IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 June 2024, accepted 24 June 2024, date of publication 27 June 2024, date of current version 5 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3419908

RESEARCH ARTICLE

RELA_Net: Upper Airway CBCT Image Segmentation Model Based on Receptive Field Expansion and Large-Kernel Attention

HONGYONG GAO[1], WEIBO SONG[1], SHULIN CUI[2], BAOKANG ZHOU[1], HUACHAO TAN[1], AND QIANG WANG[3]

[1]School of Information Engineering, Dalian Ocean University, Dalian 116023, China
[2]Dalian Municipal Central Hospital, Dalian 116033, China
[3]School of Locomotive and Rolling Stock Engineering, Dalian Jiaotong University, Dalian 116028, China

Corresponding author: Weibo Song (swb@dlou.edu.cn)

This work was supported in part by the Educational Department of Liaoning Province under Grant LJKMZ20221110, and in part by the Capital Project of Dalian Municipal Health Commission under Grant 2111004 and Grant 2211006.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the Central Hospital of Dalian University of Technology (Dalian Central Hospital).

ABSTRACT The structure of the upper airway is variable and complex due to its environmental and physiological factors. Currently, doctors mainly rely on manual outlining and segmentation from images. This method is time-consuming and relies heavily on the doctor's experience. To solve this problem, we propose a fully automatic segmentation model for upper airway Cone Beam Computed Tomography (CBCT) images based on U-Net. The receptive field expansion module (RFEM) is used to replace the last three convolutional blocks of the encoder in the original U-Net model to improve the feature information extraction capability. And a large kernel attention module (LKA) is added to the skip connection part to dynamically adjust the receptive field of the feature extraction backbone, to alleviate the feature loss and redundancy of the skip connection. The dataset used in this paper is one created by us and the clinicians themselves, totaling 1345 CBCT images. Which were taken from 53 patients with airway obstruction. The imaging experts guided and delineated the label images. Experimental results show that the IoU and Dice score of the upper airway segmentation predicted by the RELA_Net network model in this article on the test sets are 94.39% and 97.10% respectively. Based on the prediction maps of the test set images, the segmentation model proposed in this article demonstrates an improvement in comparison to U-Net and other models, particularly in reducing over- and under-segmentation in the upper airway. This contributes to improving the diagnostic accuracy for patients with airway obstruction, thereby enhancing patient care and treatment planning.

INDEX TERMS U-net neural network, upper airway, medical image segmentation.

I. INTRODUCTION

For a long time, the upper airway has been an area of interest in respiratory medicine, as it is primarily responsible for respiratory, vocal, and swallowing functions [1]. The structure and dimensions of the upper airway are determined by a various of factors [2], [3], such as the soft tissues, muscles, and craniofacial skeleton surrounding the pharynx, all of which influence airway volume and facial growth patterns. When soft tissue laxity, adenoid, and tonsil hypertrophy are present, they predispose individuals to partial or total upper airway obstruction, which leads to the risk of obstructive

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

sleep apnea (OSA) [4], [5]. OSA is a common sleep breathing disorder that refers to episodes of apnea during sleep caused by partial or complete obstruction of the upper airway. It significantly affects patients' quality of life and may even be life-threatening [6].

One study [7] showed that pharyngeal airway dimensions were significantly larger in those who breathed through the nose than in those who breathed through the mouth. In growing individuals with oropharyngeal or nasopharyngeal obstruction caused by enlarged tonsils or adenoids, airway capacity is significantly increased after tonsillectomy or adenoidectomy. Therefore, it is important to perform regular airway examinations and undertake surgical interventions when necessary. Upper airway image segmentation allows for the visualization of lesion diagnosis, observation of treatment effects and helps to reduce the risk of OSA. Currently, for the segmentation of the upper airway, most doctors use the patient's CBCT images as the basis for effective segmentation of the upper airway region in images by manual delineation, which takes a lot of time and is inefficient. In addition, it is heavily dependent on the physician's experience, and for the same patient's CBCT image, different physicians may obtain different outlining results due to differences in experience [8]. There is a great deal of subjectivity in this manual segmentation, and these differences directly affect the effectiveness of the treatment and increase the risk of side effects for the patient. Therefore, in order to improve the efficiency of diagnosis, reduce the workload of doctors, and quantitatively assess the effect before and after treatment, it is important to establish a fast and accurate automatic upper airway segmentation model for the diagnosis and treatment of OSA patients [9].

II. RELATED WORK

In recent years, with the rapid improvement of computer hardware performance, deep learning technology has been rapidly developed. Compared with traditional machine learning and computer vision methods [10], [11], [12], methods based on deep learning have achieved good results in the field of image segmentation. Their excellent feature extraction and representation capabilities offer advantages in segmentation accuracy and speed. Deep learning models such as Convolutional Neural Network (CNN) [13], [14], [15] are becoming increasingly popular in medical image segmentation.

In the field of medical image segmentation, U-Net [16] is the base model used by most of the current segmentation algorithms. It uses skip connections to link the semantic features in the encoder and decoder paths, and finally obtains a feature map that contains both low-level semantic information and incorporates high-level semantic information, thus achieving accurate segmentation of medical images. Inspired by the success of U-Net, most of the recently proposed leading models are built on top of the U-Net architecture to cope with a variety of problems in different image segmentation tasks. Zhou et al. introduced the Unet++ [17] network model, which revised dense skip connection paths, resulting in a more adaptable network structure. This modification effectively mitigates the semantic divide issue associated with direct connections in U-Net. Huang et al. proposed the UNet3+ [18] network model based on Unet++ to add another full-scale skip connection method, which achieves the fusion of low- and high-resolution information at different scales.

In many images processing researches, combining specific attention can enhance the performance, such as the MRAUNet [19] model, which introduces a specially designed module called Multi-Resolution Attention that significantly improves the image quality and recognition accuracy. In paper [20], an attention-guided CNN architecture is proposed to combine feature maps with local details to improve classification performance. Similarly, attention mechanisms have been introduced in many medical image segmentation efforts [21], [22], [23], For instance, Oktay et al. integrated the attention mechanism into U-Net and introduced the Attention U-Net [24] network. This network learns the relationship between pixels during up-sampling through the attention gate module, increasing model sensitivity to foreground pixels. Ruan et al. proposed the MALUNet [25] network model, which incorporates multiple attention mechanisms. This model combines channel attention and spatial attention with a lightweight design to better focus and adjust the channel and positional information of the feature map. When trying to solve the problem that deeper networks are prone to gradient vanishing, He et al. [26] proposed the ResNet network model, which introduces a deep residual architecture that has been widely used in different segmentation networks. Some other studies [27] applied Transformers to UNet architectures, for example, the TransUNet [28] model applies Transformers to the encoder stage of the U-Net architecture. The combination of Transformer's ability to interact with global information and U-Net's ability to fuse multi-scale features is utilized, resulting in significant performance gains in image segmentation tasks.

In the task of segmenting of upper airway, accurate segmentation of upper airway structures is important for medical image analysis and diagnosis. However, existing methods [29] have some limitations when dealing with upper airway images with complex details and high noise inter-ference. In order to solve these problems, this paper proposes a model that combines a RFEM module and LKA module for upper airway image segmentation. The experimental results show that the network in this paper is able to obtain more accurate segmentation and achieve excellent performance. The main contributions of this paper are as follows:

1) We proposed RELA_Net network is based on the U-Net structure, replacing the conventional convolutional block with a RFEM module in the encoder part and adding a LKA module on the skip connections for accurate segmenting the upper airway parts in CBCT images.

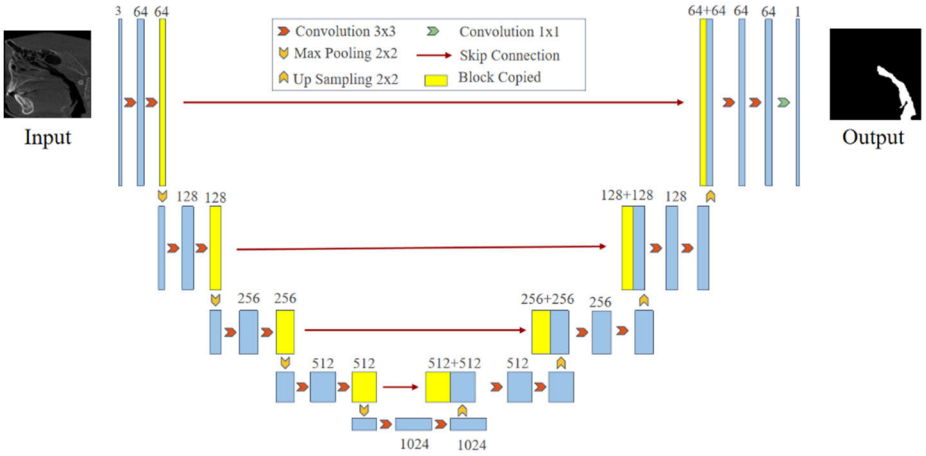

FIGURE 1. U-Net network model structure.

2) We have created a dataset of the upper airway under the guidance of a professional pathologist. The experimental results on this dataset, through comparison experiments with other methods and ablation experiments, confirm that RELA_Net achieves better performance in upper airway segmentation than other state-of-the-art methods.

3) To evaluate the generalization ability of our proposed RELA_Net, we validated our model on the publicly available dataset DSB 2018.

The rest of the paper is organized as follows: Section III describes in detail the network structure proposed in this paper. Section IV presents the dataset used in this study and provides a detailed description of the training process setup. Section V presents all the experimental results of this paper. Finally, Section VI concludes our findings and suggests potential avenues for future research.

III. METHODS

A. U-Net OVERVIEW

U-Net is the current benchmark model for mainstream image segmentation tasks and has had a profound impact on semantic segmentation since it was proposed, as shown in Fig. 1, the U-Net backbone network consists of three parts: encoder, decoder, and skip connections. In the encoding pathway, there are primarily feature extraction blocks and pooling layers. The feature extraction segment employs 3×3 convolutions and ReLU activation functions. A 2×2 pooling operation ensures maximum receptive field while simultaneously reducing the image resolution. In the decoding path, up-sampling and convolution are used for feature reconstruction, and the skip connection is the highlight of U-Net, which fuses information from different scales, thus preserving the original features of the image. Its lightweight structure and excellent performance have made it a classic model in the field of medical image segmentation.

B. RELA_Net ARCHITECTURE

Due to the problems of noise, blurring, and low contrast in medical images, their feature extraction is more challenging than ordinary images, and the ordinary convolutional module in the U-Net decoding path may lead to insufficient feature information extraction. Therefore, this paper proposes a new model called RELA_Net, and the detailed structure is shown in Fig. 2. We replace the last three layers in the encoding path from normal convolution blocks to RFEM modules fusing different dilated convolutions, each with a different dilation rate. The RFEM module helps to understand the context of the CT image by increasing the receptive fields extracted from the convolutional kernel, and can obtain the contextual information of different receptive fields while ensuring computational efficiency. In addition, the feature extraction process is prone to lose the target information due to the complex background of medical images. Therefore, a LKA attention mechanism is integrated at skip connections, allowing the model to dynamically prioritize significant image features, thereby enhancing feature extraction and selection.

1) RECEPTIVE FIELD EXPANSION MODULE

To more comprehensively extract advanced features from the image, there is an urgent need to capture non-local contextual information with a large receptive field, serving as an important complement and compensation. Drawing inspiration from KCPNet [30], this paper introduces an RFEM module based on dilated convolutions of densely connected voids, replacing the convolution blocks of the last three layers of U-Net. The key distinction from other expansive convolution modules, such as Atrous Spatial Pyramid Pooling (ASPP) in DeepLab, is that we assign greater weight to the initial feature map, thereby retaining more original information.

The dilation convolution operation used in the RFEM module is intended to expand the receptive field to capture broader context information. The dilation convolution is shown in

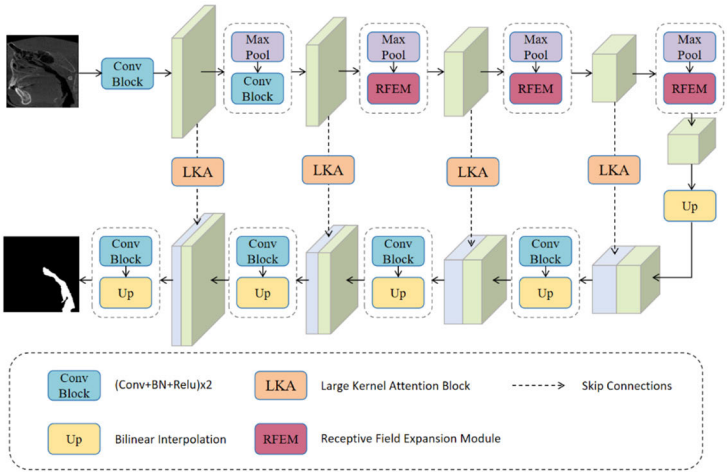

FIGURE 2. RELA_Net architecture.

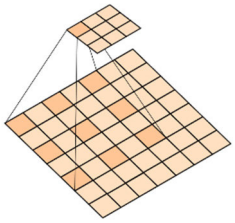

FIGURE 3. Dilation convolution.

Figure 3, by introducing a jump interval in the convolution kernel, which represents the jump step length between the elements in the convolution kernel, known as the dilated rate. In conventional convolution operations, each element of the convolution kernel is multiplied with the corresponding position of the input data. In contrast, in dilated convolution, the elements of the convolution kernel have a fixed spacing between them and are no longer densely arranged. This interval is determined by the dilation rate, and by increasing the dilation rate, dilated convolution can expand the receptive field of the convolutional kernel, enabling it to capture a broader range of input data. This mechanism is particularly useful in computer vision tasks. The dilated convolution is able to increase the network's receptive field while keeping the number of parameters constant, allowing the model to better handle features with different scales and hierarchies, compared to traditional convolution operations.

The RFEM module expands the receptive field by utilizing various dilated convolutional layers, and this module achieves this through multiple concatenation operations. The RFEM module effectively captures both non-local contextual features and local target features, enabling it to gather feature information across different scales. This capability facilitates enhanced fusion of global image information and strengthens the model's feature extraction capabilities. These features are useful for processing complex image data and extracting advanced features, and are therefore more suitable for use in deep networks.

The RFEM module consists of 3×3 dilated convolution layers and 1×1 ordinary convolution layers, and the specific RFEM module structure is shown in Fig. 4. The RFEM module sequentially processes the input features using dilated convolutions with different dilation rates and merges the processed features with the input features. Firstly, the dilated convolutions with dilation rates of 2, 4, 8 and 16 are performed sequentially for feature extraction. The two ordinary 1×1 convolutions in the RFEM module are placed before the dilated convolutions with dilation rates of 4, 8, and 16, respectively, which are used to reduce the number of feature channels and help to reduce the number of model parameters to avoid overfitting. By repeatedly concatenating with the original feature maps, this process avoids the significant dilation of original feature map information in subsequent convolutions.

The entire RFEM dilated convolution module operation can be summarized as follows:

$$R_1 = DConv3 \times 3_{r=2}(Input) \tag{1}$$

$$R_2 = DConv3 \times 3_{r=4}(Conv1 \times 1(R_1 \copyright Input)) \tag{2}$$

$$R_3 = DConv3 \times 3_{r=8}(Conv1 \times 1\,(R_1 \copyright R_2 \copyright Input)) \tag{3}$$

$$R_4 = DConv3 \times 3_{r=16}(Conv1 \times 1\,(R_1 \copyright R_2 \copyright R_3 \copyright Input)) \tag{4}$$

$$Output = R_1 \copyright R_2 \copyright R_3 \copyright R_4 \copyright Input \tag{5}$$

In the formula, "©" represents the concatenation of feature maps extracted from various dilated convolutions based on their channel dimensions. "Input" and "Output" correspondingly denote the initial input feature maps and the final output feature maps.

2) LARGE KERNEL ATTENTION MODULE

Transformer [31] based models such as Vision Transformer (ViT) [32] and others [33] have gained popularity in computer

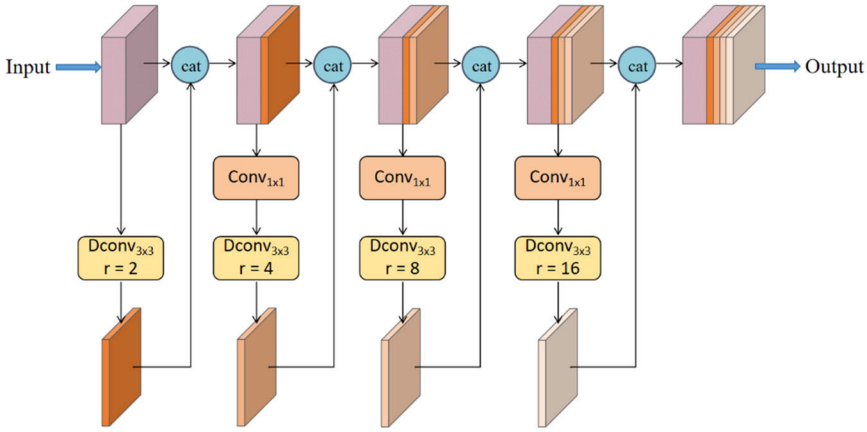

FIGURE 4. RFEM module structure.

vision due to their effectiveness in image recognition tasks. Studies [34], [35] have shown that a large receptive field is a key factor in their success. For example, ConvNeXt [36] uses 7×7 deep convolution in its backbone, which significantly improves the performance of downstream tasks. Additionally, D-LKANet [37] is a simplified attention mechanism using a large convolutional kernel that fully appreciate the volumetric context. Similarly, SegNeXt [38] demonstrated that large kernel convolution plays an important role in modulating convolutional features with richer context. LSKNet [39] can dynamically adjust its large spatial receptive field to better simulate the ranging context of various objects in a remote sensing scene. These studies show that convolutional networks with designs with large receptive fields can also achieve comparable performance to transformer-based models.

Due to the relatively high variability of different slices of the same site in medical images, which often requires a wide range of receptive fields and different ranges of contextual information. This is because parts of the upper airway that are not part of the upper airway may be incorrectly segmented without reference to a sufficiently distant context. Moreover, the majority of existing segmentation models employ a standard skip connection between the encoder and decoder, often overlooking the issue of image feature gaps and losses that occur during these connections.

In order to solve this problem, in this paper, we design a method of hybrid attention selection mechanism for large kernel called LKA. The module is designed to be able to capture features from different receptive fields by adapting a large convolutional kernel, combined with an attentional mechanism for filtering, in order to more efficiently deal with complex background changes in and around the object. This approach helps to improve the model's target recognition accuracy while reducing the sensitivity to background interference, making the model perform better when dealing with complex scenes. The hybrid attention selection mechanism simultaneously considers both channel and spatial information to weight the features extracted by convolutional kernels at different depths, thereby achieving precise feature extraction. This mechanism goes about dynamically adjusting the weights of each kernel based on inputs, enables the network to adaptively select kernels of different sizes and flexibly adjust the adapted receptive field. In this way, the network can autonomously choose the appropriate convolutional kernel size and receptive field range according to the characteristics of the input data and the task requirements, so as to process complex scenes and tasks more effectively.

Given the necessity for a broader contextual understanding in upper airway images, we use this characteristic for the segmentation task by adapting its larger spatial receptive field to blend and filter features effectively. The structure of the LKA module is shown in Figure 5. Each LKA module consists of a large kernel selection (LK) sub-block, which dynamically adjusts the network's receptive field as needed, and an attention sub-block, which consists of a mixture of channel attention (CA) and spatial attention (SA). CA uses globally averaged information to re-weight feature channels, while the SA module enhances the ability of cyberspace masks to model contextual information. By harnessing the strengths of both approaches, we sequentially aggregate information from large kernels across both spatial and channel dimensions. This enables us to capture richer information, enhancing the localization capability while preserving crucial details and suppressing noise and irrelevant information.

The overall module of LKA can be summarized as follows: Initially, the input feature maps traverse through various kernel functions of different sizes to acquire features with distinct receptive fields.

$$Attn1 = Conv1 \times 1(conv5 \times 5(Input)) \quad (6)$$

$$Attn2 = Conv1 \times 1(conv7 \times 7\,(conv5 \times 5(Input))) \quad (7)$$

where conv5 $\times$ 5 has a convolution padding of 2, conv7 $\times$ 7 has a dilated convolution rate of 3 and padding of 9, and 1×1 convolution is used to resize the number of channels to reduce the computation.

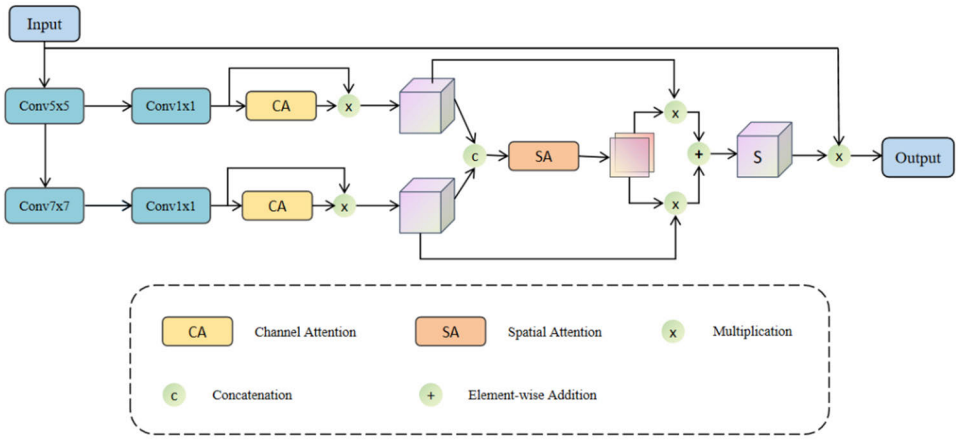

FIGURE 5. LKA module structure.

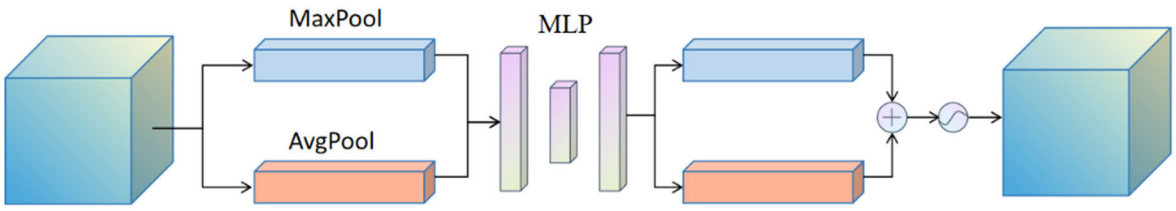

FIGURE 6. Channel attention module.

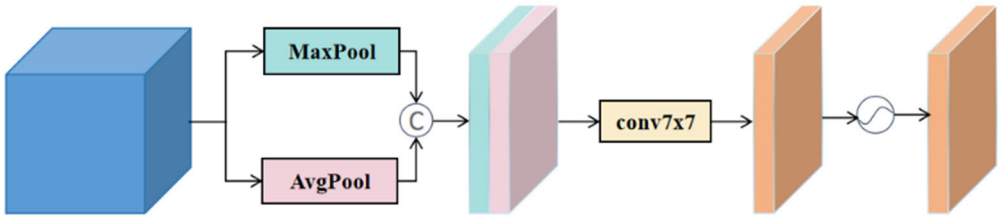

FIGURE 7. Spatial attention module.

Then, the extracted features are selected by CA. The CA is shown in Fig. 6 and is divided into two parts. Firstly, global average pooling (Avgpool) and global maximum pooling (Maxpool) are performed for the individual feature layers coming in as input, respectively. After that the obtained results are processed using the shared fully connected layer, the two processed results are summed and the Sigmoid is taken to fix the value to between 0-1 to obtain this weight.

The resulting weights are then multiplied by the original input feature layer Attn1:

$$Attn1 = CA\,(Attn1) \cdot Attn1 \tag{8}$$

$$Attn2 = CA\,(Attn2) \cdot Attn2 \tag{9}$$

Subsequently, feature concatenation takes place:

$$Attn = Attn1 \copyright Attn2 \tag{10}$$

After the feature maps are concatenated, the resulting feature map, denoted as Attn, undergoes feature selection using the SA module illustrated in Fig. 7, whose input feature maps are the concatenated feature maps of the CA outputs described above. The SA operation involves both averaging and max-pooling across channels of the input features, efficiently extracting spatial relations.

$$SAavg = Pavg(Attn), SAmax = Pmax(Attn) \tag{11}$$

where SA_{avg} and SA_{max} are the average and maximum pooled spatial feature descriptors. To allow information interaction between different spatial descriptors, the spatially merged features are connected, and the merged features are converted into 2 spatial feature maps using a convolutional layer F, after which the spatial features are obtained by applying a sigmoid activation function:

$$SA = \sigma(F([SAavg + SAmax])) \tag{12}$$

where $\sigma\ (\cdot)$ denotes the sigmoid function.

The features in the decomposed large kernel sequence are then weighted with the corresponding spatially-selective masks and fused with a 1 × 1 convolution to obtain the attentional features S:

$$S = F(SA[0] \cdot ttn1 + SA[1] \cdot ttn2) \tag{13}$$

The final output of the LKA module is the elemental product between the input features Input and S:

$$Output = Input \cdot S \tag{14}$$

IV. EXPERIMENTS

A. DATASET

1) UPPER AIRWAY CBCT DATASET

CBCT images of the upper airway are usually stored in the Digital Imaging and Communications for Medical (DICOM) file format, which is highly structured and contains a wealth of medical imaging information, but may sometimes contain unnecessary redundant data. When processing these CBCT scan images, the DICOM format was converted to png image format to facilitate subsequent analysis and processing. By comparing the three views, it was found that the sagittal plane of the upper airway CBCT image could demonstrate the overall airway area, and the nasopharyngeal, oropharyngeal, and hypopharyngeal airways could be observed sequentially from top to bottom, which was easy for the doctor to observe the narrow part of the airway. Therefore, the sagittal image of the CBCT image was chosen to label the upper airway.

The labelling tasks were as follows: firstly, clinicians from the Otorhinolaryngology Head and Neck Surgery Clinic of a certain affiliated central hospital demonstrated the labelling of CBCT images of the upper airway, and explained the knowledge of nasopharyngeal, oropharyngeal and hypopharyngeal parts of the upper airway. Subsequently, members of the group labeled the upper airway CBCT image data. The doctor then screened and optimized the labeling results to complete the upper airway CBCT image data labeling task. The dataset for this paper was selected from 53 patients with respiratory dyspnea, and CBCT data from 9 patients were randomly chosen as the test set, with the rest used for the training set. A total of 1345 upper airway CBCT images were labeled in the dataset, with 950 images allocated to the training set, 238 images to the validation set, and 157 images to the test set. Some of the patient input images and corresponding ground truth labels are displayed as shown in Figure 8.

In terms of data preprocessing, to prevent overfitting, we also applied some data augmentation techniques, including horizontal flipping, vertical flipping, and random rotation.

The source of the data has been authorized not to involve personal privacy.

2) DATA SCIENCE BOWL (DSB) 2018 DATASET

The DSB 2018 dataset is a public dataset from the Kaggle competition [40], which includes 670 cell nucleus images. We partitioned the original dataset into three parts: a training set of 428 images, a validation set of 107 images, and a test set of 135 images.

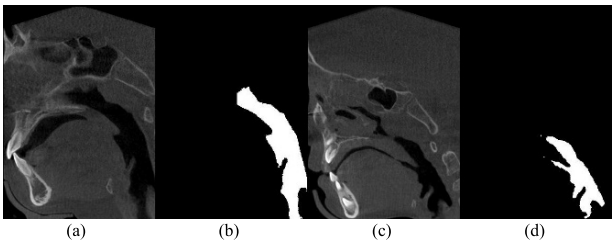

FIGURE 8. Dataset examples: (a) Original Image 1, (b) Ground Truth 1, (c) Original Image 2, (d) Ground Truth 2.

B. IMPLEMENTATION DETAILS

The primary experimental setup is as follows: The operating system used for the experiments is Windows 11. The GPU utilized is an NVIDIA GeForce RTX 3080 Ti with 16GB of RAM. The experiments were conducted using Python 3.8 and the PyTorch 2.0.1 framework. During the experimentation process, the batch size was set to 8. The input image size was uniformly adjusted to 256 × 256 pixels. We employed the Dice loss function as our segmentation loss. Training was conducted using the Adam optimizer with a learning rate of 0.0001 for all models. The training was carried out for 100 epochs.

All experiments in this study were conducted using the same loss function and parameter settings for training. During the training process, we did not utilize any pre-trained weights for the mentioned models.

C. EVALUATION METRICS

In this study, precision (Pre), recall (Rec), intersection over union (IoU), and Dice similarity coefficient (DSC) were selected as evaluation metrics.

Precision indicates the proportion of samples predicted as positive by the model that are indeed positive, relative to all samples predicted as positive. This criterion assesses the accuracy of predicting correct pixel samples, calculated using the formula:

$$Pre = \frac{TP}{TP + FP} \tag{15}$$

Recall represents the proportion of correctly predicted samples by the model relative to the total number of actual correct pixel samples. The formula is as follows:

$$Rec = \frac{TP}{TP + FN} \tag{16}$$

IoU, represents the intersection of the segmentation prediction result and the ground truth segmentation label divided by their union. A higher IoU value closer to 1 indicates better segmentation performance. The calculation formula is as follows:

$$IoU = \frac{TP}{FP + TP + FN} \tag{17}$$

TABLE 1. Segmentation results of different models on upper airway dataset.

Method	Pre	Rec	IoU	Dice
U-net	0.9870	0.9451	0.9332	0.9646
Atten-Unet	0.9842	0.9536	0.9390	0.9683
ResUnet	0.9822	0.9463	0.9300	0.9632
Unet++	**0.9874**	0.9377	0.9266	0.9614
DeepLabv3+	0.9830	0.9511	0.9353	0.9663
DCSAU_Net	0.9826	0.9556	0.9396	0.9685
TransUnet	0.9806	0.9580	0.9401	0.9688
RELA_Net	0.9831	**0.9597**	**0.9439**	**0.9710**

Dice similarity coefficient is a function used to measure the similarity of two samples and can be used to calculate the similarity between the predicted image and the labelled image, taking a value between 0 and 1. The formula is shown below:

$$Dice = \frac{2TP}{FP + 2TP + FN} \tag{18}$$

True Positive (TP) represents correctly predicting upper airway pixels. False Positive (FP) represents falsely predicting background pixels as upper airway pixels. True Negative (TN) represents correctly predicting background pixels. False Negative (FN) represents falsely predicting upper airway pixels as background pixels.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we will report two types of experimental results. We will compare our proposed model with previous methods on two datasets. Additionally, we will conduct an ablation study on the upper airway dataset to analyze the effectiveness of each module in RELA_Net. We will evaluate performance using Pre, Rec, IoU, and Dice coefficient.

A. COMPARISON WITH EXISTING METHODS

1) RESULT ON THE UPPER AIRWAY DATASET

In order to validate the overall segmentation performance of RELA_Net on medical image segmentation tasks, we compared RELA_Net with other UNet-based methods, including U-Net, ResUnet [41], Atten-Unet, DeepLabv3+ [42], DCSAU_Net [43], and other segmentation network models on airway segmentation datasets under the perspective of quantitative evaluation, and experimentally evaluated the various quantitative metrics for the prediction results of different models were obtained. The results of the assessment are shown in Table 1, where the best results are in bold. The experimental results demonstrate that the RELA_Net model proposed in this paper outperforms other segmentation network models across most evaluation metrics, achieving the highest scores in terms of Recall, Dice coefficient, and IoU metrics. The average IoU and Dice scores achieved by our method are 94.39% and 97.10%, respectively, showcasing improved segmentation accuracy. When compared with classical and recent neural network benchmarks such as U-Net, Atten-Unet, and ResUnet, our method exhibits enhancements in IoU and Dice coefficients ranging from 0.38% to 1.73% and 0.22% to 0.96%, respectively. These results indicate the superior accuracy of our segmentation network model in delineating the upper airway in CBCT images.

The visualizations of our segmentation model on this dataset are shown in Figure 9, where areas of missegmentation or omission are annotated with deep red boxes. The results indicate that segmenting the overall airway in upper airway images is relatively straightforward, while delineating the edges surrounding the upper airway poses challenges. Some models exhibit instances of under-segmentation and over-segmentation at the edges of the airway. For instance, U-Net and Unet++ show under-segmentation in certain image segments, while ResUnet and DCSAU_Net demonstrates erroneous segmentation in some segmentation outputs. In contrast, the model proposed in this paper achieves more precise edge segmentation of the upper airway CBCT images.

2) RESULT ON THE DSB 2018 DATASET

To evaluate the generalization ability of our proposed RELA_Net, we validated the model on the DSB 2018 dataset. The test results are presented in Table 2. In comparison with other methods, our model achieved the highest scores in Precision, IoU, and Dice, with scores of 91.21%, 84.96%, and 91.50% respectively, demonstrating the model's capability to enhance feature extraction and focus on relevant information, thereby exhibiting superior generalization performance.

Figure 10 displays the segmentation visualization results of several relevant networks on the DSB 2018 dataset. RELA_Net can more accurately depict the contours and shapes of cell nuclei. For instance, DCSAU_Net in the first row exhibits over-segmentation. From the results shown in the third row, our proposed model can better approximate the ground truth labels, while ResUnet shows under-segmentation and Deeplabv3+ demonstrates significant over-segmentation. These experimental results indicate that the features extracted by the U-shaped network integrating receptive field expansion and large-kernel attention modules contain better long-range semantic information and local details. Compared to other mainstream models,

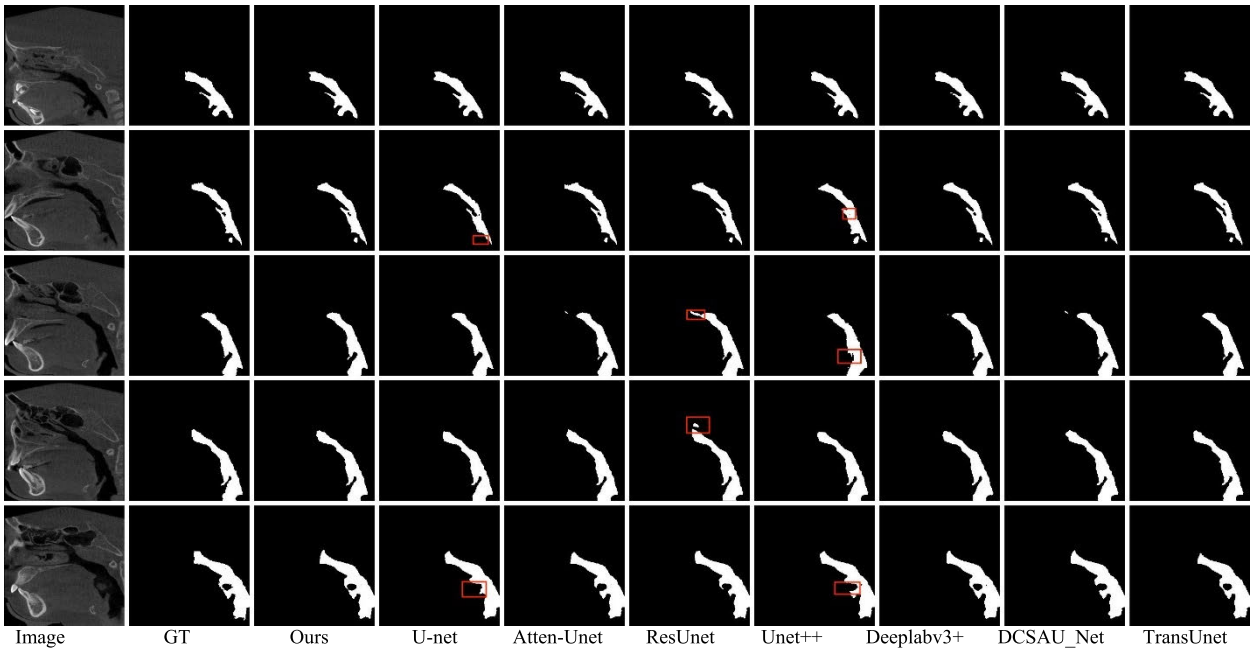

FIGURE 9. Visualization of the segmentation results of the different methods on upper airway CBCT segmentation dataset.

TABLE 2. Segmentation results of different models on DSB2018 dataset.

Method	Pre	Rec	IoU	Dice
U-net	0.9102	0.9217	0.8426	0.9103
Atten-Unet	0.9109	0.9256	0.8479	0.9140
ResUnet	0.8839	0.9162	0.8266	0.8951
Unet++	0.8977	**0.9338**	0.8420	0.9110
DeepLabv3+	0.8827	0.8933	0.8053	0.8827
DCSAU_Net	0.8963	0.9215	0.8311	0.9011
TransUnet	0.9093	0.9285	0.8455	0.9124
RELA_Net	**0.9121**	0.9263	**0.8496**	**0.9150**

RELA_Net can learn fine structures more effectively, leading to the generation of more precise contours.

B. ABLATION STUDY

To explore the influence of different factors on model performance, this study conducted ablation experiments with various variables on the upper airway segmentation dataset. The ablation experiments included: 1) Investigating the impact of replacing different layers of the encoder with dilated convolution blocks on network performance; 2) Comparative experiments with the addition of different attention; 3) Assessing the influence of added modules on network performance.

1) IMPACT OF REPLACING DIFFERENT LAYERS OF THE ENCODER WITH RFEM MODULES ON NETWORK PERFORMANCE

Table 3 illustrates the results of ablation experiments where the original convolution blocks were replaced with RFEM dilated convolution blocks at various layers of the encoder. According to the experimental findings, configuring the last three layers of the U-Net encoder as dilated convolution blocks facilitates deep feature extraction, offering a broader receptive field and contextual information. Compared to adding RFEM modules to shallow layers or all layers, this configuration demonstrates superior performance, notably enhancing the optimization stability of the model.

Noted: "√" indicates the adoption of the dilated convolution block, "-" indicates otherwise. The optimal data is highlighted in black.

2) COMPARATIVE EXPERIMENTS WITH THE ADDITION OF DIFFERENT ATTENTION

To validate the effectiveness of the large kernel attention mechanism we designed, we replaced different attention modules before and after the large kernel feature extraction and concatenation.

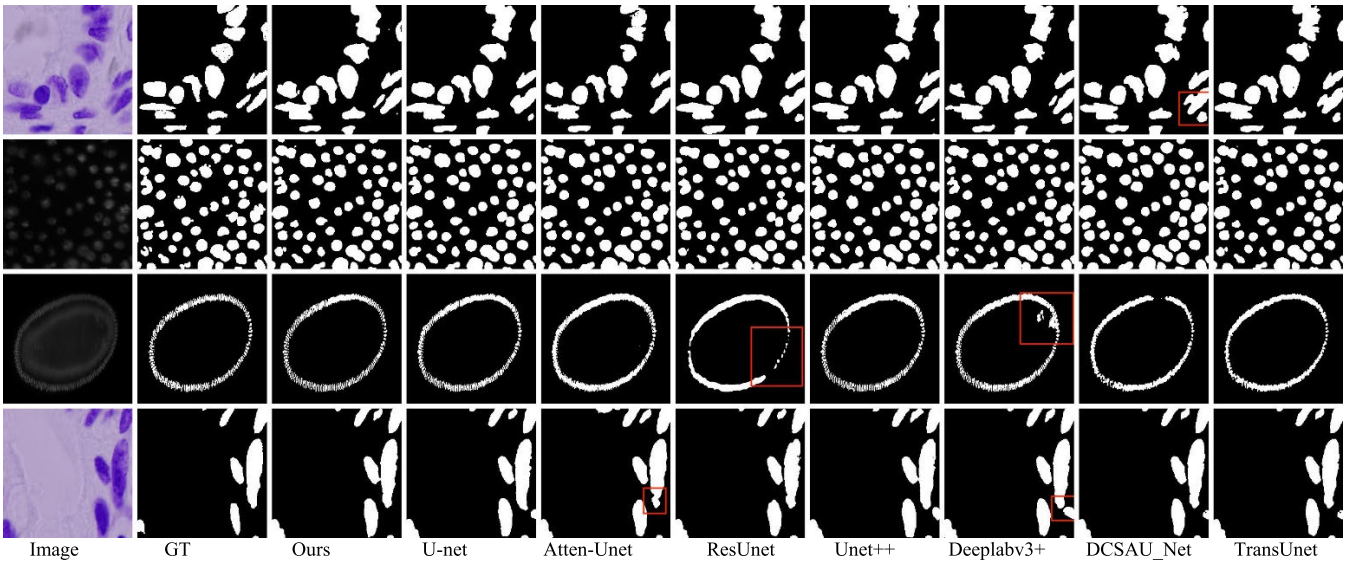

FIGURE 10. Visualization of the segmentation results of the different methods on DSB2018 dataset.

TABLE 3. Replacement of dilated convolution blocks in different layers of the encoder.

Layer1	Layer2	Layer3	Layer4	Layer5	Pre	Rec	IoU	Dice
-	-	-	-	-	**0.9870**	0.9451	0.9332	0.9646
√	√	-	-	-	0.9840	0.9525	0.9377	0.9676
-	-	√	√	√	0.9819	**0.9588**	**0.9418**	**0.9698**
√	√	√	√	√	0.9817	0.9567	0.9395	0.9685

TABLE 4. Comparison of experimental results with different attention added to LKA module.

Method	Pre	Rec	IoU	Dice
Baseline	**0.9870**	0.9451	0.9332	0.9646
Model 1	0.9833	0.9526	0.9371	0.9667
Model 2	0.9855	0.9468	0.9334	0.9647
Model 3	0.9846	**0.9570**	**0.9427**	**0.9702**
Model 4	0.9866	0.9525	0.9400	0.9688
Model 5	0.9864	0.9545	0.9419	0.9698

Table 4 presents a comparative analysis of incorporating different large kernel attention modules into skip connections. In this table, ''Baseline'' denotes the U-Net model, while the remaining five models integrate various large kernel attention modules within the skip connections. Model1 represents the incorporation of large kernel attention into the U-Net model, with the addition of the SA mechanism following the extraction of features using different large convolutional kernels. Structurally similar to the proposed LKA module, it does not include the CA module. Model2, however, replaces the SA attention in Model1 with CA attention. Model3 is our proposed LKA module, which is a CA module added before the feature concatenation of the Model1 module. In Model4, the CA module preceding the feature concatenation in the LKA module is replaced with the CoordAtt [44] module, while in Model5, it's substituted with the SE [45] module before the feature concatenation in the LKA module. The experimental results demonstrate that our designed LKA module achieved IoU and Dice scores of 94.27% and 97.02%, respectively. This validates its effectiveness in reducing information redundancy for skip connections. Compared to other added attention mechanisms, our LKA module exhibits superior performance in CBCT respiratory tract image segmentation.

3) THE IMPACT OF ADDED MODULES ON SEGMENTATION RESULTS

To further validate the effectiveness of the proposed modules, we conducted ablation studies by gradually adding dilated convolution blocks and large kernel attention modules to the U-Net model as the baseline network, while keeping the experimental setup and parameters consistent.

TABLE 5. Comparison of experimental results before and after adding different modules.

Method	Pre	Rec	IoU	Dice
Baseline	**0.9870**	0.9451	0.9332	0.9646
Baseline + RFEM	0.9819	0.9588	0.9418	0.9698
Baseline + LKA	0.9846	0.9570	0.9427	0.9702
RELA_Net	0.9831	**0.9597**	**0.9439**	**0.9710**

Firstly, in the U-Net network encoder, we replaced ordinary convolutional layers with the RFEM module, referred to as Baseline + RFEM. Secondly, we integrated the LKA module into the network, known as Baseline + LKA, to evaluate the performance of segmenting upper airway images. Lastly, we fused these two modules into the baseline network, forming the proposed RELA_Net network. Experiments were conducted on the upper airway dataset, and the experimental data is shown in Table 5. It can be observed that by adding the RFEM module and LKA module separately to the baseline model, the IoU increased by 0.86% and 0.95%, respectively, while the Dice score increased by 0.52% and 0.56%, respectively. It validates that both modules can focus on the detailed features of the images, enabling the model to fully extract image features and improve segmentation performance.

After incorporating the RFEM module into the model, the LKA module is added. It employs Large Kernel Attention to extract features, enhancing or suppressing different channels through a combination of SA and CA. This process aims to improve the feature extraction and screening capabilities of the image. Experimental results show that compared to the baseline network, the addition of these two modules leads to an improvement of 1.07% and 0.64% in Dice and IoU scores, respectively, on the upper airway dataset.

The experimental results demonstrate that the integration of the RFEM module not only enhances the extraction of deep-level hierarchical information at the encoding end but also, through the LKA skip connection structure, assigns corres-ponding weights to features at different scales. It enables more effective fusion of the features extracted by the encoder, thereby contributing to achieving better segmentation performance.

VI. CONCLUSION

This paper proposes a novel network model for 2D medical image segmentation based on U-Net, named RELA_Net. The model combines an encoder with ordinary convolutions and dilated convolutions, along with a decoder composed of convolutional upsampling layers. Additionally, to mitigate potential feature loss and redundancy in skip connections, the model incorporates a large-kernel attention mechanism. Experimental results demonstrate that the RELA_Net model outperforms several existing classical and state-of-the-art models in evaluation metrics such as IoU and Dice similarity index. These findings affirm the effectiveness and feasibility of the proposed approach for segmenting upper airway CBCT images.

Lastly, due to the complex and variable nature of airway structures, there is still room for improvement in the network's ability to extract features from this area. In future research, the primary tasks will be to continuously refine and optimize the model's architecture. Additionally, more effective image preprocessing algorithms could be introduced to ensure that the model adequately learns image features, thus comprehensively enhancing its segmentation capabilities for upper airway image segmentation tasks. In subsequent work, we will explore multi-class three-dimensional segmentation methods to identify and accurately classify regions of the upper airway, including the oropharynx and nasopharynx. We aim to significantly advance the field of upper airway image segmentation, thereby making important contributions to clinical diagnosis and the development of medical imaging.

REFERENCES

[1] Q. Xu, X. Wang, N. Li, Y. Wang, X. Xu, and J. Guo, "Craniofacial and upper airway morphological characteristics associated with the presence and severity of obstructive sleep apnea in Chinese children," *Frontiers Pediatrics*, vol. 11, Mar. 2023, Art. no. 1124610, doi: 10.3389/fped.2023.1124610.

[2] I. Indriksone and G. Jakobsone, "The influence of craniofacial morphology on the upper airway dimensions," *Angle Orthodontist*, vol. 85, no. 5, pp. 874–880, Sep. 2015, doi: 10.2319/061014-418.1.

[3] L. Zhang and H. Liu, "Influence of adenoid hypertrophy on malocclusion and maxillofacial development in children," *Evidence-Based Complementary Alternative Med.*, vol. 2022, pp. 1–6, Jul. 2022, doi: 10.1155/2022/2052359.

[4] C.-Y. Lin, C.-N. Chen, K.-T. Kang, T.-Y. Hsiao, P.-L. Lee, and W.-C. Hsu, "Ultrasonographic evaluation of upper airway structures in children with obstructive sleep apnea," *J. Amer. Med. Assoc. Otolaryngol., Head Neck Surg.*, vol. 144, no. 10, p. 897, Oct. 2018, doi: 10.1001/jamaoto.2018.1809.

[5] B. C. Neelapu, O. P. Kharbanda, H. K. Sardana, R. Balachandran, V. Sardana, P. Kapoor, A. Gupta, and S. Vasamsetti, "Craniofacial and upper airway morphology in adult obstructive sleep apnea patients: A systematic review and meta-analysis of cephalometric studies," *Sleep Med. Rev.*, vol. 31, pp. 79–90, Feb. 2017, doi: 10.1016/j.smrv.2016.01.007.

[6] M. P. Pase et al., "Sleep architecture, obstructive sleep apnea, and cognitive function in adults," *J. Amer. Med. Assoc. Netw. Open*, vol. 6, no. 7, Jul. 2023, Art. no. e2325152, doi: 10.1001/jamanetworkopen.2023.25152.

[7] C. N. Liu, K. T. Kang, C. J. Yao, Y. J. Chen, P. L. Lee, W. C. Weng, and W. C. Hsu, "Changes in cone-beam computed tomography pediatric airway measurements after adenotonsillectomy in patients with OSA," *J. Amer. Med. Assoc. Otolaryngol Head Neck Surg.*, vol. 148, no. 7, pp. 621–629, 2022, doi: 10.1001/jamaoto.2022.0925.

[8] N. Alsufyani, C. Flores-Mir, and P. Major, "Three-dimensional segmentation of the upper airway using cone beam CT: A systematic review," *Dentomaxillofacial Radiol.*, vol. 41, no. 4, pp. 276–284, May 2012, doi: 10.1259/dmfr/79433138.

[9] Ç. Sin, N. Akkaya, S. Aksoy, K. Orhan, and U. Öz, "A deep learning algorithm proposal to automatic pharyngeal airway detection and segmentation on CBCT images," *Orthodontics Craniofacial Res.*, vol. 24, no. S2, pp. 117–123, Dec. 2021, doi: 10.1111/ocr.12480.

[10] S. S. Al-Amri, N. V. Kalyankar, and S. D. Khamitkar, "Image segmentation by using edge detection," *Int. J. Comput. Sci. Eng.*, vol. 2, no. 3, pp. 804–807, 2010.

[11] S. S. Al-amri, N. V. Kalyankar, and S. D. Khamitkar, "Image segmentation by using threshold techniques," 2010, *arXiv:1005.4020*.
[12] H. P. Ng, S. H. Ong, K. W. C. Foong, P.-S. Goh, and W. L. Nowinski, "Medical image segmentation using k-means clustering and improved watershed algorithm," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Mar. 2006, pp. 61–65, doi: 10.1109/SSIAI.2006.1633722.
[13] X. Yan, H. Tang, S. Sun, H. Ma, D. Kong, and X. Xie, "AFTer-UNet: Axial fusion transformer UNet for medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 3270–3280.
[14] A. G.-U. Juarez, H. A. W. M. Tiddens, and M. de Bruijne, "Automatic airway segmentation in chest CT using convolutional neural networks," in *Proc. 3rd Int. Workshop RAMBO, 4th Int. Workshop BIA, 1st Int. Workshop TIA 2018, Held Conjunction MICCAI*, Granada, Spain. Springer, Sep. 2018, pp. 238–250.
[15] M.-L. Huang and Y.-Z. Wu, "Semantic segmentation of pancreatic medical images by using convolutional neural network," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103458, doi: 10.1016/j.bspc.2021.103458.
[16] 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.*, Munich, Germany. Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[17] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. 4th International Workshop DLMIA*, vol. 2018, Granada, Spain. Springer, Sep. 20, 2018, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.
[18] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," 2020, *arXiv:2004.08790*.
[19] A. Fateh, R. T. Birgani, M. Fateh, and V. Abolghasemi, "Advancing multilingual handwritten numeral recognition with attention-driven transfer learning," *IEEE Access*, vol. 12, pp. 41381–41395, 2024, doi: 10.1109/ACCESS.2024.3378598.
[20] C. Öksüz, O. Urhan, and M. K. Güllü, "An integrated convolutional neural network with attention guidance for improved performance of medical image classification," *Neural Comput. Appl.*, vol. 36, no. 4, pp. 2067–2099, Feb. 2024, doi: 10.1007/s00521-023-09164-x.
[21] Y. Wu, H. Jiang, and W. Pang, "MSRA-Net: Tumor segmentation network based on multi-scale residual attention," *Comput. Biol. Med.*, vol. 158, May 2023, Art. no. 106818, doi: 10.1016/j.compbiomed.2023.106818.
[22] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104815.
[23] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," in *Proc. 24th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Strasbourg, France. Springer, Oct. 2021, pp. 36–46, doi: 10.1007/978-3-030-87193-2.
[24] 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*.
[25] J. Ruan, S. Xiang, M. Xie, T. Liu, and Y. Fu, "MALUNet: A multi-attention and light-weight UNet for skin lesion segmentation," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2022, pp. 1150–1156.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[27] S. Soni, N. D. Londhe, R. Raj, and R. S. Sonawane, "TransUnet for psoriasis lesion segmentation," in *Proc. IEEE Bombay Sect. Signature Conf. (IBSSC)*, Dec. 2022, pp. 1–6, doi: 10.1109/IBSSC56953.2022.10037394.
[28] 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*.
[29] A. Javed, Y.-C. Kim, M. C. K. Khoo, S. L. D. Ward, and K. S. Nayak, "Dynamic 3-D MR visualization and detection of upper airway obstruction during sleep using region-growing segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 2, pp. 431–437, Feb. 2016, doi: 10.1109/TBME.2015.2462750.
[30] Y. Han, J. Liao, T. Lu, T. Pu, and Z. Peng, "KCPNet: Knowledge-driven context perception networks for ship detection in infrared imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5000219, doi: 10.1109/TGRS.2022.3233401.
[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. NeurIPS*, vol. 3, 2017, Art. no. 5000219.
[32] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[33] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
[34] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 12159–12168.
[35] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. S. Torr, and L. Zhang, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6877–6886.
[36] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.
[37] R. Azad, L. Niggemeier, M. Hüttemann, A. Kazerouni, E. K. Aghdam, Y. Velichko, U. Bagci, and D. Merhof, "Beyond self-attention: Deformable large kernel attention for medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2024, pp. 1287–1297.
[38] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "SegNeXt: Rethinking convolutional attention design for semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 1140–1156.
[39] Y. Li, Q. Hou, Z. Zheng, M.-M. Cheng, J. Yang, and X. Li, "Large selective kernel network for remote sensing object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 16748–16759.
[40] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, and A. E. Carpenter, "Nucleus segmentation across imaging experiments: The 2018 data science bowl," *Nature Methods*, vol. 16, no. 12, pp. 1247–1253, Dec. 2019.
[41] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
[42] 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.
[43] Q. Xu, Z. Ma, N. He, and W. Duan, "DCSAU-net: A deeper and more compact split-attention U-Net for medical image segmentation," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106626, doi: 10.1016/j.compbiomed.2023.106626.
[44] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13713–13722.
[45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

HONGYONG GAO received the B.E. degree in automation from Qingdao University of Technology, in 2022. He is currently pursuing the master's degree in control science and engineering with Dalian Ocean University, China. His current research interests include medical image segmentation, deep learning, and computer vision.

WEIBO SONG was born in Dalian, Liaoning, China, in 1981. He received the Ph.D. degree from Dalian University of Technology. He is currently with the College of Information Engineering, Dalian Ocean University. His current research interests include the intersection of medicine and engineering.

SHULIN CUI was born in Dalian, Liaoning, China, in 1980. He received the M.S. degree from Dalian Medical University. He is currently engaged in the research of upper airway obstructive diseases.

BAOKANG ZHOU received the B.E. degree in software engineering from Henan University of Engineering, Henan, China, in 2022. He is currently pursuing the master's degree in control science and engineering with Dalian Ocean University, Dalian, China. His current research interests include medical image segmentation and software engineering.

HUACHAO TAN received the B.S. degree in electronic information engineering from Shandong University of Technology, in 2020, and the M.S. degree in control science and engineering from Dalian Ocean University, Dalian, China, in 2022. His research interests include deep learning, computer vision, and object detection.

QIANG WANG received the Ph.D. degree from Dalian University of Technology, China. He is currently a Teacher with Dalian Jiaotong University, China. His research interests include artificial intelligence, big data analysis, machine learning, and deep learning.