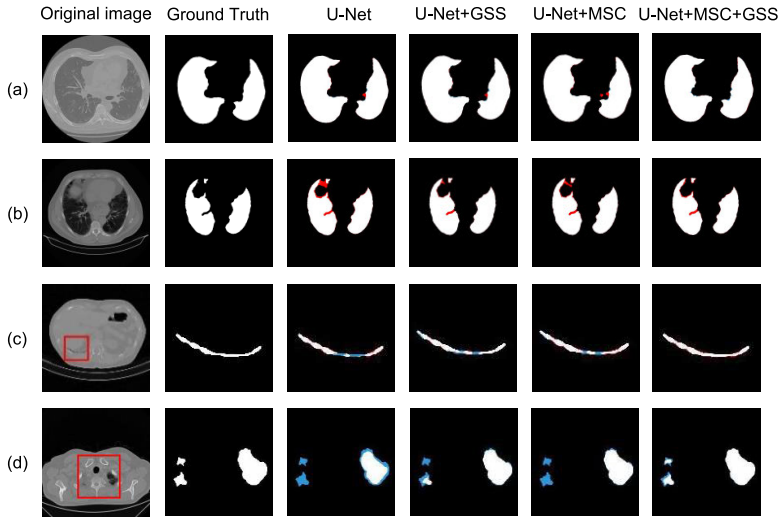


FIGURE 7. Test examples of the ablation experiment. White areas indicate the correctly segmented lung parenchyma, red areas represent false positive regions, and blue areas correspond to false negative regions. In order to magnify the segmentation performance of each network on relatively small lung parenchyma, only the corresponding masks and results within red outlined areas are displayed.

the segmentation performance of the network with both MSC and GSS is the best. The proposed network not only enables the network to perform well in the segmentation task of large and clear lung parenchyma but also shows significant progress in the segmentation task of small and blurry lung parenchyma.

The results on a patient-by-patient basis are shown in Table 2. It can be observed that the addition of GSS or MSC

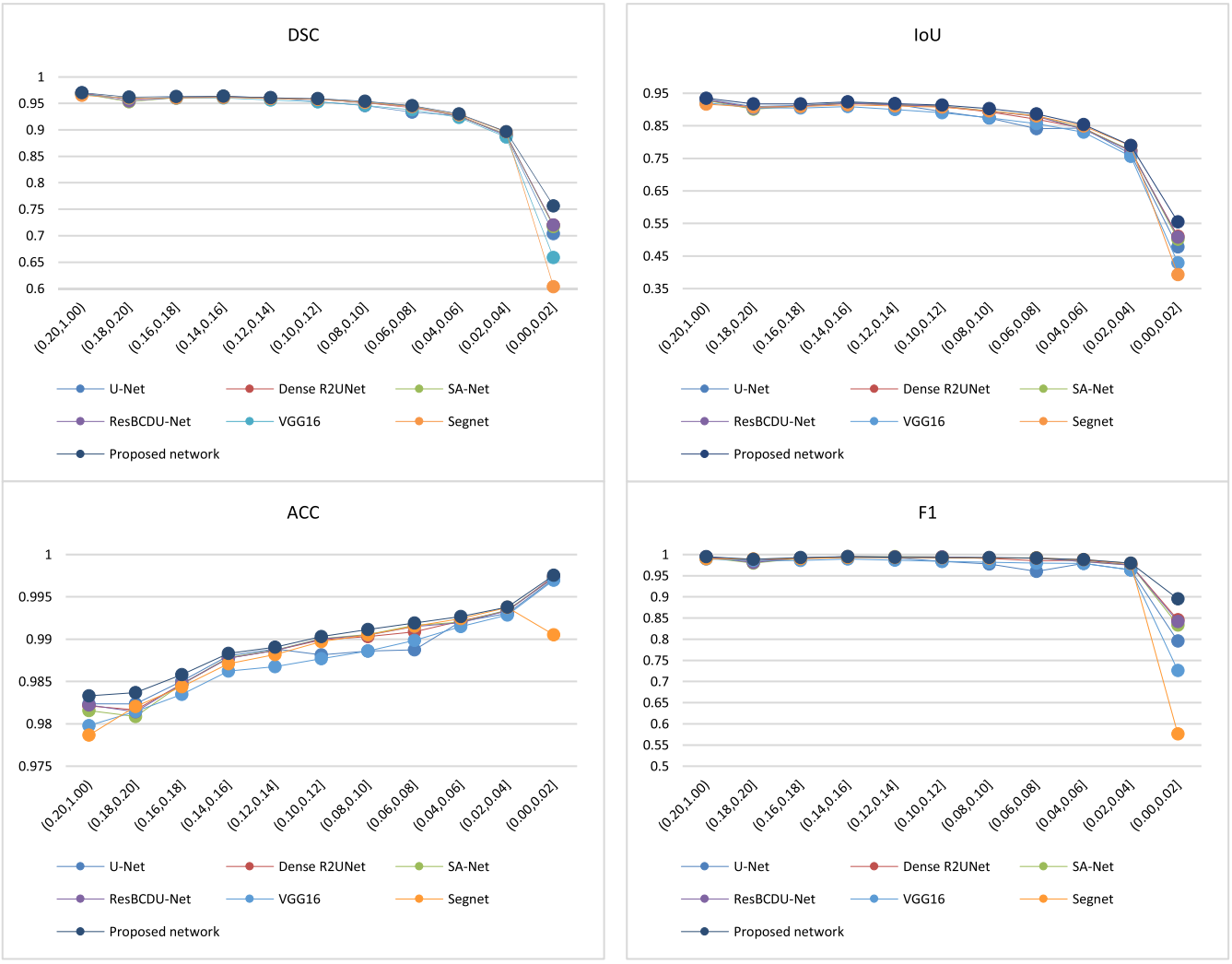


FIGURE 8. Segmented statistical results of different networks. It is found that the pixel proportion of relatively large lung parenchyma in CT images is typically above 0.2 but does not exceed 0.3. In light of this, this study employs a dynamic upper limit set at values greater than 0.2, with a step size of 0.02, to systematically assess the average segmentation performance of various networks on different-sized lung parenchyma segmentations.

TABLE 2. Results of the ablation experiment on a patient-by-patient basis. The scores in the table represent the mean of each metric.

Network	DSC	3DDSC
U-Net	0.921023	0.841090
U-Net+GSS	0.941518	**0.843843**
U-Net+MSC	0.937714	0.842166
U-Net+MSC+GSS	0.942720	0.842687

improves the segmentation results of the complete lungs, as evidenced by the notable increase in mean DSC and mean 3DDSC. Moreover, the simultaneous addition of GSS and MSC further enhances the overall lung segmentation performance. The proposed network achieves a mean DSC and mean 3DDSC that are 2.17% and 1.7% higher than the baseline U-Net, respectively. This illustrates that the proposed network promotes complete lung segmentation.

C. COMPARISON WITH OTHER NETWORKS

To validate the superiority of the proposed network, it is compared with recent deep learning-based lung parenchyma segmentation networks. To ensure a fair comparison, the same training-validation set is used, and their performances are evaluated on the test set. The experimental results are presented in Table 3, Fig. 8, Fig. 9, Fig. 10, Fig. 11, and Table 4.

Table 3 shows the mean segmentation performance of each network over the entire test set. According to Table 3, it can be observed that all metrics of the proposed network are the highest. In particular, the IOU of the proposed network is 1.2% higher than that of the second-best network. This suggests that among these advanced networks, the proposed network still maintains its overall superiority. Similarly, with all P-values below the significance level, the differences between the proposed network and other networks are significant.

TABLE 3. Overall results of different networks. The scores in the table represent the mean of each metric. P-values are reported for comparison with other networks. P-values are reported for the comparison of the proposed network with other networks. The significance level was set at 0.05.

Network	DSC	IOU	ACC	F1
U-Net [3]	0.923899 (p=0.018)	0.847280 (p=0.002)	0.987554 (p<0.001)	0.963181 (p<0.001)
Dense R2Unet [29]	0.926538 (p=0.041)	0.853992 (p=0.042)	0.987712 (p=0.002)	0.973167 (p=0.027)
SA-Net [30]	0.925878 (p=0.032)	0.853034 (p=0.031)	0.987582 (p<0.001)	0.971469 (p=0.009)
ResBCDU-Net [31]	0.926421 (p=0.033)	0.854195 (p=0.038)	0.987752 (p<0.001)	0.972797 (p=0.023)
VGG16 [41]	0.916888 (p=0.002)	0.834657 (p<0.001)	0.986300 (p<0.001)	0.953657 (p<0.001)
Segnet [42]	0.912005 (p=0.001)	0.837496 (p<0.001)	0.985999 (p<0.001)	0.940275 (p<0.001)
Proposed network	**0.93290**	**0.865887**	**0.988562**	**0.980520**

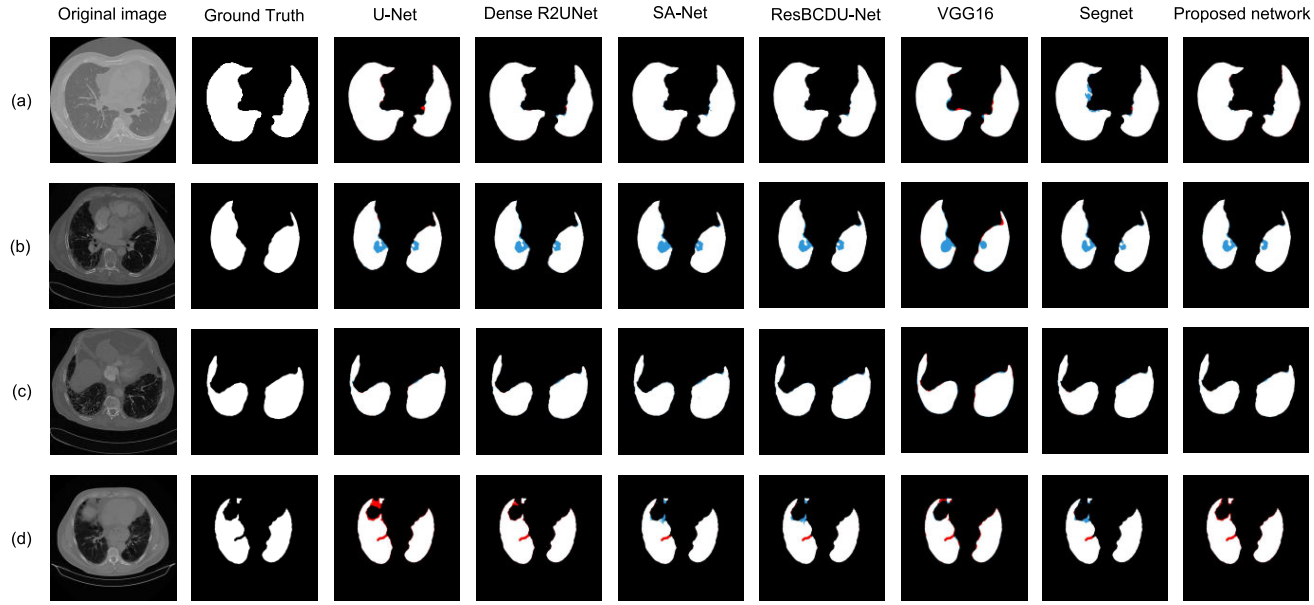


FIGURE 9. Segmentation cases of different networks for relatively large lung parenchyma. White areas indicate the correctly segmented lung parenchyma, red areas represent false positive regions, and blue areas correspond to false negative regions.

From the segmented statistical results presented in Fig. 9, it is evident that the proposed network exhibits an advantage in segmentation performance across all sizes. In particular, as the lung parenchyma gradually decreases in size, the DSC, IoU, and F1 metrics of the proposed network gradually widen the gap with other networks. This gap reaches its peak within the range of 0-0.02 lung parenchyma proportion. This effectively demonstrates the meaning of the proposed network: achieving segmentation results that cater to lung parenchyma of varying sizes.

Fig. 9 to 11 present visualized examples of different networks. From Fig. 9, it can be seen that the proposed network performs similarly to other networks in the segmentation of relatively large lung parenchyma regions. Fig. 10 illustrates the segmentation results of each network on punctate lung parenchyma regions. For relatively small lung parenchyma regions, Segnet exhibits severe over-segmentation. Other networks are more prone to under-segmentation when dealing with such cases of ambiguous boundaries. This phenomenon is most evident in the image (d). In contrast, the proposed network can segment the target area more comprehensively and accurately. Fig. 11 illustrates the segmentation results of each network on elongated lung parenchyma regions. In addition to the still evident over-segmentation issue in Segnet, other networks tend to exhibit incomplete segmentation at both ends when dealing with the segmentation of elongated lung parenchyma regions (as evident in images (b) and (d)). There is also a problem of fragmented segmentation results

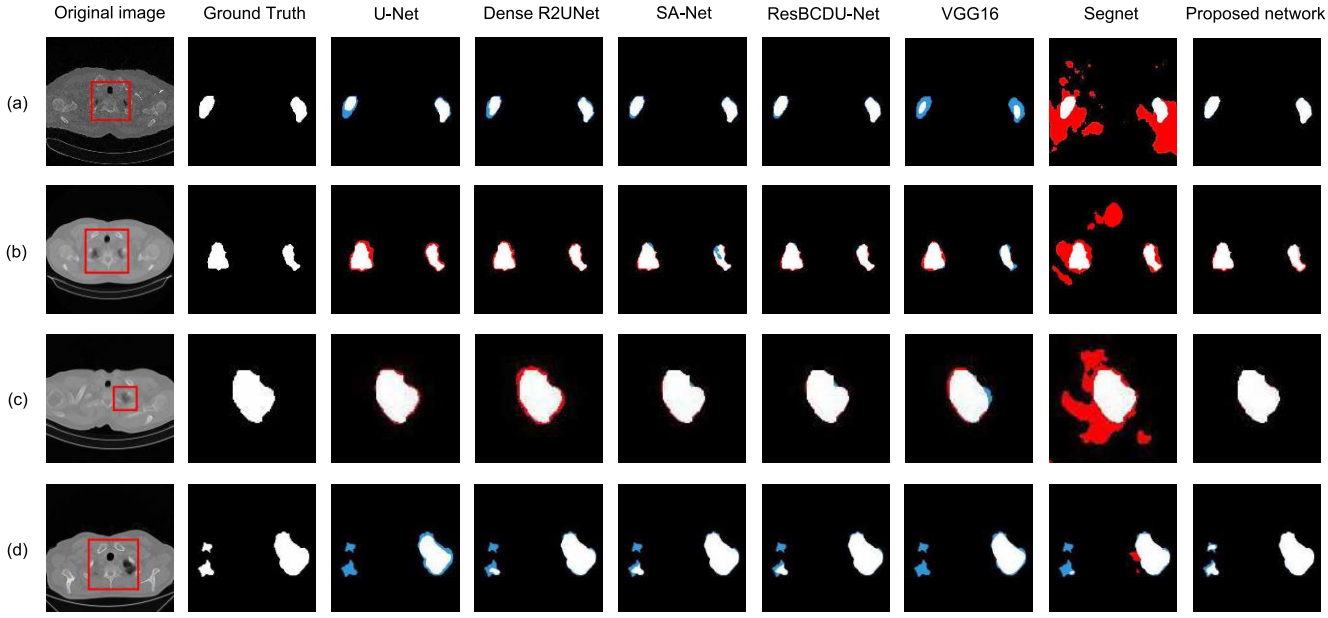


FIGURE 10. Segmentation cases of different networks for punctate lung parenchyma. White areas indicate the correctly segmented lung parenchyma, red areas represent false positive regions, and blue areas correspond to false negative regions. In order to magnify the segmentation performance of each network on relatively small lung parenchyma, only the corresponding masks and results within red outlined areas are displayed.

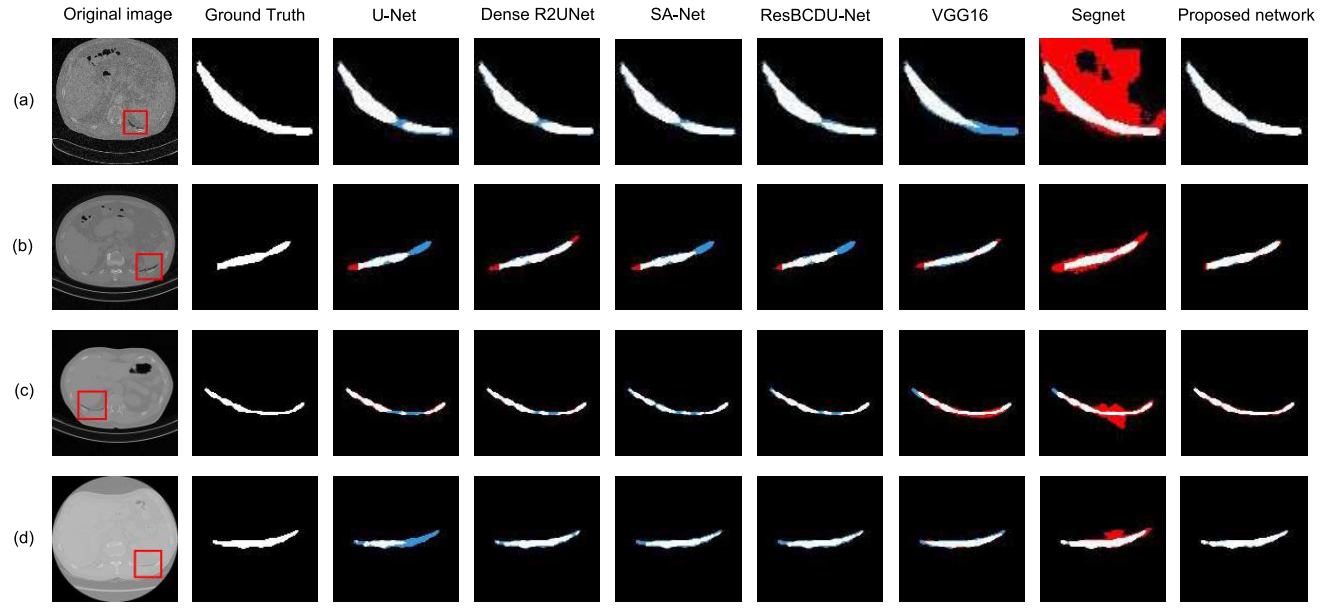


FIGURE 11. Segmentation cases of different networks for elongated lung parenchyma. White areas indicate the correctly segmented lung parenchyma, red areas represent false positive regions, and blue areas correspond to false negative regions. In order to magnify the segmentation performance of each network on relatively small lung parenchyma, only the corresponding masks and results within red outlined areas are displayed.

(as noticeable in images (a) and (c)). It is obvious that the proposed network handles these issues most effectively by contrast.

The results on a patient-by-patient basis are shown in Table 4. The mean DSC of the proposed network is 0.63% higher than that of the second-best network. While the mean 3DDSC of the proposed network is only 0.16% lower than that of the best network. The results indicate that the proposed network demonstrates good performance in segmenting complete lungs.

In addition, this paper compares the parameter size and the execution time on the test dataset of each network, as shown in Table 5. As the MSC does not increase the number of filters in each layer and GSS is merely a shallow and simple network

TABLE 4. Results of the different networks on a patient-by-patient basis. The scores in the table represent the mean of each metric.

Network	DSC	3DDSC
U-Net [3]	0.921023	0.841090
Dense R2Unet [29]	0.936361	**0.844315**
SA-Net [30]	0.927065	0.841066
ResBCDU-Net [31]	0.930564	0.841757
VGG16 [41]	0.933733	0.841984
Segnet [42]	0.923618	0.826571
Proposed network	**0.942720**	0.842687

TABLE 5. Comparison of computational costs for different networks.

Network	Parameter size	Execution time(s)
U-Net [3]	7472289	336.76
Dense R2Unet [29]	28728483	450.93
SA-Net [30]	11326017	403.21
ResBCDU-Net [31]	127457283	423.79
VGG16 [41]	21757057	395.71
Segnet [42]	29442433	379.50
Proposed network	7901311	371.77

structure, the proposed network does not have significantly more parameters compared to U-Net. It can be seen that the parameter size and the execution time of the proposed network are only higher than those of U-Net, with a parameter size increase of only 5.7%. In contrast, the parameter size of other networks is higher by one to two orders of magnitude compared to the proposed network. Therefore, not only does the proposed network outperform other networks in terms of segmentation performance, but it also incurs relatively lower computational costs.

V. DISCUSSION

Existing lung parenchyma segmentation methods often work well for CT images with a large proportion of lung parenchyma. Although CT images with a large proportion of lung parenchyma typically contain most of the pathological information, it does not mean that CT images with a small proportion of lung parenchyma can be ignored. The work presented in this paper addresses these considerations.

This paper presents an enhanced lung parenchyma segmentation network based on U-Net. The network significantly improves the segmentation accuracy of small and blurry lung parenchyma regions while maintaining the segmentation accuracy of large and clear regions. The network improvement consists of two aspects: first, adding a shape stream branch to the encoding path of the backbone network; second, replacing the convolutional blocks in the encoding path of the backbone network with MSC blocks.

Traditional segmentation methods rely on manually setting thresholds for pixel-level classification, suitable for high-contrast and simple scene images. The proposed method leverages deep neural networks to automatically learn complex features and structures in images, making it suitable for handling complex scenes and large-scale datasets. Take the work of Walczak et al. [6] as an example. To achieve good segmentation results, their method consists of numerous steps, including false-positive elimination, lung segmentation, and post-processing. While this method executes faster due to its lower parameter count, it struggles to handle complex scenes effectively. On the other hand, the proposed network contains a large number of parameters and can adapt to all possible complex scenes through automatic training. This method also eliminates the need for intricate preprocessing and postprocessing. A single network is sufficient to perform the segmentation task effectively.

The original U-Net employs only two 3×3 convolutional layers in each layer. This simplistic feature extraction module structure is limited in its ability to capture image features across multiple scales. A large receptive field allows each feature point to perceive a large region of information from the input image, which is beneficial for handling large objects. However, a large receptive field also tends to downplay fine-grained details and overlook small objects. In the task of lung parenchyma segmentation, large receptive fields help improve the segmentation accuracy of large and clear lung parenchyma regions while reducing the receptive field appropriately enhances the segmentation accuracy of small and blurry lung parenchyma regions. In this paper, multiple convolutional kernels of different sizes are used in each convolutional block of the network's encoding path to address the limitation of a single receptive field. As a result, the proposed network can accommodate lung parenchyma segmentation tasks of varying sizes and shapes.

Traditional medical image segmentation networks typically have a single encoding-decoding path, where color, shape, and texture features are processed together within a single CNN. Such networks fail to efficiently handle all types of features, leading to limitations in the performance of segmentation networks. There is a duality between edge detection and semantic segmentation tasks. High-level semantic features from the semantic segmentation task can help exclude irrelevant edge information in edge detection, and accurate edge information can assist the semantic segmentation task in accurately locating the target region. In this paper, a shape stream is added to the encoding path of the backbone network, specifically designed for handling shape-related information. The shape stream is a shallow and simple network structure. It interacts with multiple deep layers of the backbone network, extracting the correlation of edge information from high-level features of the backbone network. Irrelevant edge information is filtered out by utilizing GCLs to deactivate activations, allowing the simple architecture to achieve clear and accurate target edge information. The edge detection result from the shape stream is fused with the backbone network to provide clear boundary information for the segmentation task, resulting in more accurate segmentation of the lung parenchyma region.

One limitation of this network is that the segmentation accuracy for small and blurry lung parenchyma regions has not yet reached the level attained for large and clear regions. Moreover, the proposed network is a two-stream structure. Under the same parameter size, the network with a multi-stream structure will occupy more graphics card memory. This is also a drawback of GSCNN. This work not only improves the segmentation performance of small and blurry lung parenchyma but also inspires other medical image segmentation tasks involving small targets, such as nodule and tumor segmentation. Research related to medical science demands high levels of rigor and precision, and scholars are always pursuing more sophisticated CAD systems. By improving the shortcomings of lung parenchyma segmentation tasks through this work, we can make the CAD system for lung diseases more sophisticated. In the future, this work can help people pay more attention to the ''ignored'' parts of medical images, and the proposed network can be applied to other segmentation tasks involving small targets.

VI. CONCLUSION

In lung parenchyma segmentation tasks, the segmentation performance of small and blurry lung parenchyma regions is often overlooked. This paper addresses this issue by incorporating a shape stream into the backbone network and improving the convolutional blocks in the encoding path to be MSC blocks. The proposed network's performance has been validated on the Kaggle OSIC pulmonary fibrosis progression prediction competition dataset. Experimental results demonstrate that the network significantly improves the segmentation performance of small and blurry lung parenchyma regions while maintaining the segmentation performance of large and clear lung parenchyma regions.

ACKNOWLEDGMENT

The authors would like to thank the Open Source Imaging Consortium (OSIC) for providing the OSIC Pulmonary Fibrosis Progression dataset by the Kaggle website.

REFERENCES

[1] H. P. Ng, S. Huang, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, ''Medical image segmentation using watershed segmentation with texture-based region merging,'' in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 4039–4042.

[2] C. Kotropoulos and I. Pitas, ''Segmentation of ultrasonic images using support vector machines,'' *Pattern Recognit. Lett.*, vol. 24, nos. 4–5, pp. 715–727, Feb. 2003.

[3] B. Ait Skourt, A. El Hassani, and A. Majda, ''Lung CT image segmentation using deep neural networks,'' *Proc. Comput. Sci.*, vol. 127, pp. 109–113, Feb. 2018.

[4] G. Wang, S. Guo, L. Han, and A. B. Cekderi, ''Two-dimensional reciprocal cross entropy multi-threshold combined with improved firefly algorithm for lung parenchyma segmentation of COVID-19 CT image,'' *Biomed. Signal Process. Control*, vol. 78, Sep. 2022, Art. no. 103933.

[5] X. Yang, G. Xu, and T. Zhou, ''An effective approach for CT lung segmentation using region growing,'' *J. Phys., Conf. Ser.*, vol. 2082, no. 1, Nov. 2021, Art. no. 012001.

[6] M. Walczak, I. Burda, J. Nalepa, and M. Kawulok, ''Segmenting lungs from whole-body CT scans,'' in *Proc. Int. Conf. Bey. Data., Archit. Struc.* Cham, Switzerland: Springer, 2017, pp. 403–414.

[7] S. Hu, E. A. Hoffman, and J. M. Reinhardt, ''Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images,'' *IEEE Trans. Med. Imag.*, vol. 20, no. 6, pp. 490–498, Jun. 2001.

[8] I. Siddique, I. S. Bajwa, M. S. Naveed, and M. A. Choudhary, ''Automatic functional brain MR image segmentation using region growing and seed pixel,'' in *Proc. 4th Int. Conf. Inf. Commun. Technol.*, Dec. 2006, pp. 1–2.

[9] W. Deng, W. Xiao, H. Deng, and J. Liu, ''MRI brain tumor segmentation with region growing method based on the gradients and variances along and inside of the boundary curve,'' in *Proc. 3rd Int. Conf. Biomed. Eng. Informat.*, vol. 1, Oct. 2010, pp. 393–396.

[10] S. Poonguzhali and G. Ravindran, ''A complete automatic region growing method for segmentation of masses on ultrasound images,'' in *Proc. Int. Conf. Biomed. Pharmaceutical Eng.*, Dec. 2006, pp. 88–92.

[11] Y. Pathak, P. K. Shukla, A. Tiwari, S. Stalin, S. Singh, and P. K. Shukla, ''Deep transfer learning based classification model for COVID-19 disease,'' *IRBM*, vol. 43, no. 2, pp. 87–92, Apr. 2022.

[12] P. Campadelli, E. Casiraghi, and D. Artioli, ''A fully automated method for lung nodule detection from postero-anterior chest radiographs,'' *IEEE Trans. Med. Imag.*, vol. 25, no. 12, pp. 1588–1603, Dec. 2006.

[13] A. M. Mendonça, J. A. da Silva, and A. Campilho, ''Automatic delimitation of lung fields on chest radiographs,'' in *Proc. 2nd IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2004, pp. 1287–1290.

[14] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, ''Segment anything in medical images,'' 2023, *arXiv:2304.12306*.

[15] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, ''Segment anything,'' 2023, *arXiv:2304.02643*.

[16] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[17] J. Dolz, C. Desrosiers, and I. B. Ayed, ''IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet,'' in *Proc. Int. Workshop Challenge Comput. Methods Clin. Appl. Spine Imag.* Cham, Switzerland: Springer, 2018, pp. 130–143.

[18] X. Li, W. Qian, D. Xu, and C. Liu, ''Image segmentation based on improved unet,'' *J. Phys., Conf. Ser.*, vol. 1815, no. 1, Feb. 2021, Art. no. 012018.

[19] G. Xu, X. Wu, X. Zhang, and X. He, ''LeViT-UNet: Make faster encoders with transformer for medical image segmentation,'' 2021, *arXiv:2107.08623*.

[20] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, ''RIC-Unet: An improved neural network based on unet for nuclei segmentation in histology images,'' *IEEE Access*, vol. 7, pp. 21420–21428, 2019.

[21] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, ''ResUNet—A: A deep learning framework for semantic segmentation of remotely sensed data,'' *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020.

[22] Y. Cao, S. Liu, Y. Peng, and J. Li, ''DenseUNet: Densely connected UNet for electron microscopy image segmentation,'' *IET Image Process.*, vol. 14, no. 12, pp. 2682–2689, Oct. 2020.

[23] Y. Weng, T. Zhou, Y. Li, and X. Qiu, ''NAS-Unet: Neural architecture search for medical image segmentation,'' *IEEE Access*, vol. 7, pp. 44247–44257, 2019.

[24] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, ''UNet3+: A full-scale connected UNet for medical image segmentation,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[25] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, ''NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,'' *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[26] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, ''Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal,'' *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 568–576, Feb. 2020.

[27] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, ''TransUNet: Transformers make strong encoders for medical image segmentation,'' 2021, *arXiv:2102.04306*.

[28] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, ''Swin-Unet: Unet-like pure transformer for medical image segmentation,'' 2021, *arXiv:2105.05537*.

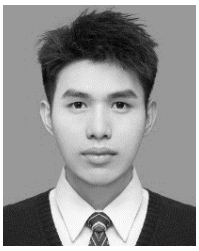
[29] K. Dutta, "Densely connected recurrent residual (dense R2UNet) convolutional neural network for segmentation of lung CT images," 2021, *arXiv:2102.00663*.
[30] J. Hu, H. Wang, J. Wang, Y. Wang, F. He, and J. Zhang, "SA-Net: A scale-attention network for medical image segmentation," *PLoS One*, vol. 16, no. 4, Apr. 2021, Art. no. e0247388.
[31] Y. Jalali, M. Fateh, M. Rezvani, V. Abolghasemi, and M. H. Anisi, "ResBCDU-Net: A deep learning framework for lung CT image segmentation," *Sensors*, vol. 21, no. 1, p. 268, Jan. 2021.
[32] A. Aghamohammadi, R. Ranjbarzadeh, F. Naiemi, M. Mogharrebi, S. Dorosti, and M. Bendechache, "TPCNN: Two-path convolutional neural network for tumor and liver segmentation in CT images using a novel encoding approach," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115406.
[33] K. Gao, J. Su, Z. Jiang, L.-L. Zeng, Z. Feng, H. Shen, P. Rong, X. Xu, J. Qin, Y. Yang, W. Wang, and D. Hu, "Dual-branch combination network (DCN): Towards accurate diagnosis and lesion segmentation of COVID-19 using CT images," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101836.
[34] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 695–704.
[35] M. Y. Ansari, Y. Yang, S. Balakrishnan, J. Abinahed, A. Al-Ansari, M. Warfa, O. Almokdad, A. Barah, A. Omer, A. V. Singh, P. K. Meher, J. Bhadra, O. Halabi, M. F. Azampour, N. Navab, T. Wendler, and S. P. Dakua, "A lightweight neural network with multiscale feature enhancement for liver CT segmentation," *Sci. Rep.*, vol. 12, no. 1, p. 14153, Aug. 2022, doi: 10.1038/s41598-022-16828-6.
[36] S. Gao, M. Cheng, K. Zhao, X. Zhang, M. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.
[37] H. Xia, W. Sun, S. Song, and X. Mou, "Md-Net: Multi-scale dilated convolution network for CT images segmentation," *Neural Process. Lett.*, vol. 51, no. 3, pp. 2915–2927, Apr. 2020.
[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[39] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5228–5237, doi: 10.1109/ICCV.2019.00533.
[40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.
[41] L. Geng, S. Zhang, J. Tong, and Z. Xiao, "Lung segmentation method with dilated convolution based on VGG-16 network," *Comput. Assist. Surg.*, vol. 24, no. 2, pp. 27–33, Oct. 2019.
[42] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

LUN ZHU was born in 1974. He received the B.Eng. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China.

He is currently a Senior Engineer with Changzhou University, Changzhou, China. His main research interests include data mining, embedded systems, and the application of the Internet of Things technology.

YINGHUI CAI was born in 1997. She received the B.Eng. degree from Zhejiang Shuren University, Hangzhou, China. She is currently pursuing the M.Eng. degree with Changzhou University, under the supervision of Prof. Lun Zhu.

Her main research interests include machine learning, biometrics, image processing, and deep learning.



JIAHAO LIAO was born in 1998. He received the B.Eng. degree from Zhejiang Shuren University, Hangzhou, China. He is currently pursuing the M.Eng. degree with Changzhou University, under the supervision of Prof. H. Y. Wang.

His main research interests include image processing, deep learning, and image segmentation.

FAN WU was born in 1978. He received the M.Eng. degree from the Zhejiang University of Technology, Hangzhou, China.

He is currently an Associate Professor with Zhejiang Shuren University, Hangzhou. His main research interests include digital twins and artificial intelligence.