RESEARCH ARTICLE

ResDSda_U-Net: A Novel U-Net-Based Residual Network for Segmentation of Pulmonary Nodules in Lung CT Images

ZHANLIN JI[1,2], **(Member, IEEE), ZIHENG ZHAO**[1]**, XINYI ZENG**[1]**, JINGKUN WANG**[3]**,**
LI ZHAO[3]**, XUEJI ZHANG**[4]**, AND IVAN GANCHEV**[2,5,6]**, (Senior Member, IEEE)**

[1]Hebei Key Laboratory of Industrial Intelligent Perception, North China University of Science and Technology, Tangshan 063210, China
[2]Telecommunications Research Centre (TRC), University of Limerick, Limerick, V94 T9PX Ireland
[3]Beijing National Research Center for Information Science and Technology, Institute for Precision Medicine, Tsinghua University, Beijing 100084, China
[4]School of Biomedical Engineering, Shenzhen University Health Science Center, Shenzhen, Guangdong 518060, China
[5]Department of Computer Systems, University of Plovdiv "Paisii Hilendarski," 4000 Plovdiv, Bulgaria
[6]Institute of Mathematics and Informatics—Bulgarian Academy of Sciences, 1040 Sofia, Bulgaria

Corresponding authors: Xinyi Zeng (zhangxueji@szu.edu.cn) and Ivan Ganchev (ivan.ganchev@ul.ie)

ABSTRACT The timely detection and segmentation of pulmonary nodules in lung computed tomography (CT) images can aid in the early diagnosis and treatment of lung cancer. However, manual segmentation of pulmonary nodules by doctors is highly demanding in terms of operational requirements and efficiency. To effectively improve the pulmonary nodule segmentation, this paper proposes a novel neural network, called ResDSda_U-Net, based on the U-Net network with the following improvements: (1) combining a Depthwise Over-parameterized Convolutional layer (DO-Conv) with a simple parameter-free attention module (SimAM), in the form of a newly designed ResDS block; (2) incorporating a denser Dense Atrous Spatial Pyramid Pooling (DASPP) module, between the encoder and decoder, using modified dilated rates to extract multi-scale information more effectively; and (3) adding channel and spatial attention mechanisms to the decoder, in the form of newly designed Convolution and Channel Attention (CCA) and Convolution and Spatial Attention (CSA) blocks, to enhance global pixel attention, fully capture global contextual information, and enable the decoder to better eliminate differences between pixels. The conducted experiments demonstrate that the proposed ResDSda_U-Net network outperforms all existing U-Net based networks (according to all evaluation metrics used) and all considered state-of-the-art networks (according to half of the evaluation metrics), by achieving corresponding values of 86.65% for the Dice Similarity Coefficient (DSC), 76.73% for Intersection over Union (IoU), 86.30% for sensitivity, and 87.22% for precision.

INDEX TERMS Computed tomography image, pulmonary nodule segmentation, residual neural network, attention mechanism.

I. INTRODUCTION

Pulmonary nodules are one of the leading causes of high mortality and healthcare utilization worldwide, [1]. Pulmonary

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

nodules refer to irregular circular shadows of various sizes, with clear or blurred edges, and a diameter of less than or equal to 3 cm in medical computed tomography (CT) pulmonary images. The development of pulmonary nodules is the result of the interplay between unknown antigens and the body's cellular and humoral immune functions, [2], [3], [4].

Most of the pulmonary diseases that people commonly suffer from are caused by the formation of pulmonary nodules. According to the basic classification of pulmonary nodule diseases, we can categorize the etiology of pulmonary nodules into benign and malignant conditions. The mortality rate of malignant pulmonary nodules is 23% among patients with pulmonary nodules. Despite significant research advancements and the development of modern society, the overall cure rate for pulmonary nodules remains at only 10% [5]. Therefore, it is crucial to detect, diagnose, and treat pulmonary nodules disease as early as possible. However, in the early stages of diagnosis, small pulmonary nodules in pulmonary CT images are considered a key clue to diagnosing primary diseases, but not all pulmonary nodules are malignant, so the identification and detection of pulmonary nodules is the most important part. Currently, chest CT and positron emission tomography-computed tomography (PET-CT) are the most reliable and sensitive imaging methods for detecting pulmonary nodules, [6], [7], [8]. Conventional CT scanning has extremely high sensitivity for detecting pulmonary nodules, and it is mostly used for examining high-risk patients, [9]. CT-enhanced scanning can reveal the basic conditions of pulmonary nodules and their surrounding environment. By clearly observing their subtle structures and combining corresponding examination data, doctors can diagnose the location, texture, size, micro-level details, malignancy degree, internal structure, spiculation, calcification, lobulation, sphericity, and edge conditions of pulmonary nodules, thereby grasping the basic situation of the nodules, [10]. However, this technology not only has low detection efficiency but also can cause fatigue and affect diagnosis if doctors use it for a long time, leading to missed or misdiagnosed cases, [11], [12].

Therefore, many researchers have developed more accurate and time-saving algorithms for the segmentation of pulmonary nodules. Earlier programs relied on manual segmentation, but it has been found through practice that manual segmentation is not only time-consuming and tedious, but also requires a large number of medical experts to perform the task. Additionally, disease annotation is a highly subjective and variable task, often influenced by clinical experience and other factors, which can affect standardization, [13], [14]. Researchers have been studying algorithms for automatic segmentation in order to alleviate the burden of manual segmentation. In recent years, more and more research has focused on the segmentation of pulmonary nodules in pulmonary CT images. However, due to the complex structure of the pulmonary and the small size of pulmonary nodules, various obstacles have not been solved, such as pulmonary structure segmentation, segmentation near branching and intersection, and pulmonary nodule segmentation, [15]. Recent studies have shown that deep learning [16] exhibits excellent performance in medical image segmentation tasks.

In this paper, we propose a novel U-Net [17] based neural network, named ResDSda_U-Net, which utilizes skip connections to combine low-resolution semantic in-formation with high-resolution local spatial information. However, existing U-Net architectures have limitations such as a relatively smaller number of backbone layers and inferior feature extraction capabilities compared to other networks, which can lead to lower accuracy in identifying adjacent pixels of different classes during segmentation. The identification of non-nodule structures in the surrounding environment can lead to erroneous recognition and subsequently decrease the accuracy of the segmentation process. To address these issues, we incorporate a Depthwise Over-parameterized Convolutional layer (DO-Conv) and multiple attention mechanisms into U-Net to enhance the feature extraction, which is identified in [18] as ''the single most important factor in achieving high performance''. As pulmonary nodules and non-nodules are highly similar, there is a significant risk of misdiagnosis, along with other factors influencing correct diagnosis. Therefore, we use a Dense Atrous Spatial Pyramid Pooling (DASPP) module [19], which was slightly modified (i.e., made denser) to better capture contextual features [20], accurately identify the contour of pulmonary nodules [21], and more precisely recognize the shape and location of pulmonary nodule lesions.

The main contributions of this paper are the following:

1) Due to the weak generalization ability and poor feature extraction capability of the network backbone, a novel ResDS block is elaborated for addition to the U-Net backbone. The use of several such ResDS blocks in the proposed ResDSda_U-Net network allows it to improve its convolutional layer using a deep residual DO-Conv layer, which enhances the network's generalization ability and adaptability, without increasing computational complexity, thereby allowing it to better handle various types and sizes of input images. Furthermore, a simple parameter-free attention module (SimAM) [22] is incorporated into each ResDS block, which enables enhancing the feature extraction ability of the encoder without increasing computational workload.

2) A denser DASPP module is employed after the lowest-level feature extraction, as it has the ability to effectively maintain intricate spatial data and make up for the absence of intricate spatial features with high precision. Unlike the original DASPP module, however, the denser DASPP uses modified dilation rates to better capture long-range and multi-scale lost information.

3) Channel and spatial attention mechanisms are incorporated into the decoder, in the form of newly designed Convolution and Channel Attention (CCA) and Convolution and Spatial Attention (CSA) blocks, as to enhance the network's focus on pixels and

fully leverage contextual information. In addition, this improvement allows to better integrate targets at different spatial locations and scales, thereby improving network accuracy.

II. RELATED WORK

A. DEEP LEARNING FOR MEDICAL IMAGE PROCESSING

In the past, medical imaging technologies such as CT [23], magnetic resonance imaging (MRI) [24], positron emission tomography (PET) [25], and others [26], [27], [28], have been extensively utilized for early detection and diagnosis of diseases. However, long-term continuous work of physicians is inevitable prone to fatal errors and fatigue. With the introduction of deep learning, the situation is continuously improving [29], as careful observation or learning can effectively describe fixed patterns of useful features, which play a crucial role in various tasks of medical image analysis. Many techniques and methods were elaborated in the past, including image thresholding [30], active contouring [31], region growing [32], de-formable models [33], and machine learning-based approaches for segmenting pulmonary nodules and pulmonary CT [34], [35], [36]. With the invent of deep neural networks (DNNs) [37], automatic feature learning can be more effectively achieved.

In the field of deep learning, particularly in the areas of computer vision and medical imaging, convolutional neural networks (CNNs) [38] have achieved significant success. In 2014, the fully convolutional network (FCN) proposed by Long et al. [39] became the main framework for image segmentation. By replacing the fully connected layers in CNN models with convolutional layers, it consists entirely of convolutional and pooling layers. FCN has an encoder-decoder architecture, where the encoder is utilized for feature extraction, and the decoder is employed for up-sampling to restore the final segmentation output to the original resolution, achieving pixel-level classification. With the advancement of deep learning, researchers have made modifications to FCN. In 2018, Zhao et al. proposed a combination of FCN and Conditional Random Fields (CRFs) [40] for brain tumor segmentation, which effectively addressed the limitations of FCN. In 2015, Ronneberger et al. [17] proposed the U-Net network architecture as an improvement over FCN. The main enhancements of U-Net include the addition of skip connections and the employment of a symmetric encoder-decoder structure, which enables the network to better handle details and spatial information. Skip connections allow the network to fuse more low-level feature information in the decoder part, thereby improving segmentation accuracy. By utilizing more feature channels and employing symmetric convolution and pooling operations, U-Net is capable of recovering better resolution and fine details. Bai et al. [41] integrated spatial and temporal information into a network using U-Net, thereby improving feature extraction and contextual connections. Oktay et al. [42] proposed an Attention Gate (AG)

model specifically designed for medical images, where the network can implicitly learn salient features of the image during training. With minimal computational overhead, AGs can be easily integrated into U-Net, allowing to increase its sensitivity and prediction accuracy. The main improvement of the resultant model, called Attention U-Net, over U-Net lies in the introduction of attention mechanisms to enhance focus on important feature regions. In 2017, Badrinarayanan et al. [43] proposed SegNet, which provided a more advanced framework for deep learning. Building upon the FCN architecture, SegNet preserves multi-scale extracted features and contextual information, thus retaining more details during the image segmentation process. In 2018, Zhou et al. [44] proposed the U-Net++ segmentation network to address the limitations of fully convolutional semantic segmentation. Building upon the U-Net architecture, U-Net++ improved skip connections in the decoder part, as to enable the aggregation of multi-scale features for achieving greater flexibility in segmentation. Additionally, supervised learning was employed to search for the optimal depth, resulting in superior segmentation performance. In 2018, Alom et al. proposed a segmentation network, called R2U-Net [45], which improved the cyclic residual convolutional layer on the basis of U-Net. This ensured the feature representation during the segmentation process, while maintaining the same number of network parameters as U-Net.

In 2018, Poap et al. [46] employed a nature-inspired algorithm for segmenting chest X-ray images to detect lung diseases such as pneumonia, tumors, and emphysema. The objective was to automate the analysis of X-ray screenings for expediting the process of medical examinations and initiate treatment promptly. Traditional methods utilize image segmentation to extract regions of interest (RoIs) for further analysis of deviations from normality. However, in [46] a method is proposed that extends segmentation using only essential elements. The presented research results demonstrated the effectiveness of the heuristic approach in detecting abnormalities in aggregated X-ray images for segmentation. However, it should be noted that the proposed method is based on simulating biological behavior, which may not always be accurate in detecting potential degenerative tissues.

In 2021, Dong et al. [47] proposed a CT lung nodule segmentation method, based on the combination of ResNeXt and U-Net++ with SCSE attention modules. Their method demonstrated a powerful capability for extracting lung nodule features and achieved excellent performance in segmenting lung nodules. However, the authors in [47] do not compare their proposed method with other advanced lung nodule segmentation methods.

In 2021, Rehman et al. [48] proposed a framework for detecting 15 chest diseases, including COVID-19, using chest X-ray examination patterns. The proposed framework includes a CNN architecture, based on deep learning, with a SoftMax classifier. Transfer learning is employed in this framework to accelerate training, improve model

performance, address data imbalance issues, and provide better abstract feature representations by leveraging existing knowledge and feature representations. The utilization of transfer learning enhances the accuracy of COVID-19 detection and improves the predictability of other chest diseases. Additionally, batch normalization is used to enhance the model's robustness to noise and variations in input data. By normalizing information and speeding up the CNN training between convolutional layers and Re-LU layers, the proposed framework improves its robustness. Compared to other state-of-the-art (SOTA) frameworks used for diagnosing COVID-19 and other chest diseases, the proposed framework demonstrates improved accuracy in COVID-19 detection and increased predictability for other chest diseases, [48].

In 2021, Khan et al. [49] proposed a model, called VGG-SegNet, which utilizes the first five layers of the VGG19 model as an encoder. These layers are capable of extracting low-level features from images such as edges and textures. The decoder, implemented with a reversed VGG19 model, is used to up-sample the encoded features, allowing the reconstruction of segmentation images with the same size as the original image. This architecture effectively enhances the accuracy of image segmentation. The model combines handcrafted features with deep features, improving the accuracy of lung nodule detection. It can also be applied to CT images without removing artifacts. Therefore, this method contributes to the diagnosis and treatment of lung cancer and provides valuable insights for further advancements in the field of lung nodule detection.

In 2022, Zhang et al. [50] proposed the multi-scale segmentation squeeze-and-excitation (SE) U-Net with a conditional random field (M-SegSEUNet-CRF) for automated lung tumor segmentation from CT images. M-SegSEUNet-CRF employs a multi-scale strategy to address the issue of variable tumor sizes. By leveraging a spatial adaptive attention mechanism and incorporating segmentation SE blocks embedded within the 3D U-Net, the model highlights tumor regions effectively and surpasses other existing models such as U-Net and U-Net++, highlighting the potential of its advanced approach for lung tumor segmentation. It could inspire further research and advancements in the field, leading to the development of more accurate and efficient models for lung tumor analysis.

In 2022, Jaszcz et al. [51] proposed a model for lung X-ray image segmentation using the Heuristic Red Fox Optimization (HRFO) algorithm. Based on the Red Fox Optimization algorithm, HRFO introduces adaptive weights and random perturbation strategies to enhance the algorithm's diversity and global search capability. Traditional manual selection of segmentation threshold parameters requires a significant amount of labor and time, and is prone to subjective errors. By using the HRFO algorithm to automatically select segmentation threshold parameters, the model improves the accuracy and efficiency of lung X-ray image segmentation.

This model achieves fast computation and analysis by reducing the search space, thereby enhancing the efficiency of disease identification and classification. However, the performance of the model may be influenced by image quality and noise, which may impose certain limitations when processing low-quality images. The utilization of this model provides new ideas and methods for research in this field, thereby promoting its development.

In 2023, Jennifer et al. [52] proposed a model for automatic detection of lung infections using an ensemble approach. The model utilizes α-means and β-augmentation operations to reduce ensemble uncertainty and enhance image edges, thus improving accuracy. The enhanced images are then fed into transfer learning architectures such as ResNet-50, VGG-16, and XG-Boost for training and testing, to distinguish between different types of lung infections. Using pre-trained ResNet-50 and VGG-16 models reduces the training time and computational costs, while improving the model's performance and accuracy. Additionally, incorporating the XG-Boost model as a classifier enhances the model's classification accuracy and robustness. By combining the advantages of these models, a more accurate and robust model for diagnosing lung infections can be constructed, thereby improving the accuracy and efficiency of medical diagnosis. The introduction of this model has had a positive impact on the field, providing a new approach for detecting lung infections and assisting doctors in diagnosing patients more quickly and accurately.

The network proposed in the current paper is based on an improved U-Net architecture, utilizing an encoder-decoder framework, which allows it to effectively capture contextual information in images. The encoder gradually extracts high-level features of the images through successive downsampling operations, while the decoder restores the features to the original resolution through up-sampling operations, generating segmentation results containing rich contextual information. Skip connections play a crucial role in U-Net by connecting low-level features with high-level features, facilitating the decoder in leveraging detailed information from the encoder to enhance the accuracy of the segmentation results. Based on a literature review, it has been discovered that attention mechanisms (briefly presented in the next section) provide advantages in helping models focus on key information within the input data. Attention mechanisms dynamically adjust weights based on the importance of different parts of the input data, enabling the model to better recognize and utilize important features, thereby improving its performance. Furthermore, it has been found that residual structures can propagate gradients more effectively, facilitating network convergence and increasing the training speed. Therefore, attention mechanisms and residual structures are employed by the proposed network as described further in this paper.

B. ATTENTION MECHANISMS

The attention is a research field that investigates human visual perception. From the perspective of cognitive science, due

to the existence of an irreducible cognitive bottleneck in information processing, people tend to selectively focus on the most salient aspects of information while overlooking some relevant information. In recent years, attention mechanisms [53] have been widely employed in deep learning (DL) to enhance the network performance. Attention mechanisms can be classified into different types, mainly including soft attention, hard attention, and self-attention. With the incorporation of attention mechanisms, a DL model is able to selectively focus on certain information within the data in two aspects: (1) determining which parts of the input to pay attention to, and (2) allocating limited processing resources to important parts, thereby enhancing the model's performance. A soft attention mechanism can help a model to focus on key regions in the input data and extract task-relevant information. A hard attention mechanism can be used to accurately localize the RoIs, such as nodules. A self-attention mechanism can aid a model in capturing long-range dependencies within input sequences.

SENet [54] is a representative work in computer vision that applies attention mechanism to channel dimension. Its structure is remarkably simple; yet it yields significant performance improvement. From the perspective of contextual modeling, GENet [55] makes full use of the mechanism of spatial attention to better integrate contextual relationships. RANet [56] improves network performance by focusing attention on RoIs or salient regions. DANet [44] leverages spatial pixels and channel features as query statements for context modeling, adaptively integrating local features and global dependencies.

In this paper, newly designed CCA and CSA attention blocks, and the recently proposed SimAM attention module [22], are incorporated into U-Net, resulting in the proposed ResDSda_U-Net network described in the following section. The combination of multiple attention mechanisms enhances the performance of the network and allows for more precise capturing of contextual information, which in turn enables achieving better accuracy, stability, and interpretability in the segmentation of pulmonary nodules.

III. PROPOSED NETWORK–ResDSda_U-NET

Figure 1 illustrates the overall architecture of the proposed ResDSda_U-Net network, which is an improved version of the U-Net consisting of an encoder and a decoder. Upon receiving an input image, the encoder extracts image features to obtain important information, which forms a compact feature input. Once entered into the encoder, the features are used to generate pixel-level segmentation masks. In addition to these two pathways, the original U-Net architecture also includes skip connections, which connect the output layer of each encoding layer to the corresponding input of the decoder. Because the encoder may lose some important information while extracting features, this information can be integrated through skip connections to better facilitate the generation of segmentation masks.
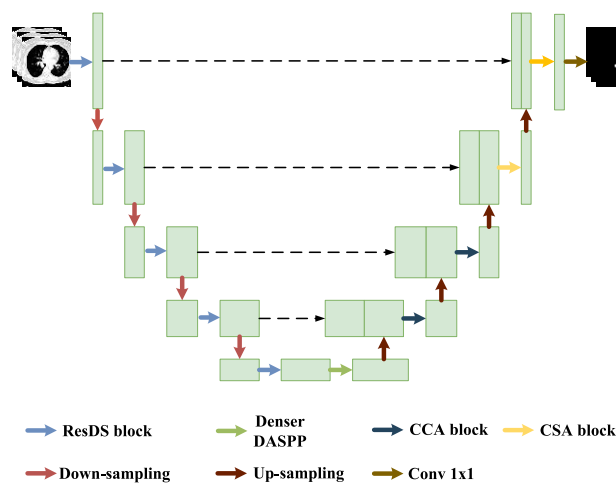

FIGURE 1. The architecture of the proposed ResDSda_U-Net network.

A. OVERALL ARCHITECTURE

Similarly to FCN and SegNet, U-Net employs convolutional blocks for semantic segmentation. Its network architecture is symmetric, with the encoder used for feature extraction and decomposing the image into a combination of feature maps at different levels to obtain contextual information. The convolutional block consists of two sequences of 3×3 convolutional operations, and utilizing this block effectively doubles the number of feature channels and enhances information extraction.

The proposed ResDSda_U-Net network adds newly designed ResDS blocks to the encoder part of the original U-Net architecture. The encoder is used for feature encoding of the original image, acquiring multi-scale contextual features. After a series of convolutions, a max-pooling operation [57] with a pool size of 2×2 and stride of 2 is performed to halve the size of the feature maps. These convolutions and pooling are repeated four times in total. Subsequently, an additional convolutional block is used to connect the encoder and decoder. In addition, a denser DASPP module is added after this convolutional block to enable the network to capture larger receptive fields more easily, thereby enhancing its ability to process multiscale pathological regions. The decoder utilizes the features extracted by the encoder to construct segmentation maps. Following the feature extraction from the encoder, up-sampling is immediately performed to restore the feature maps. Additionally, after each up-sampling operation, the feature map is merged with the corresponding feature maps output by the encoder layers through skip connections to better integrate shallow and deep features. The concatenated feature maps are then propagated to convolutional layers to undergo a series of convolutions for further processing. Similarly to the encoding, the decoding process is also repeated four times in the proposed ResDSda_U-Net network, utilizing both CSA blocks and CCA blocks. Finally, a 1×1 convolution is performed to obtain the final segmented image. Except for this last convolutional layer which uses
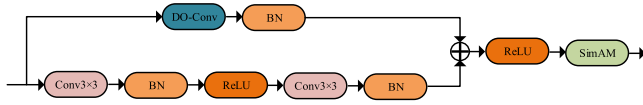
FIGURE 2. The architecture of the proposed ResDS block.

a *Sigmoid* activation function [58], a Rectified Linear Unit (ReLU) activation function is used in all other convolutional layers.

B. ResDS BLOCKS

Efficient feature extraction is a prerequisite for medical image segmentation. As a fundamental part of the pulmonary nodule segmentation processing, the convolutional layers must possess strong feature extraction capabilities. The encoder part of the proposed ResDSda_U-Net network employes newly designed ResDS blocks (Figure 2), based on the original U-Net encoder's residual structure (Res) fused with a DO-Conv layer and a SimAM attention module (Figure 3). The DO-Conv layer was proposed in [59] to enhance the feature extraction capability of convolutional layers by incorporating additional depth convolutions. In addition, to prevent an increase in computational complexity, a SimAM attention module [22] is used at the end of the ResDS block in order to infer attention weights and evaluate the importance of neurons after the convolutional layers.

The SimAM attention module can be viewed as a computational unit that enhances the feature extraction capability of CNN without increasing the computational cost. It can take any intermediate feature tensor as an input and can output a feature map of the same size through various transformations. In order to effectively utilize SimAM, it is imperative to evaluate the significance of each individual neuron. The process of information-rich neurons interacting with their surrounding neurons exhibits significant differences. Neurons that display spatial inhibitory effects are expected to be more critical. The specific methods used are the following, [72]:

1) The feature space mean d is obtained based on the input feature mapping X, as follows:

$$d = (X - X.mean(dim = [2, 3]))^2, \qquad (1)$$

where $X.mean$ denotes the mean operation performed on the variables in tensor X and $dim = [2, 3]$ denotes the operation performed along the second and third dimensions of tensor X.

2) The variance of the feature map width W and height H in the channel direction is calculated based on the feature space mean d, as follows:

$$v = \frac{d.sum(dim = [2, 3])}{H \times W - 1}. \qquad (2)$$

3) The energy distribution of the obtained feature map is given by the energy factor E, calculated as follows:
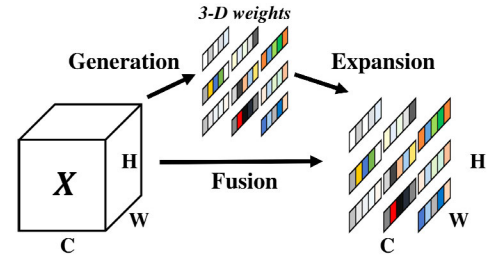
$$E = \frac{d}{4(v + r)} + 0.5. \qquad (3)$$

FIGURE 3. The architecture of the SimAM attention module.

4) The final result is an enhanced feature map, as shown below:

$$X^* = sigmoid\left(\frac{1}{E}\right) \odot X. \qquad (4)$$

Ultimately, the encoder of the proposed ResDSda_U-Net network consists of five ResDS blocks and four pooling layers with a stride of 2.

C. DENSER DASPP MODULE

The denser DASPP module, proposed here, is added in between the encoder and decoder of ResDSda_U-Net. It is a slightly modified version of the DASPP module [19] which, as a combination of DenseNet [60] and an ASPP module [61], can effectively extract objects of different scales and more effectively segment high-resolution lung images with complex textures. DASPP captures multi-scale contextual information by applying dilated convolutions with different rates. The one-dimensional dilated convolution is performed as per [19], as follows:

$$\Upsilon(i) = \sum_{n=1}^{N}(X[i + r \times n] \times W(n)), \qquad (5)$$

where X denotes the input feature map, W denotes the convolution kernel, N denotes the size of the filter, and r denotes the dilation rate of the convolution block. By incorporating dilation into different convolutions, while also preserving the resolution of the feature map, the network is able to improve its segmentation capabilities by having a larger receptive field while still maintaining high resolution.

DASPP consists of three layers of dilated convolutions, 1×1 convolutional blocks, and a global average pooling layer. The cascaded computation of the five feature extraction operations is given in [19] as:

$$\Upsilon = Concat\left(I_{pool}(X), C(X), \Upsilon_0, \Upsilon_1, \Upsilon_2\right), \qquad (6)$$

where Concat(·) denotes the concatenation of the channel dimension of five output feature maps, $I_{pool}(\cdot)$ denotes the average pooling operation, and $C(\cdot)$ denotes the convolution operation with a 1×1 kernel.

In the proposed ResDSda_U-Net network, a newly designed, denser DASPP module is used, which utilizes modified dilated rates. The dilation rate refers to the distance between pixels in the convolutional kernel during the convolutional operation. Different dilation rates are used at different
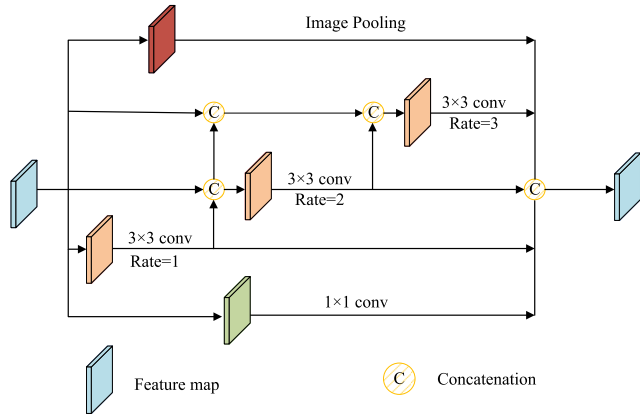
FIGURE 4. The architecture of the denser DASPP module.

convolutional layers to expand the size of the receptive field. By using different dilation rates, different receptive field sizes can be achieved in different convolutional layers. Smaller dilation rates allow to capture detailed information, while larger dilation rates can capture a broader range of contextual information. Due to the small size of pulmonary nodules, the use of smaller dilation rates enables better detection of their size. More specifically, in the denser DASPP module, the original DASPP dilated rates of 6, 12, and 18, used at the corresponding layers of dilated convolutions, are respectively replaced with 1, 2, and 3, as to better capture the image location information and improve the segmentation accuracy of the network. The values (1, 2, 3) of the dilated rates were experimentally obtained as the optimal ones, as shown further in Subsection V-B. Thus, in the denser DASPP module shown in Figure 4, used by the proposed ResDSda_U-Net network, the output Υ is obtained (as an expression for each layer of dilated convolution) in slightly different from [19] manner, as follows:

$$
\begin{cases}
\Upsilon_0 = H_{1,3}(X) \\
\Upsilon_1 = H_{2,3}\left(Concat\left(X, \Upsilon_0\right)\right) \\
\Upsilon_2 = H_{3,3}\left(Concat\left(X, \Upsilon_0, \Upsilon_1\right)\right),
\end{cases}
\tag{7}
$$

where $H_{r,n}(\cdot)$ denotes a dilated convolution with dilated rate r and convolution kernel size n.

D. CCA AND CSA BLOCKS
In the field of medical image segmentation, the use of images for segmentation is limited by their local receptive fields due to the relatively uniformity in size, shape, and texture of medical images. This limitation results in the emergence of local features, which diminishes the ability to capture contextual relationships and leads to potential differences between features corresponding to pixels with the same label. Ultimately, this can have a detrimental effect on the performance of the network. Therefore, we propose to use an attention mechanism in the decoder to capture the correlations between features. Specifically, a spatial attention module (SAM) and a channel attention module (CAM) [62] are utilized to build
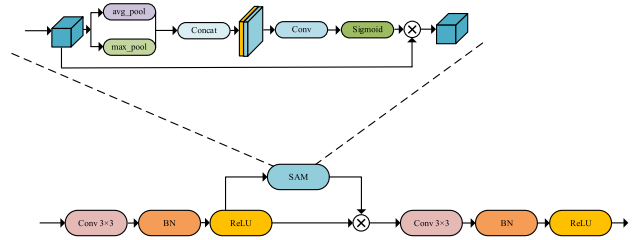

FIGURE 5. The architecture of the proposed CSA block.

two new types of attention blocks (a CCA block and a CSA block) for use in the proposed ResDSda_U-Net network as to improve its performance.

The characteristics of the network become increasingly rich with its depth. However, due to the cascading convolutional and down-sampling operations, important in-formation may be lost during the network's operation. To mitigate this issue, incorporating these two newly designed attention blocks in the decoder can significantly reduce information loss during the forward propagation process.

In addition, when dealing with scenarios where adjacent boundaries and ambiguous contours are present, the concatenation of high-dimensional features after up-sampling may lead to confusion of features at different scales. This can be prevented by the attention mechanism utilized in the proposed ResDSda_U-Net network in the form of the CCA and CSA blocks, described in the next subsections. Due to the relatively high resolution of shallow feature maps, the channel feature distribution has a greater impact on feature fusion, [72]. Therefore, two CCA blocks are utilized in the first two up-sampling convolutional blocks to better fuse features, followed by two CSA blocks used in the last two up-sampling convolutional blocks.

1) CSA BLOCK
The structure of the proposed CSA block is shown in Figure 5. Introducing it as a part of the decoder for image segmentation allows it to utilize spatial relationships between features and generate a spatial attention map. In order to accurately calculate the spatial attention, SAM [62] first performs max pooling and average pooling operations on the input feature map and connects them to generate an efficient feature descriptor (Figure 5). The results of maximum and average pooling are concatenated along the channel dimension to obtain a feature map with a dimension of H × W × 2. Subsequently, a convolution operation is performed on the concatenated results to obtain a feature map with a dimension of H × W × 1, which is then processed by an activation function. The computation of the output features of SAM is performed as follows:

$$
\begin{aligned}
M_s(F) &= \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right) \\
&= \sigma\left(f^{7\times7}\left(\left[F_{avg}^s; F_{max}^s\right]\right)\right)
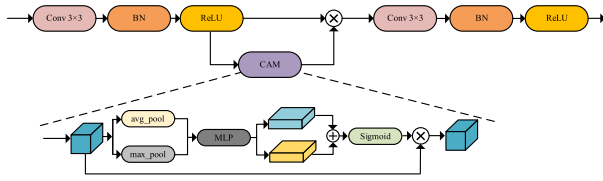\end{aligned}
\tag{8}
$$

FIGURE 6. The architecture of the proposed CCA block.

2) CCA BLOCK

The structure of the CCA block is shown in Figure 6. It utilizes CAM [62] whose input is a feature map with dimensions of H × W × C, where H and W are the height and width of the feature map, respectively, and C is the number of channels. CAM first performs global pooling and average pooling operations on the feature map in the spatial dimension, thereby reducing the spatial size. After performing pooling operations, the feature map is fed into a multilayer perceptron (MLP) for learning. Subsequently, the outputs of MLP are summed and mapped through a *Sigmoid* function to obtain the final attention values. The calculation of channel-wise attention output features is shown below:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$
$$= \sigma\left(W_1\left(W_0\left(F_{avg}^c\right)\right) + W_1\left(W_0\left(F_{max}^c\right)\right)\right) \quad (9)$$

IV. EXPERIMENTS

A. DATASET

The experimental dataset utilized in this study was sourced from the publicly available Lung Image Database Consortium / Image Database Resource Initiative (LIDC/IDRI) dataset [63], which provides the largest public resource to assess the performance of models for the detection of pulmonary nodules, [64]. The LIDC/IDRI dataset is a collection of pulmonary nodule images formed through the efforts of the National Institutes of Health in the United States, comprising 1018 helical thoracic CT scans collected from seven academic centers. Each case in the dataset was annotated by four radiologists for the contour and other important information of pulmonary nodules, and the annotations were saved in XML files. Four independent teams of medical experts evaluated and interpreted the lung nodules, discovered in the images, and provided detailed characteristics and annotation information about them. The process included a series of meetings for collaborative discussions and consistent validation. Therefore, the creation of the LIDC/IDRI dataset involved collaboration and clinical validation from medical experts whose knowledge and experience were crucial for annotating nodules, interpreting images, and providing information about nodule characteristics. Overall, the LIDC/IDRI dataset not only provides clinical validation but also enhances the reliability and credibility of research results by collaborating with medical experts, thus enabling researchers to interpret findings effectively.
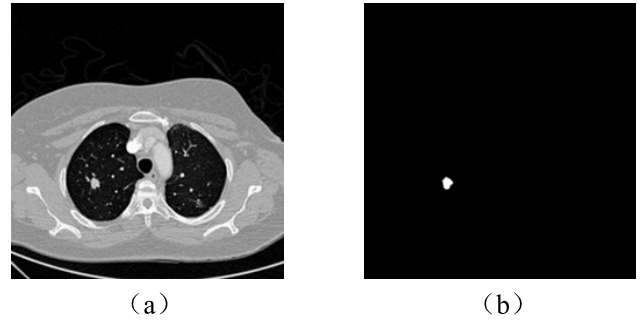

（a）   （b）

FIGURE 7. Samples of LIDC/IDRI dataset: (a) Original CT image; (b) Corresponding ground-truth mask.

In the conducted experiments on this dataset, we selected pulmonary nodules with a diameter greater than 3 mm and at least three expert annotations, while excluding nodules with inconsistent slice thickness or missing slices. Ultimately, we identified 7,603 usable case images through this screening process. A sample LIDC/IDRI image and corresponding ground-truth mask are shown in Figure 7.

B. DATA PREPROCESSING

The reflection of tissue absorption of X-rays in medical images is measured in Hounsfield Units (HU). As the concept of grayscale values pertains to computer science, prior to conducting the experiments, it was necessary to convert medical images to HU values. In the field of medical imaging, the windowing technique is used to alter the window width and window level, thereby selecting the CT value range of interest. Given that CT values for different tissue structures or pathological sites in medical CT images vary, it is necessary to choose an appropriate window width and window level when studying a particular site. To this end, we selected the optimal window level and width values for the segmentation of pulmonary nodules on the LIDC/IDRI dataset. Specifically, we used a window level of −300 and a window width of 600. Finally, we normalized the data range of the pulmonary images in the interval of 0 to 1.

Due to the specific structure of pulmonary CT images, we utilized a center enlargement approach for image enhancement without altering the image shape, which allows to significantly increase the network's generalization ability and reduce the overfitting in the network.

The steps of the data pre-processing process are illustrated in Figure 8.

C. PARAMETER SETTING

In the experiments, we used a total of 7603 CT images, which were randomly divided into training, validation, and test sets at a ratio of 8:1:1. During the training process, we set the number of training epochs to 250, the initial learning rate to 0.001, and the batch size to 16. We utilized a random seed number of 1234, and employed the Stochastic Gradient
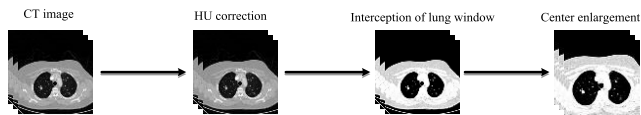
FIGURE 8. The data pre-processing process.

Descent (SGD) optimization algorithm to train the network, as shown in Table 1.

D. EVALUATION METRICS

Four evaluation metrics were used to measure the networks' segmentation performance, namely the Dice similarity coefficient (DSC), intersection over union (IoU), sensitivity (Sen), and precision (Pre). DSC is used to compute the similarity between two sets/samples. IoU is a measure of the degree of overlap between a target mask and a ground truth mask, commonly employed as a criterion to determine if a segmentation is a true positive (TP) sample. These two metrics, widely used in the field of medical image segmentation, are defined as follows:

$$IoU = \frac{TP}{TP + FP + FN}; \qquad (10)$$

$$DSC = \frac{2TP}{2TP + FP + FN}, \qquad (11)$$

where *TP*, *FP*, and *FN* respectively denote the accurate segmentation of pulmonary nodule lesions as pulmonary nodules, incorrect segmentation of background regions as pulmonary nodule lesions, and incorrect segmentation of pulmonary nodule lesions as background regions.

Precision is the measure of the proportion of correct positive predictions out of the total positive predictions made by a network. Sensitivity, on the other hand, refers to the degree of recognition exhibited by a network on the dataset. These two metrics are defined as follows:

$$Sen = \frac{TP}{TP + FN}; \qquad (12)$$

$$Pre = \frac{TP}{TP + FP}. \qquad (13)$$

E. COMBINED LOSS FUNCTION

In the task of pulmonary nodule segmentation, due to the fact that pixels can only be labelled with a probability of being or not being part of a nodule, this can be treated as a binary classification problem. Binary cross-entropy (BCE) loss is commonly used for binary segmentation tasks. It is a loss function that measures the difference between the predicted probability distribution and the actual labels. The main advantage of the BCE loss function is that it provides a smooth loss curve, which helps to train models faster. It is defined as follows:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} \left(s_i ln\left(g_i \right) + \left(1 - s_i \right) ln\left(1 - g_i \right) \right), \quad (14)$$

where g_i denotes the actual pixel value and s_i denotes the pixel value produced by a network.

TABLE 1. Experimental environment's parameters.

Component	Name/Value
Operating system	Linux 3.10.0-1062.el7.bclinux.x86_64
CPU	Intel(R) Xeon(R) Gold 6240 CPU @ 2.60GHz
GPU	NVIDIA Tesla V100S PCIe 32GB
Input image size	256×256
Batch size	16
Epoch	250
Optimizer	SGD
Initial learning rate	0.001
Random seed	1234

However, the BCE loss function is not suitable for standalone use in pulmonary nodule segmentation, as non-nodule pixels in pulmonary nodule images are far more abundant than nodule pixels, which may cause the network to have a bias towards segmenting background pixels. Thus, in the conducted experiments a combined approach is utilized, which combines the BCE loss with the Dice loss to form a BCE+Dice loss [73]. The advantage of using the Dice loss function is related to its ability to effectively handle class imbalance in terms of foreground and background pixel counts. It prompts a network to generate images that closely resemble the mask. Due to its inherent maximization of DSC and its beneficial impact on class imbalance, it is gradually becoming a popular choice. The Dice loss function is defined as follows:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^{N} s_i g_i}{\sum_{i=1}^{N} s_i^2 + \sum_{i=1}^{N} g_i^2}. \qquad (15)$$

The combined BCE+Dice loss function enables generating better predictions by a network, thus performing better pulmonary nodule segmentation. It is defined as follows:

$$L = \frac{1}{2} L_{BCE} + L_{Dice}. \qquad (16)$$

V. EXPERIMENTAL RESULTS AND ANALYSIS
A. SEGMENTATION PERFORMANCE COMPARISON OF NEURAL NETWORKS

First, the segmentation performance of the proposed ResDSda_U-Net network was compared to that of existing open-source U-Net based networks (U-Net++ [44], U-Net [17], AttentionU-Net [42], R2U-Net [45]) and Seg-Net [43], all used for medical image segmentation, based on experiments conducted on the LIDC/IDRI dataset. In the experiments, we applied the same training strategy to ResDSda_U-Net and to the open-source networks to ensure fair comparison. Table 2 presents the obtained results (the best value achieved among the networks for a particular metric is shown in **bold**).

As shown in Table 2, the proposed ResDSda_U-Net network outperforms all other network according to all evaluation metrics used. This demonstrates that ResDSda_U-Net

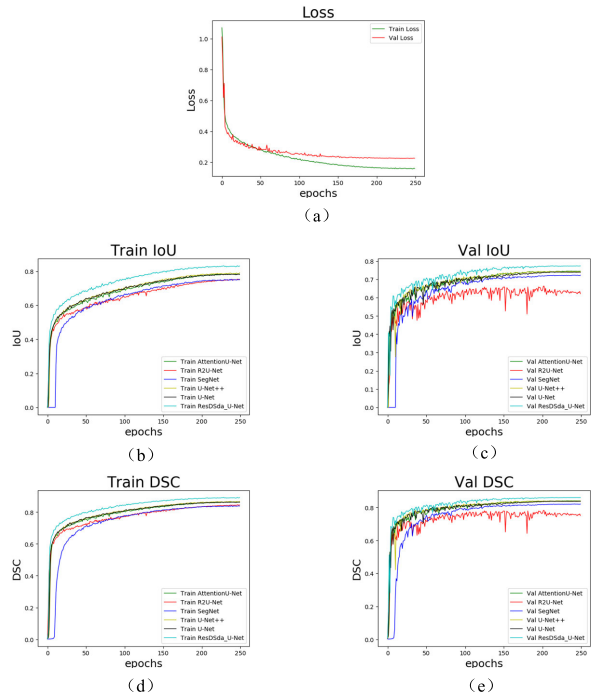**TABLE 2.** Segmentation performance comparison results with U-Net based networks and SegNet.

Network	DSC	IoU	Sen	Pre
SegNet [43]	82.89%	71.26%	80.08%	86.56%
U-Net++ [44]	83.72%	72.40%	80.97%	87.17%
U-Net [17]	83.84%	72.54%	81.19%	87.13%
AttentionU-Net [42]	84.46%	73.50%	82.52%	87.00%
R2U-Net [45]	77.95%	64.35%	82.67%	74.29%
ResDSda_U-Net (*proposed*)	**86.65%**	**76.73%**	**86.30%**	**87.22%**
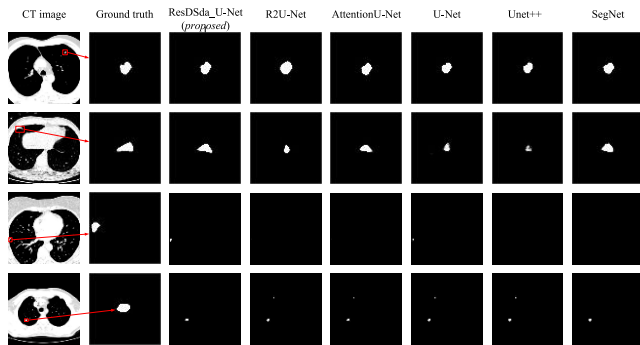
FIGURE 9. An illustration of pulmonary nodule segmentation, performed by U-Net based networks and SegNet.

FIGURE 10. Training and validation process of ResDSda_U-Net and other existing U-Net-based open-source networks and SegNet: (a) training and validation loss curves of ResDSda_U-Net; (b) IoU training curves of considered networks; (c) IoU validation curves of considered networks; (d) DSC training curves of considered networks; (e) DSC validation curves of considered networks.

is capable of efficiently enhancing the ability to detect pulmonary nodule lesions compared not only to the baseline network (U-Net), but also to other U-Net-based networks and SegNet. More specifically, ResDSda_U-Net is ahead of the second-best performing network respectively by 2.19 points according to DSC, 3.23 points according to IoU, 3.63 points according to sensitivity, and 0.05 points according to precision. As illustrated in Figure 9, the proposed ResDSda_U-Net network can effectively differentiate the scope of pulmonary nodule lesions from the background region. Compared to other networks, used in the comparison, ResDSda_U-Net can better address the segmentation problem and produce masks with clearer lesion boundaries, making it more sensitive and effective in segmenting pulmonary nodules. This also indicates that the proposed network exhibits stronger robustness and effectiveness in performing the segmentation task. Additionally, to provide a clearer visualization of the convergence of the proposed network in comparison to other existing U-Net-based open-source networks and SegNet during the training process, Figure 10 illustrates the loss variation curves of ResDSda_U-Net on both the training and validation sets, as well as the DSC and IoU training and validation curves of all considered networks.

Next, the segmentation performance of the proposed ResDSda_U-Net network was compared to that of existing SOTA networks used for medical image segmentation. In the conducted experiments, we applied the same training strategy to ResDSda_U-Net, JOSHUA [65], UTNet [66], FusionNet [67], and AUNet [68] to ensure fairness. The other

results, shown in Table 3 for the rest of the networks, are obtained from the corresponding literature sources (the values not found in these sources are marked with '-'). The best value achieved among the networks for a particular metric is shown in **bold**.

As can be seen from Table 3, the proposed ResDSda_U-Net network outperforms all SOTA networks according to DSC and IoU. More specifically, ResDSda_U-Net is ahead of the second-best performing network by 1.55 points according to DSC and 2.6 points according to IoU, respectively. According to sensitivity, ResDSda_U-Net takes third place by scoring 3.7 points less than the leader (the network proposed in [69]), and according to precision it takes fourth place by scoring 0.76 points less than the leader (AUNet [68]).

Figure 11 shows examples of pulmonary nodule segmentation, where ResDSda_U-Net outperforms all SOTA networks used in the experiments.

Additionally, to provide a clearer visualization of the convergence of the proposed network in comparison to the SOTA networks during the training process, Figure 12 illustrates their DSC and IoU training and validation curves.

B. COMPUTATIONAL RESOURCES COMPARISON OF NEURAL NETWORKS

Table 4 presents a comparison of computational resources employed by ResDSda_U-Net and other SOTA networks. Despite the fact that the proposed network uses the largest

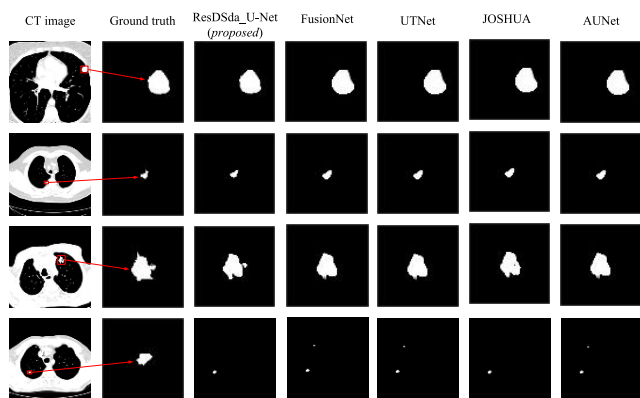**TABLE 3.** Segmentation performance comparison results with SOTA networks.

Network	DSC	IoU	Sen	Pre
JOSHUA [65]	83.45%	71.96%	81.79%	85.59%
UTNet [66]	84.02%	72.91%	81.40%	87.24%
FusionNet [67]	84.36%	73.38%	82.13%	87.32%
AUNet [68]	84.85%	74.13%	82.42%	**87.98%**
Liu *et al*. [70]	82.00%	69.00%	87.00%	-
Dong *et al*. [47]	83.43%	72.51%	-	-
Roy *et al*. [69]	84.00%	72.00%	**90.00**%	-
Qin *et al*. [71]	85.00%	73.00%	85.00%	-
Zhang *et al*. [50]	85.10%	-	82.70%	-
ResDSda_U-Net (*proposed*)	**86.65%**	**76.73%**	86.30%	87.22%

FIGURE 11. An illustration of pulmonary nodule segmentation, performed by SOTA networks.

FIGURE 12. Training and validation process of ResDSda_U-Net and existing SOTA networks: (a) IoU training curves; (b) DSC training curves; (c) IoU validation curves; (d) DSC validation curves.

number of parameters among the networks compared, it ranks second in terms of computational complexity (measured in FLOPs) after the leading network (i.e., JOSHUA for both indicators). However, ResDSda_U-Net exhibits outstanding

TABLE 4. Computational resources comparison results with SOTA networks.

Network	Number of parameters (million)	FLOPs (million)
JOSHUA [65]	**17.45**	**40750.22**
UTNet [66]	57.45	81202.18
AUNet [68]	75.50	106797.97
FusionNet [67]	78.53	134212.81
ResDSda_U-Net (*proposed*)	89.13	67402.52

TABLE 5. Results obtained by ResDSda_U-Net with different rates of dilation convolution.

Rate	DSC	IoU	Sen	Pre
（2,4,6）	86.54%	76.53%	85.83%	**87.46%**
（3,6,9）	86.48%	76.47%	85.75%	87.42%
（4,8,12）	85.97%	75.71%	85.11%	87.13%
（6,12,18）	86.37%	76.38%	85.88%	87.13%
（1,2,3）	**86.65%**	**76.73%**	**86.30%**	87.22%

segmentation performance, surpassing all networks presented in Table 4 (except for AUNet based on precision).
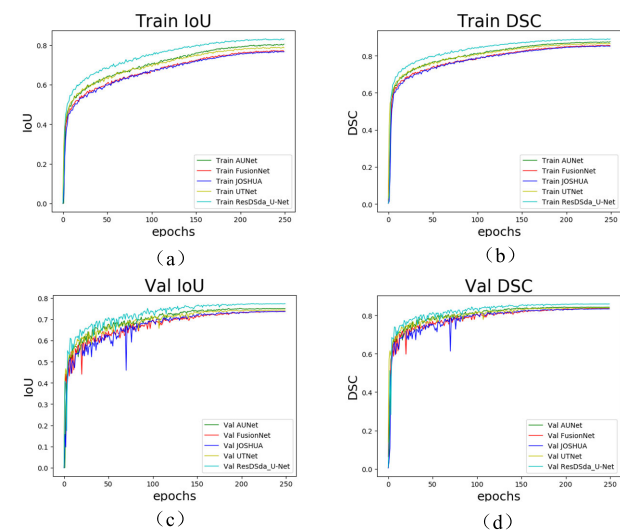
C. STUDY OF DILATED CONVOLUTION RATES

In a third set of experiments, we aimed at finding the optimal values of the dilated convolution rates allowing to increase the receptive field as much as possible while maintaining the feature map size. Due to the varying sizes of lesions and complex textures in pulmonary nodule segmentation, dilated convolution uses filters with holes to increase the receptive field while keeping the feature map size unchanged. As shown in Table 5, through multiple trials, these experiments have shown that revising the dilated rates of the original DASPP module to (1,2,3) values allows achieving the best segmentation performance, according to three out of the four evaluation metrics used, which leads to stronger network robustness and optimal receptive field effect.

D. STUDY OF OPTIMIZERS

In a fourth set of experiments, we studied the use of different optimizers for the proposed ResDSda_U-Net network. As shown in Table 6, under the same training strategy, ResDSda_U-Net achieved the best performance with the SGD optimizer, according to all evaluation metrics.

E. ABLATION STUDY

In order to validate the performance improvement that can be achieved due to the incorporation of various modules in the baseline network (U-Net), we conducted an ablation study on the LIDC/IDRI dataset using the newly designed ResDS, CCA, and CSA blocks, and a denser DASPP module. In the conducted experiments, these blocks/modules were gradually added to the original U-Net network, as shown in Table 7.

TABLE 6. Results obtained by ResDSda_U-Net with different optimizers.

Optimizer	DSC	IoU	Sen	Pre
Adam	75.28%	61.20%	76.47%	74.91%
AdamW	80.72%	68.21%	81.13%	80.88%
ASGD	83.88%	72.66%	82.12%	86.10%
SGD	**86.65%**	**76.73%**	**86.30%**	**87.22%**

TABLE 7. Ablation study results.

Network	DSC	IoU	Sen	Pre
U-Net	83.84%	72.54%	81.19%	87.13%
U-Net + ResDS	84.45%	73.45%	81.99%	87.46%
U-Net + CCA + CSA	85.40%	74.87%	85.21%	85.91%
U-Net + Denser DASPP	85.49%	75.05%	83.08%	**88.46%**
U-Net + ResDS + CCA + CSA + Denser DASPP (i.e., ResDSda_U-Net)	**86.65%**	**76.73%**	**86.30%**	87.22%

During the experiments, parameters such as training epochs and learning rate were kept constant.

First, by incorporating ResDS blocks into the original U-Net network, its segmentation performance was improved by 0.61, 0.91, 0.80, and 0.33 points according to DSC, IoU, sensitivity, and precision, respectively. The reason for this improvement is that the ResDS blocks can emphasize important features and suppress irrelevant features. Then, by incorporating CCA and CSA blocks into U-Net, the network performance was further improved by 1.56, 2.33, and 4.02 points, according to DSC, IoU, and sensitivity, respectively, even though according to precision the network performance dropped by 1.22 points. The reason for this improvement is the addition of an attention mechanism, which allows the network to focus on specific information of key data, thereby enhancing its segmentation performance. The incorporation of a denser DASPP module into U-Net resulted in an increase of all evaluation metrics by 1.65, 2.51, 1.89, and 1.33 points, respectively for DSC, IoU, sensitivity, and precision. The reason for this performance improvement is that DASPP can help the model better capture contextual information, thereby enhancing the accuracy of multi-scale pulmonary nodule segmentation. Finally, the simultaneous incorporation of all these types of blocks/modules into U-Net, resulting in the proposed ResDSda_U-Net network, allowed to achieve indeed the best performance results, based on three out of four evaluation metrics used, namely DSC, IoU, and sensitivity.

VI. CONCLUSION

Segmentation of pulmonary nodules is an important task performed on lung CT images, which aims to delineate nodules from images and assist physicians in the diagnosis and treatment of pulmonary nodules. However, due to the presence of complex anatomical structures, such as pulmonary vessels, and the fuzzy boundaries between different tissues, effective segmentation of pulmonary nodules in CT images is a challenging problem. Deep learning-based lung nodule

segmentation plays a significant role in solving this problem, as it can quickly and accurately identify pathological areas.

This paper has proposed a novel residual network for segmentation of pulmonary nodules in lung CT images, named ResDSda_U-Net, based on improvements of the U-Net network. More specifically, newly designed ResDS, CCA, and CSA blocks, and a denser DASPP module have been proposed for incorporation into U-Net to improve the network's utilization of feature information. Firstly, by incorporating residual connections into the network, in the form of a newly designed ResDS block, the issues of gradient vanishing and exploding in deep networks was addressed. Secondly, the introduction of a denser DASPP module allows to achieve better semantic information acquisition and higher resolution. Thirdly, incorporating channel and spatial attention mechanisms into the network, in the form of the newly designed CCA and CSA blocks, enables enhancing network efficiency and interpretability. Overall, the incorporation of ResDS, CCA, and CSA blocks, and the denser DASPP module into U-Net positively influences its decision-making process by improving its feature representation capability and obtaining multi-scale contextual information. The utilization of these newly designed blocks and modules allows to enhance the U-Net performance in semantic segmentation tasks, enabling it to better capture the semantic information of images, and improve the quality and accuracy of segmentation results.

Experiments, conducted on the LIDC-IDRI public dataset, demonstrated the total superiority of the proposed ResDSda_U-Net network compared to other existing U-Net based networks. Moreover, another set of experiments confirmed that ResDSda_U-Net outperforms also all SOTA networks considered, according to the DSC and IoU metrics. Results of additional studies, focused on the dilated convolution rates, optimizers, and ablation study, have also been presented. In conclusion, the proposed network can achieve superior performance in pulmonary nodule segmentation, and thus may assist doctors in diagnosing pulmonary nodular lesions quickly and reliably.

There are limitations to the proposed network, though, as the detection of lesion contours, scale, and range, based on the CT images of the LIDC/IDRI public dataset, was not precise enough in some cases. To further address these limitations, in the future we will consider integrating multiple imaging modalities, such as CT, MRI, and PET, as to provide a more comprehensive set of information, thereby enhancing the network's performance in nodule segmentation. Additionally, the incorporation of other relevant data beyond image data, such as clinical data and medical records, can be contemplated as supplementary information to further improve the segmentation performance.

In the future, it is hoped that the proposed ResDSda_U-Net network will not be limited to pulmonary nodule segmentation alone but can be extended to common medical image segmentation. However, this will require some adjustment and adaptation of the network's architecture, e.g., modifying the network's depth, width, convolution kernel size, pooling

strategy, etc., as to meet the requirements of different medical imaging tasks. Moreover, when extending the application of the proposed network to other tasks, it is necessary to reassess its performance and make additional adjustments accordingly. Performance evaluation metrics may differ across different tasks and may require specific evaluation procedures. Therefore, one must select appropriate performance evaluation metrics based on the task requirements, ensure the use of consistent standards aligned with the task objectives during the evaluation process, and make necessary adjustments.

Additionally, it is desired to design lightweight blocks/modules for incorporation into the network to accelerate the training process and reduce the training time.

REFERENCES

[1] T. Franquet, "Imaging of pulmonary viral pneumonia," *Radiology*, vol. 260, no. 1, pp. 18–39, Jul. 2011.

[2] K. Stefanidis, E. Konstantelou, G. T. Yusuf, A. Oikonomou, K. Tavernaraki, D. Karakitsos, S. Loukides, and I. Vlahos, "Radiological, epidemiological and clinical patterns of pulmonary viral infections," *Eur. J. Radiol.*, vol. 136, Mar. 2021, Art. no. 109548.

[3] G. Wei, M. Qiu, K. Zhang, M. Li, D. Wei, Y. Li, P. Liu, H. Cao, M. Xing, and F. Yang, "A multi-feature image retrieval scheme for pulmonary nodule diagnosis," *Medicine*, vol. 99, no. 4, 2020, Art. no. e18724.

[4] Y. Takahashi, T. Onodera, K. Kobayashi, and T. Kurosaki, "Primary and secondary B-cell responses to pulmonary virus infection," *Infectious Disorders-Drug Targets*, vol. 12, no. 3, pp. 232–240, May 2012.

[5] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *Lancet*, vol. 395, no. 10223, pp. 470–473, Feb. 2020.

[6] D. P. Bahner, E. J. Adkins, R. Nagel, D. Way, H. A. Werman, and N. A. Royall, "Brightness mode quality ultrasound imaging examination technique (B-QUIET): Quantifying quality in ultrasound imaging," *J. Ultrasound Med.*, vol. 30, no. 12, pp. 1649–1655, Dec. 2011.

[7] S. L. Fernandes, U. J. Tanik, V. Rajinikanth, and K. A. Karthik, "A reliable framework for accurate brain image examination and treatment planning based on early diagnosis support for clinicians," *Neural Comput. Appl.*, vol. 32, no. 20, pp. 15897–15908, Oct. 2020.

[8] I. K. Ip, K. J. Mortele, L. M. Prevedello, and R. Khorasani, "Repeat abdominal imaging examinations in a tertiary care hospital," *Amer. J. Med.*, vol. 125, no. 2, pp. 155–161, Feb. 2012.

[9] A. McWilliams et al., "Probability of cancer in pulmonary nodules detected on first screening CT," *New England J. Med.*, vol. 369, no. 10, pp. 910–919, Sep. 2013.

[10] R. Y. Kim, J. L. Oke, L. C. Pickup, R. F. Munden, T. L. Dotson, C. R. Bellinger, A. Cohen, M. J. Simoff, P. P. Massion, C. Filippini, F. V. Gleeson, and A. Vachani, "Artificial intelligence tool for assessment of indeterminate pulmonary nodules detected with CT," *Radiology*, vol. 304, no. 3, pp. 683–691, Sep. 2022.

[11] A. Gudur, H. Sivaraman, and V. Vimal, "Deep learning-based detection of lung nodules in CT scans for cancer screening," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 7s, pp. 20–28, 2023.

[12] H. Macmahon, R. Engelmann, F. M. Behlen, K. R. Hoffmann, T. Ishida, C. Roe, C. E. Metz, and K. Doi, "Computer-aided diagnosis of pulmonary nodules: Results of a large-scale observer test," *Radiology*, vol. 213, no. 3, pp. 723–726, Dec. 1999.

[13] Y. Ni, Z. Xie, D. Zheng, Y. Yang, and W. Wang, "Two-stage multitask U-Net construction for pulmonary nodule segmentation and malignancy risk prediction," *Quant. Imag. Med. Surg.*, vol. 12, no. 1, pp. 292–309, Jan. 2022.

[14] B. C. Lassen, C. Jacobs, J.-M. Kuhnigk, B. van Ginneken, and E. M. van Rikxoort, "Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans," *Phys. Med. Biol.*, vol. 60, no. 3, pp. 1307–1323, Feb. 2015.

[15] E. M. van Rikxoort, B. de Hoop, S. van de Vorst, M. Prokop, and B. van Ginneken, "Automatic segmentation of pulmonary segments from volumetric chest CT scans," *IEEE Trans. Med. Imag.*, vol. 28, no. 4, pp. 621–630, Apr. 2009.

[16] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100379.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Munich, Germany. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[18] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *Proc. 4th Int. Conf. Adv. Comput. Commun. Technol.*, Rohtak, India, Feb. 2014, pp. 5–12.

[19] S. Yin, H. Deng, Z. Xu, Q. Zhu, and J. Cheng, "SD-UNet: A novel segmentation framework for CT images of lung infections," *Electronics*, vol. 11, no. 1, p. 130, Jan. 2022.

[20] C. N. D. Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proc. 25th Int. Conf. Comput. Linguistics, Tech. Papers (COLING)*, Dublin, Ireland, Aug. 2014, pp. 69–78.

[21] Y. K. Tandon, B. J. Bartholmai, and C. W. Koo, "Putting artificial intelligence (AI) on the spot: Machine learning evaluation of pulmonary nodules," *J. Thoracic Disease*, vol. 12, no. 11, pp. 6954–6965, Nov. 2020.

[22] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "Simam: A simple, parameter-free attention module for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jul. 2021, pp. 11863–11874.

[23] G. D. Rubin, "Computed tomography: Revolutionizing the practice of medicine for 40 years," *Radiology*, vol. 273, no. 2S, pp. S45–S74, Nov. 2014.

[24] V. P. B. Grover, J. M. Tognarelli, M. M. E. Crossey, I. J. Cox, S. D. Taylor-Robinson, and M. J. W. McPhail, "Magnetic resonance imaging: Principles and techniques: Lessons for clinicians," *J. Clin. Exp. Hepatol.*, vol. 5, no. 3, pp. 246–255, Sep. 2015.

[25] T. Jones and D. Townsend, "History and future technical innovation in positron emission tomography," *J. Med. Imag.*, vol. 4, no. 1, Mar. 2017, Art. no. 011013.

[26] D. D. Patil and S. G. Deore, "Medical image segmentation: A review," *Int. J. Comput. Sci. Mobile Comput.*, vol. 2, no. 1, pp. 22–27, 2013.

[27] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, no. 1, pp. 315–337, Aug. 2000.

[28] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019.

[29] S. Masood, M. Sharif, A. Masood, M. Yasmin, and M. Raza, "A survey on medical image segmentation," *Current Med. Imag. Rev.*, vol. 11, no. 1, pp. 3–14, Apr. 2015.

[30] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun./Jul. 2016, pp. 2424–2433.

[31] B. Al-Diri, A. Hunter, and D. Steel, "An active contour model for segmenting and measuring retinal vessels," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1488–1497, Sep. 2009.

[32] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," *Proc. SPIE*, vol. 4322, pp. 1337–1346, Jul. 2001.

[33] J. Zhang, Y. Zhong, and C. Gu, "Deformable models for surgical simulation: A survey," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 143–164, 2018.

[34] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015.

[35] M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," *J. Intell. Learn. Syst. Appl.*, vol. 9, no. 1, pp. 1–16, 2017.

[36] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron. Mark.*, vol. 31, no. 3, pp. 685–695, 2021.

[37] J. Li, R. Zhao, J.-T. Huang, and Y. Gong, "Learning small-size DNN with output-distribution-based criteria," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, Singapore, Sep. 2014, pp. 14–18.

[38] D. Wei, B. Zhou, A. Torrabla, and W. Freeman, "Understanding intra-class knowledge inside CNN," 2015, *arXiv:1507.02379*.

[39] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

[40] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018.

[41] W. Bai, H. Suzuki, C. Qin, G. Tarroni, O. Oktay, P. M. Matthews, and D. Rueckert, "Recurrent neural networks for aortic image sequence segmentation with sparse annotations," in *Proc. 21st Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Granada, Spain. Cham, Switzerland: Springer, Sep. 2018, pp. 586–594.

[42] O. Oktay, J. Schlemper, L. Le 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," 2018, *arXiv:1804.03999*.

[43] 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.

[44] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. 4th Int. Workshop Deep Learn. Med. Image Anal. (DLMIA), 8th Int. Workshop Multimodal Learn. Clin. Decis. Support (ML-CDS)*, Granada, Spain. Cham, Switzerland: Springer, Sep. 2018, pp. 3–11.

[45] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.

[46] D. Poap, M. Wozniak, R. Damaševicius, and W. Wei, "Chest radiographs segmentation by the use of nature-inspired algorithm for lung disease detection," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 2298–2303.

[47] L. Dong and H. Liu, "Segmentation of pulmonary nodules based on improved UNet++," in *Proc. 14th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Shanghai, China, Oct. 2021, pp. 1–5.

[48] N.-U. Rehman, M. S. Zia, T. Meraj, H. T. Rauf, R. Damaševičius, A. M. El-Sherbeeny, and M. A. El-Meligy, "A self-activated CNN approach for multi-class chest-related COVID-19 detection," *Appl. Sci.*, vol. 11, no. 19, p. 9023, Sep. 2021.

[49] M. A. Khan, V. Rajinikanth, S. C. Satapathy, D. Taniar, J. R. Mohanty, U. Tariq, and R. Damaševičius, "VGG19 network assisted joint segmentation and classification of lung nodules in CT images," *Diagnostics*, vol. 11, no. 12, p. 2208, Nov. 2021.

[50] B. Zhang, S. Qi, Y. Wu, X. Pan, Y. Yao, W. Qian, and Y. Guan, "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images," *Comput. Methods Programs Biomed.*, vol. 222, Jul. 2022, Art. no. 106946.

[51] A. Jaszcz, D. Połap, and R. Damaševičius, "Lung X-ray image segmentation using heuristic red fox optimization algorithm," *Sci. Program.*, vol. 2022, pp. 1–8, Jul. 2022.

[52] J. S. Jennifer and T. S. Sharmila, "A neutrosophic set approach on chest X-rays for automatic lung infection detection," *Inf. Technol. Control*, vol. 52, no. 1, pp. 37–52, Mar. 2023.

[53] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021.

[54] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.

[55] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 9401–9411.

[56] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6450–6458.

[57] B. Graham, "Fractional max-pooling," 2014, *arXiv:1412.6071*.

[58] X. Yin, J. Goudriaan, E. A. Lantinga, J. Vos, and H. J. Spiertz, "A flexible sigmoid function of determinate growth," *Ann. Botany*, vol. 91, no. 3, pp. 361–371, Feb. 2003.

[59] J. Cao, Y. Li, M. Sun, Y. Chen, D. Lischinski, D. Cohen-Or, B. Chen, and C. Tu, "DO-conv: Depthwise over-parameterized convolutional layer," 2020, *arXiv:2006.12030*.

[60] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, *arXiv:1404.1869*.

[61] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[62] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 3–19.

[63] S. G. Armato et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Jan. 2011.

[64] C. Jacobs, E. M. van Rikxoort, K. Murphy, M. Prokop, C. M. Schaefer-Prokop, and B. van Ginneken, "Computer-aided detection of pulmonary nodules: A comparative study using the public LIDC/IDRI database," *Eur. Radiol.*, vol. 26, no. 7, pp. 2139–2147, Jul. 2016.

[65] J. K. Peeples, J. F. Jameson, N. M. Kotta, J. M. Grasman, W. L. Stoppel, and A. Zare, "Jointly optimized spatial histogram UNET architecture (JOSHUA) for adipose tissue segmentation," *BME Frontiers*, vol. 2022, pp. 1–14, Jan. 2022.

[66] Y. Gao, M. Zhou, and D. N. Metaxas, "UTNet: A hybrid transformer architecture for medical image segmentation," in *Proc. 24th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Strasbourg, France. Cham, Switzerland: Springer, Sep./Oct. 2021, pp. 61–71.

[67] T. M. Quan, D. G. C. Hildebrand, and W.-K. Jeong, "FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics," *Frontiers Comput. Sci.*, vol. 3, May 2021, Art. no. 613981.

[68] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D. D. Feng, and S. Wang, "AUNet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Phys. Med. Biol.*, vol. 65, no. 5, Feb. 2020, Art. no. 055005.

[69] R. Roy, P. Banerjee, and A. S. Chowdhury, "A level set based unified framework for pulmonary nodule segmentation," *IEEE Signal Process. Lett.*, vol. 27, pp. 1465–1469, 2020.

[70] H. Liu, H. Cao, E. Song, G. Ma, X. Xu, R. Jin, Y. Jin, and C.-C. Hung, "A cascaded dual-pathway residual network for lung nodule segmentation in CT images," *Phys. Medica*, vol. 63, pp. 112–121, Jul. 2019.

[71] Y. Qin, H. Zheng, X. Huang, J. Yang, and Y.-M. Zhu, "Pulmonary nodule segmentation with CT sample synthesis using adversarial networks," *Med. Phys.*, vol. 46, no. 3, pp. 1218–1229, Mar. 2019.

[72] Y. Sun, F. Bi, Y. Gao, L. Chen, and S. Feng, "A multi-attention unet for semantic segmentation in remote sensing images," *Symmetry*, vol. 14, no. 5, pp. 1–19, 2022.

[73] M. Montazerolghaem, Y. Sun, G. Sasso, and A. Haworth, "U-Net architecture for prostate segmentation: The impact of loss function on system performance," *Bioengineering*, vol. 10, no. 4, pp. 1–19, 2023.

ZHANLIN JI (Member, IEEE) received the M.Eng. degree from Dublin City University, Ireland, in 2006, and the Ph.D. degree from the University of Limerick, Ireland, in 2010. He is a Professor with the North China University of Science and Technology, China, and an Associate Researcher with the Telecommunications Research Centre (TRC), University of Limerick. He has authored/coauthored 100+ research papers in refereed journals and conferences. His research interests include ubiquitous consumer wireless world (UCWW), the Internet of Things (IoT), cloud computing, big data management, and data mining.

ZIHENG ZHAO was born in 1998. He received the B.S. degree from the North China University of Science and Technology, China, in 2021. He is currently pursuing the master's degree. His research interests include machine vision and graphic image processing.

XINYI ZENG was born in 1999. She received the bachelor's degree from the Yancheng Institute of Technology, in 2021. She is currently pursuing the master's degree with the North China University of Science and Technology. Her research interests include machine vision and intelligent medical image processing.

JINGKUN WANG received the master's degree in industrial economics from the Inner Mongolia University of Technology, in 2018. She is currently a Research Assistant with Tsinghua University, Beijing, China. Her current research interests include healthcare blockchain, the Internet of Things, and big data.

LI ZHAO received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 1997 and 2002, respectively.

He is currently an Associate Professor with the Research Institute of Information Technology, Tsinghua University. His current research interests include mobile computing, the Internet of Things (IoT), e-health systems, intelligent transportation systems (ITS), home networking, machine learning, and digital multimedia.

XUEJI ZHANG is the Vice President of Shenzhen University, China. His research interests span the disciplines of chemistry, biology, materials, and medicine, with an emphasis on studies of biosensing, biomedicine, and biomaterials. He has received numerous national and international awards and honors, including a member of the Russian Academy of Engineering, a fellow of the American Institute for Medical and Bioengineering and the Royal Chemical Society, and a Simon Fellow of ICSC-World Laboratory. He received the National Innovation Award in China and the Scientist of the Year in China. He serves as the Co-Editor-in-Chief for *Sensors & Diagnostics* and has been an editorial member of 24 international journals.

IVAN GANCHEV (Senior Member, IEEE) received the Dip.Eng. (summa cum laude) and Ph.D. degrees from the Saint Petersburg University of Telecommunications, in 1989 and 1995, respectively. He is an International Telecommunications Union (ITU-T) Invited Expert and an Institution of Engineering and Technology (IET) Invited Lecturer, currently associated with the University of Limerick, Ireland, the University of Plovdiv "Paisii Hilendarski," and IMI-BAS, Bulgaria. He was involved in more than 40 international and national research projects. He has served on the TPC of more than 390 prestigious international conferences/symposia/workshops and has authored/coauthored one monographic book, three textbooks, four edited books, and more than 300 research articles in refereed international journals, books, and conference proceedings. He is on the editorial board and has served as a guest editor for multiple renowned international journals.

• • •