*Imaging and Radiation Research* 2024, 7(1), 7128.
https://doi.org/10.24294/irr7128

Article

# COVID-19 lesions image segmentation method based on UniFormer

**Peng Geng[1,*], Ziye Tan[1], Xiao Cao[2], Xiao Wang[1], Yimeng Wang[1], Dongxin Zhao[1], Conghe Wang[1]**

[1] School of Information and Science Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China
[2] Hebei Hua Zheng Information Engineering Co., Ltd., Shijiazhuang 050043, China
*** Corresponding author:** Peng Geng, Gengpeng@stdu.edu.cn

**CITATION**

Geng P, Tan Z, Cao X, et al. COVID-19 lesions image segmentation method based on UniFormer. Imaging and Radiation Research. 2024; 7(1): 7128. https://doi.org/10.24294/jipd7128

**ARTICLE INFO**

Received: 13 April 2024
Accepted: 29 May 2024
Available online: 10 June 2024



**Abstract:** In view of the fact that the convolution neural network segmentation method lacks to capture the global dependency of infected areas in COVID-19 images, which is not conducive to the complete segmentation of scattered lesion areas, this paper proposes a COVID-19 lesion segmentation method UniUNet based on UniFormer with its strong ability to capture global dependency. Firstly, a U-shaped encoder-decoder structure based on UniFormer is designed, which can enhance the cooperation ability of local and global relations. Secondly, Swin spatial pyramid pooling module is introduced to compensate the influence of spatial resolution reduction in the encoder process and generate multi-scale representation. Multi-scale attention gate is introduced at the skip connection to suppress redundant features and enhance important features. Experiment results show that, compared with the other four methods, the proposed model achieves better results in Dice, loU and Recall on COVID-19-CT-Seg and CC-CCIII dataset, and achieves a more complete segmentation of the lesion area.

**Keywords:** convolutional neural network; COVID-19 lesion image segmentation; self-attention mechanism; multiscale attention gate; spatial pyramid pooling

## 1. Introduction

Since the end of 2019, the COVID-19 pandemic has affected all aspects of human life. COVID-19 causes multiple issues, including dry cough, fever, headache, myalgia and chest troubles [1]. The diagnostic methods for COVID-19 using medical imaging technology mainly include computed tomography (CT) [2], magnetic resonance imaging (MRI) [3], and X-ray [4]. Compared with X-ray scanning, CT images has higher resolution and higher contrast and are better than X-ray images in displaying soft tissues and small lesions [5]. COVID-19 image segmentation can be introduced to accurately diagnose diseases and provide important information for doctors [6]. However, COVID-19 image segmentation requires experienced radiologists to complete. When faced with a large number of COVID-19 CT images, the manual lesions segmentation consumes a lot of time and is labor-intensive. At the same time, the results of lesions segmentation are easily affected by the radiologist's experience. These subjective and objective factors may lead to large deviations in COVID-19 CT images segmentation [7–9]. Therefore, it is necessary to design robust and accurate COVID-19 CT image segmentation method.

In COVID-19 lung image segmentation, UNet is a commonly used lung region and lesion segmentation. Milletari et al. [9] proposed V-Net in which residual blocks are used as basic convolutional blocks. On this basis, due to the infected area with low contrast in COVID-19 images and the large differences in the infected areas of different patients, accurate segmentation of the infected area is very challenging. In response to the slight differences among healthy tissues. infected tissues and noise,

Wang et al. [10] proposed COPLE-NET for segmenting COVID-19 infected lesions, which introduced maximum pooling and average pooling together in the encoding stage.

Compared with UNet, UNet++ [11] has the advantage of capturing different levels of features. Wang et al. [12] proposed a two-stage method for separating lesions from the lung COVID-19 CT images based on UNet++. In order to segment out more complete structures and more accurate detail information, Cong et al. [13] proposed an end-to-end COVID-19 infection segmentation network

Although the importance of boundary feauture is taken into account in BSNet, it is difficult to segment infected areas because the lesions on the chest CT images are scattered and it is difficult to obtain global semantic information for CNN-based methods. Ibtehaz et al. [14] modified skip connections and the convolution blocks in the ordinary UNet to strengthen the ability of long-range dependencies and multi-level feature combination.

In summary, although CNN-based method above has extraordinary feature representation capabilities, its limited receptive field limiting the accuracy of COVID-19 CT image segmentation. Zhou et al. [15] used the encoder of U-Net to obtain feature representation, input the feature representation of each layer into the attention mechanism, reweighted along the channel direction and spatial direction to obtain the most informative representation, and finally obtain the segmentation result through the decoder. Li et al. [16] aimed at the shortcomings of COVID-19 image segmentation method, such as low contrast between ground-glass opacity and background, blurred boundaries, and difficulty in accurate segmentation. A reverse attention module [13] was added to the skip connection of Unet as a fine marker to identify infected areas in the cleaning strategy. This method can learn the details of the complementary areas and focus on the segmentation of the boundary areas. Xiao et al. [17] proposed a new improved UNet++ model, in which squeeze and excitation attention blocks are adopted to adjust channel of the feature map. The the weights of task-related pixels are strengthened and the background and noise are suppressed. Zhao et al. proposed a Unet++ variant architecture SCOATNet [18] where a new spatial and channel attention module are proposed. The attention mechanism helps to enhance the weight of the infected area in the COVID-19 image and suppress interference in non-lesion areas, thereby raising the accuracy of COVID-19 image segmentation. However, the attention mechanism model fails to capture global dependency because of its smaller receptive field [19]. Therefore, the lesion area in the COVID-19 image cannot be perceived in the global scope, making the effect of the COVID-19 image segmentation poor. To overcome these difficulties, this paper designs a new image segmentation method UniUNet based on UniFormer which has better capabilities of capturing global information. The main contributions of this paper are list as following:

(1) A U-shaped network structure based on UniFormer is proposed, which can effectively remove local redundant information of adjacent slices, and build dependencies on distant lesion areas to improve the accuracy of lesion areas segmentation.

(2) Aimed to capture information of different scales, the Swin spatial pyramid pooling module is introduced into the bottom of the encoder. The module captures the global and local features of lesions through windows with different scales, forming a

multi-scale feature representation, which is helpful to strengthen the ability to capture COVID-19 lesions features.

(3) Multi-scale attention gate is introduced into each skip connection. The module selects features adaptively through convolution of three different receptive fields, and selects valuable features through point convolution voting to further improve the segmentation effect.

## 2. UniUNet network structure

The proposed UniUNet in this paper is shown in **Figure 1**, which uses the UniFormer blocks instead of the traditional convolution blocks to form a symmetric encoder-decoder structure with skip connection. Firstly, the COVID-19 lesion images are divided into image blocks which are input into the encoder. In the encoder, UniFormer blocks are introduced to carry out local to global self-attention processing on the feature maps. Swin spatial pyramid pooling module is added at the bottom of the encoder to capture features of different scales. The decoder uses the extension layer of the image block for up-sampling, and fuses the multi-scale features from the encoder through the skip connection structure with multi-scale attention gate to enhance the valuable features. Finally, the width and height of the feature maps is restored by linear mapping, and the pixel-level COVID-19 image segmentation is realized.

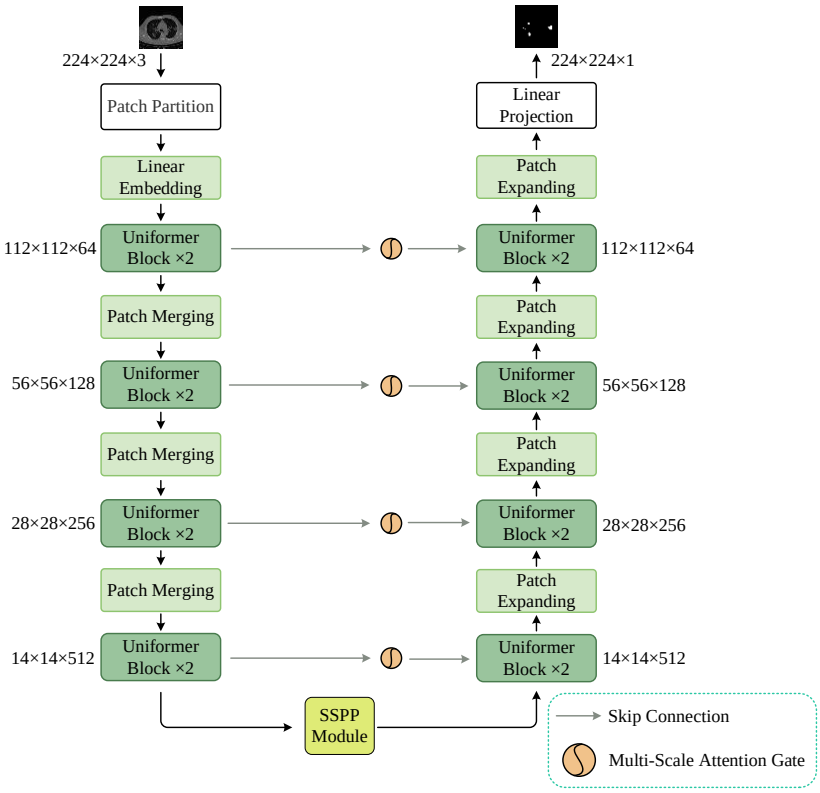


**Figure 1.** UniUNet model structure.

Patch Merging layer is used reduce resolution of feature maps with down sampling, adjusting the number of channels, and saving computation while maintaining information integrity. Unlike traditional pooling operations, Patch

Merging achieves down sampling by concatenating adjacent patches, ultimately reducing the number of channels through linear layers. In contrast to patch merging, a Patch Expanding layer is adopted to enlarge the resolution of feature maps to twice b up-sampling operation. The implementation process of Patch Merging and Patch Expanding can refer to reference [20].

## 2.1. UniFormer

To address the issue of local redundancy and global dependence in video, Li et al. [21] proposed UniFormer, which seamlessly integrated the advantages of spatio-temporal self-attention and three-dimensional convolution, A good balance between computational complexity and accuracy has been achieved. The aggregator learns local relations in the shallow layer through a small learnable parameter matrix, and learns global relations in the deep layer through similarity comparison, which effectively balances the computation cost and accuracy. Although UniFormer was originally designed for video processing, this study applied it to two-dimension COVID-19 CT image segmentation to solve the local similarity and the global dependence between infected areas, thus improving the segmentation accuracy. Specifically, the UniFormer module is composed of; three key parts: Multi-Head Relation Aggregator (*MHRA*), Dynamic Position Embedding (*DPE*) and Feed-Forward Network (*FFN*). The UniFormer module formula is as follows:

$$X = DPE(X_{in}) + X_{in} \tag{1}$$

$$Y = MHRA(Norm(X)) + X \tag{2}$$

$$Z = FFN(Norm(Y)) + Y \tag{3}$$

## 2.2. Swin spatial pyramid pool module

The encoder of UniUNet model includes UniFormer block and patch merging layer, but this will lead to the decrease of spatial resolution. Research [22] suggests that multiscale strucute is particularly effective for extracting high-level feature. In order to strengthen the spatial representation and the multi-scale representation, the Swin spatial pyramid pooling [23] module (SSPP) is inserted at the bottom of the encoder, as shown in **Figure 2**, which captures the global and local features of COVID-19 images with different windows to form the multi-scale representation. Then, these features are sent to the Cross Contextual Attention module for nonlinear fusion to capture a more comprehensive image representation.

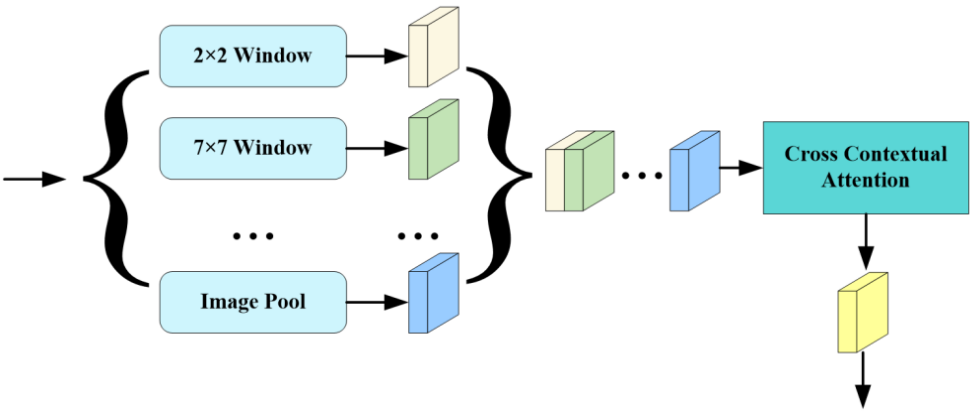


**Figure 2.** Swin spatial pyramid pooling module.

### 2.3. Multi-scale attention gate

In the original UNet network structure, not every layer of encoder features can output useful features after skip connection to help the network segment the lesion area. For different datasets and different parameter settings, some redundant information and noise may be generated, which may lead to incorrect segmentation on the infected area of the COVID-19 image, thus affecting the segmentation results of the network model. In order to suppress unimportant and redundant information and enhance valuable features during the skip connection process, this paper introduces the Multi-Scale Attention Gate (MSAG) module [24] in the skip connection part of the structure, as shown in **Figure 3**. In order to adaptively select features with different resolutions, Pointwise convolution, Standard convolution, and dilated convolution combine are combined to extract features with different receptive fields. Each convolution has a batch normalization layer. The feature maps generated by the three kinds of convolutions are of the same size. Before ReLU activation function, the feature maps are connected. Another point convolution is used to capture the important features.

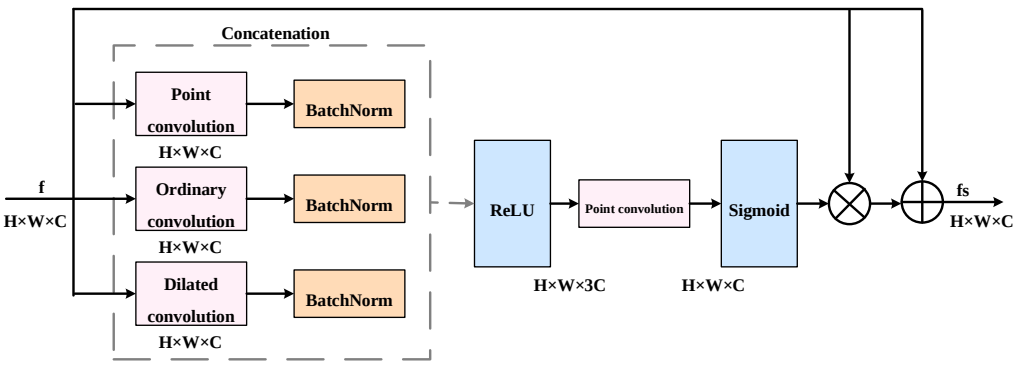


**Figure 3.** Multi-scale attention gate module.

## 3. Experimental setting and evaluation metrics

### 3.1. Dataset

To verify the effect of the proposed UniUNet, we used two publicly\COVID-19

datasets, namely the CC-CCII dataset [25] and the COVID-19-CT-Seg [26] dataset. The COVID-19-CT-Seg dataset is made up of 20 labeled COVID-19 CT sequences captured from 20 patients. The leision areas were labeled by two radiologists and verified by radiologists. The CT resolution is 512 × 512 and 680 × 680. There are 3320 slices in this dataset and 1843 slices exits lesion. We randomly selected 1394 slices from 14 patients for training and testing. The 70% slices are used as training sets and the other is set as test sets. All images were resized to 224 × 224, and all slices of each patient are only used as training or test. In CC-CCII segmentation dataset, there are 750 slices from 150 COVID-19 patients. The background, lung area, ground-glass opacity, and consolidation are manually labled in all of slices. It is worth noting that only consolidation areas and ground-glass opacity are included. There are 540 images with infected areas in CC-CCII dataset. All of images in CC-CCII dataset are randomly separated into training sets and test sets according to the ratio 7:3.

### 3.2. Experiment settings

All of the resolution is adjusted to 224 × 224. NVIDIA GeForce GTX 1660 SUPER GPU is used for training. The RMSProp optimizer is used. The weight decay and momentum are set to 1e-8 and 0.9, respectively. The two different datasets are trained for 100 and 300 epochs, respectively, the batchsize and initial learning rate were set to 6 and 0.000125, respectively. In addition, the early stopping strategy is used to train the proposed UniUnet. The training will be terminated if the metrics did not increase within 30 epoches. During the process of training, the binary cross entropy loss (BCE Loss) has the advantages of low calculation cost, and fast convergence speed, and is good at the task of binary classification. However, the BCE Loss is limited in effectiveness in dealing with class imbalance issue. The Dice Loss function has strong ability to unbalanced task, but it may be instable to segment small lesion areas in the image. In most COVID-19 images, the lesion are usually scattered and small, which is the issue of small object segmentation. Hybird Loss of Dice Loss [27] and BCE Loss [28] can effectively avoid these disadvantages and focus on detail information more stably. Therefore, the Dice loss function work together with the binary cross entropy to train the proposed UniUnet model. The hybrid loss is expressed as:

$$Loss = \alpha L_{bce} + L_{dice} \tag{4}$$

$$L_{bce} = -\frac{1}{N}\sum_{i=1}^{N}[(1 - y_i)\, log(\,1 - p_i) + y_i\, log\, p_i] \tag{5}$$

$$L_{dice} = 1 - \frac{\sum_{i=1}^{t} y_i p_i + \varepsilon}{\sum_{i=1}^{t} y_i + p_i + \varepsilon} \tag{6}$$

### 3.3. Evaluation metrics

Six metrics are used to quantitatively, evaluate the effect of different methods. They are Dice, IoU, Accuracy, Precision, Recall, and Specificity [29–30]. The definitions of the above important indicators are as follows:

(1) IoU is Intersection-over-Union ratio: also known as the Jaccard index which is one of the most commonly used indicators in medical image segmentation. It is the ratio of the overlapping area between the Ground truth and the segmentation result to the union area between them.

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{7}$$

Dice is same expression formula as $F1 - Score$ is used to evaluate the similarity between the Ground truth and the segmentation result. The larger the value of Dice, the closer the algorithm segmentation result is to the Ground truth, and the better the segmentation effect.

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \tag{8}$$

(2) Recall is used to indicate the ratio of the number of pixels in the lesion area correctly classified by the network structure to the total number of pixels in the lesion area.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

(3) Precision is used to indicate the ratio of the pixels of the lesion correctly classified by the network model to the total pixels of the lesion in the prediction result.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

(4) Accuracy is used to indicate the ratio of the pixels in the infected area and the pixels in the non-infected area accurately classified by the model to the all of pixels in the image. The higher the accuracy, the better the performance of the segmentation algorithm.

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN} \tag{11}$$

Specificity indicates the ratio of the pixels in non-infected area correctly classified by the model to all of pixels in non-lesion area. The higher the Specificity is the better the network can distinguish between lesion areas and non-lesion areas.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{12}$$

where *TP* represents True Positive which represents that the actual sample is positive and the predicted result is positive. *FP* represents False Positive which represents that the actual sample is negative, but the predicted result is positive. *TN* represents True Negative which means that actual sample is negative ads the predicted result is negative. *FN* represents False Negative which means that the actual sample is positive, but the predicted result is negative.

## 4. Experiment results and analysis

### 4.1. Ablation experiment

To verify the effectiveness of the SSPP module in the model in capturing multi-scale representation and the MSAG module in suppressing unimportant information, this section completes the ablation experiment of each module on the CC-CCII dataset. As can be seen from **Table 1**, this section uses the network structure without the SSPP module and the MASG module as the backbone network. With the SSPP module, the Dice index is improved by 3.72%, indicating that adding multi-scale representation to the network is effective, and multi-scale context extraction is conducive to capturing the global information of each lesion block in the image and the details on each lesion edge; after adding the MASG module, except for some decreases in accuracy, precision and specificity, other indicators have increased to varying degrees, among which the Dice value, IoU value and Recall value increased by 1.44%, 1.78% and 6.31%, respectively, which further proves that adding MSAG to the skip connection can suppress noise and enhance valuable features, so that the network retains more valuable global and small lesion areas during decoding. The ablation experiment results show that both SSPP and MSAG blocks can effectively increase the COVID-19 CT image segmentation performance.

In addition, FLOPsand parameter after embedding SSPP and MASG. 1 MASG, 2 MSAG, 3 MSAG, and 4 MSAG represent the with different number of MSAG moudles from bottom to top in the skip connection shown in **Figure 1**. For example, Backbone+SSPP+1 MSAG represents one MSAG module is embedded in the bottom skip connection in **Figure 2**. Backbone + SSPP + 2 MSAG represents that the other MSAG module is inserted tin the second skip connection from bottom to top in **Figure 2** based on Backbone + SSPP + 1 MSAG. From **Table 2**, it can be seen that the SSPP module has significantly increased FLops and Parmeters. Only one MSAG module resulted in a slight increase in FLops. However, adding four MSAG modules significantly increases the parameter count and FLOPS compared to only SSPP modules.

**Table 1.** Ablation experiment on CC-CCII dataset (%).

| **Methods** | **Dice** | **Recall** | **IoU** | **Precision** | **Specificity** | **Accuracy** |
|---|---|---|---|---|---|---|
| Backbone | 76.22 | 73.52 | 62.12 | 79.96 | 99.68 | 99.25 |
| Backbone + SSPP | 79.94 | 76.71 | 66.88 | 83.98 | 99.75 | 99.37 |
| Backbone + SSPP + MSAG (UniUNet) | 81.38 | 83.02 | 68.75 | 80.17 | 99.63 | 99.36 |

**Table 2.** FLOPs and Params with MSAG or not.

| **Methods** | FLOPs(G) | Params(M) |
|---|---|---|
| Backbone | 55.74 | 9.75 |
| Backbone + SSPP | 82.92 | 33.56 |
| Backbone + SSPP + 1 MSAG | 89.78 | 33.65 |
| Backbone + SSPP + 2 MSAG | 96.60 | 34.01 |
| Backbone + SSPP + 3 MSAG | 103.40 | 35.46 |
| Backbone + SSPP + 4 MSAG | 110.19 | 41.23 |

## 4.2. Comparison with other methods

Aim to prove the effectiveness of the proposed UniUNet, two COVID-19 image

segmentation methods and two medical image segmentation methods are compared. Infnet [31] is a COVID-19 image segmentation method and has been widely used as the baseline for COVID-19 image segmentation. BSNet [13] is based on modeling semantic relationship and guidance of boundary detail to segment lesion area more completely. TFCNs [32] is a state-of-the-art medical image segmentation method constructed using ResLinear Transformer and convolutional neural networks. Swin-Unet [20] is an effective medical image segmentation method using transformer technology to construct a U-shaped structure similar to the proposed UniUNet. Swin-Unet was widely thought as the baseline for medical image segmentation.

(1) Comparison on COVID-19-CT-seg dataset.

As can be seen from **Table 3**, on COVID-19-CT-Seg dataset, UniUNet is superior to Swin-Unet in all evaluation indexes, indicating that it has better training effect on small datasets and can capture the global dependence of lesion areas more effectively. Swin Unet has the lowest Dice, IoU and Precision among these five methods. The Dice, IOU, Recall and Precision of InfNet and BSNet method are relatively closer and better than Swin-Unet. But they are all lower than the proposed UniUNet in this work. Compared with the classic InfNet, UniUNet has improved Dice, Recall and IoU by 3.68%, 5.72% and 4.39% respectively, achieving better results.

In order to analyze the segmentation results of each method in a visual way, **Figure 4** demonstrates the comparison on predicted lesion area with different model. **Figure 4** illustrates that the InfNet cannot efficaciously obtain the edge details of the lesions which is an important basis for doctors to diagnose COVID-19. For the second column and the third column, UniUNet was able to accurately locate several lesion areas in the images, indicating that the proposed UniUNet can better establish the global relationship in the lesion areas. For the upper left lesion area in the fifth column, InfNet and BSNet did not segment the whole lesion area, while TFCNs segmented only part of lesion and showed obvious over segmentation. Compared with other methods, UniUNet is closest to Ground truth in segmentation accuracy, which proves the advantages of the method based on Transformer in COVID-19 lesion segmentation.

**Table 3.** Quantitative comparison on COVID-19-CT-Seg dataset (%).

| Methods | Dice | IoU | Recall | Precision | FLOPs(G) | Params(M) |
|---|---|---|---|---|---|---|
| InfNet | 74.01 | 60.22 | 78.83 | 74.14 | 31.52 | 30.91 |
| TFCNs | 73.22 | 58.91 | 74.73 | 76.02 | 168.40 | 105.79 |
| BSNet | 74.97 | 61.94 | 79.11 | 76.20 | 210.39 | 43.99 |
| Swin-Unet | 69.39 | 54.63 | 76.11 | 67.26 | 35.46 | 27.14 |
| UniUNet | 77.69 | 64.61 | 84.55 | 74.37 | 110.19 | 41.23 |

In addition, **Table 3** also provides a comparison of FLOPs and parameters for different methods. In comparison with the TFCNs, the FLOPs and parameters of the UniUNet method are remarkablely decreased. Compared with BSNet, the FLOPs of the UniUNet method are significantly reduced. The FLOPs and parameters of the UniUNet method are significantly higher than those of InfNet and Swin Unet. However, compared to these four methods, it can be seen from **Figure 4** and **Table 3** that UniUNet significantly improves segmentation accuracy.

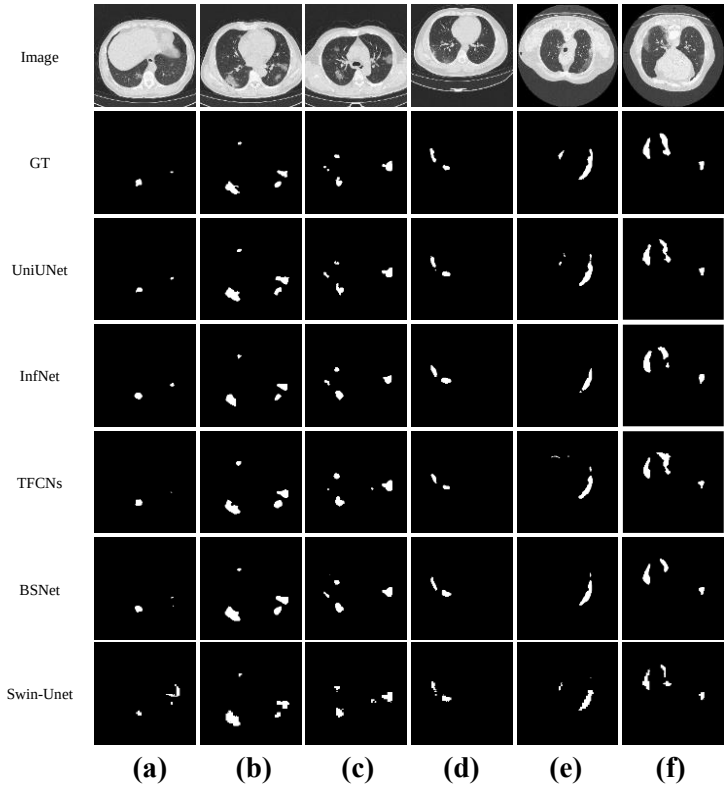


**Figure 4.** Comparison of segmentation results in COVID-19-CT-Seg dataset.

(2) Comparison on CC-CCII datasets.

In order to verify the effectiveness of the UniUNet model on CC-CCII dataset, another comparative experiment is carried out on the CC-CCII dataset, and six indicators of different methods are list in **Table 4**. The results in **Table 4** verify that UniUNet outperforms other methods in key indicators such as Dice, IoU and Recall, which are 81.38%, 68.75% and 83.02% respectively. In addition, the second or third best results have been achieved on the indicators of Precision, Accuracy and Specificity. Comprehensive analysis of these indicators shows that UniUNet has excellent segmentation performance in COVID-19 image segmentation task, which proves that it is good at COVID-19 lesion segmentation.

**Table 4.** Metrics Comparison on CC-CCII dataset (%).

| Methods | Dice | IoU | Recall | Precision | Accuracy | Specificity |
|---|---|---|---|---|---|---|
| InfNet | 77.28 | 63.12 | 75.03 | 80.55 | 99.27 | 99.68 |
| TFCNs | 76.09 | 61.73 | 78.09 | 74.64 | 99.18 | 99.53 |
| BSNet | 81.07 | 68.33 | 79.73 | 83.10 | 99.39 | 99.72 |
| Swin-Unet | 65.68 | 49.34 | 70.15 | 62.77 | 98.81 | 99.29 |
| UniUNet | 81.38 | 68.75 | 83.02 | 80.17 | 99.36 | 99.63 |

In addition, a comparison of the segmentation results of each method on the CC-CCII dataset is illustrated in **Figure 5**. **Figure 5** demonstrates that segmented results by the other methods have missing segmentation or over-segmentation. For example, the image in the first row, the InfNet, TFCNs and BSNet methods have obvious

missing segmentation on the lesion area at the lower right, and they have not divided it into connected lesion areas. The Swin-Unet method not only presents zigzag phenomenon in the segmentation result, but also has serious over-segmentation. These phenomenon of missing segmentation or over-segmentation indicates that these methods cannot effectively model the global dependency of the lesion. Only UniUNet can completely and accurately segment the larger lesion area. It shows that the proposed method can more accurately locate each part of the region and effectively remove the noise, which is because the multi-scale attention gate module is added between the skip connection to suppress the noise and enhance the valuable features. However, UniUNet is not sensitive enough to small lesions in the image, resulting in the problem of wrong segmentation.

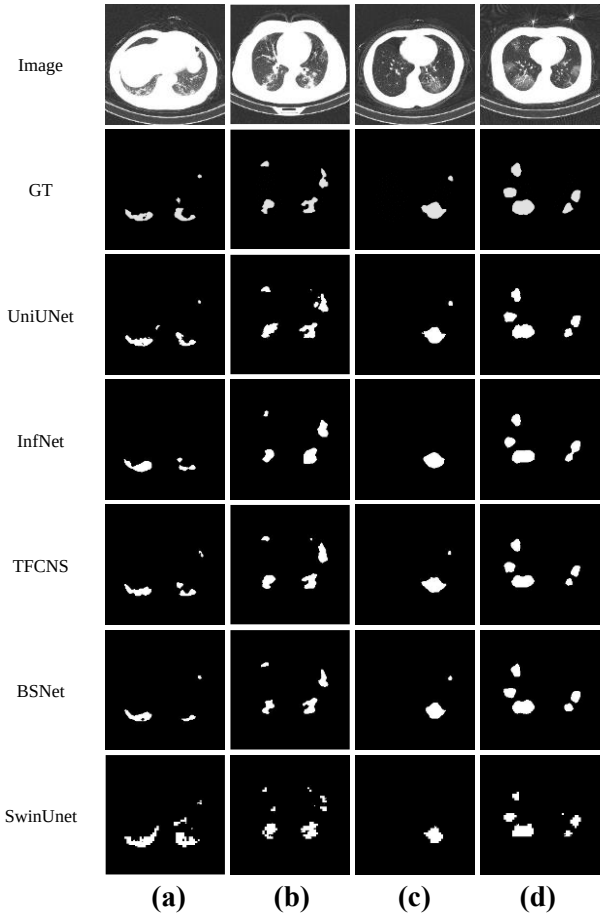


**Figure 5.** Predicted results with different methods in CC-CCII dataset.

## 5. Conclusions

Aiming at the problem that the segmentation network based on convolution neural network lacks the global dependence of modeling infected areas in COVID-19 images, which is not conducive to the complete segmentation of scattered lesion areas, a U-shaped COVID-19 image segmentation method based on UniFormer is proposed. UniFormer can establish a good correlation with the global lesion area. SSPP module can obtain multi-scale representation. The MSAG module is used to enhance valuable features in the network. The experiments illustrate that the method UniUnet in this work achieves ideal results on two COVID-19 image datasets, and a more complete

segmentation result is obtained. This method is suitable for training small datasets such as COVID-19 images, and meanwhile, it enhances the dependence of global lesion areas. However, there are still some missing and wrong segmentation phenomena in this method for some small and blurred lesion areas, because Transformer's method pays more attention to global dependence and ignores local details in the image. Therefore, it is particularly important to explore segmentation methods that can effectively retain global and local information.

**Author contributions:** Conceptualization, PG and ZT; methodology, PG and ZT; software, XC and YW; validation, ZT, PG and DZ; formal analysis, CW; investigation, ZT; resources, PG and ZT; data curation, ZT; writing—original draft preparation, ZT and PG; writing—review and editing, XW; visualization, ZT; supervision, PG; project administration, PG and XC; funding acquisition, PG. All authors have read and agreed to the published version of the manuscript.

**Conflict of interest:** The authors declare no conflict of interest.

## References

1. Geng P, Tan Z, Wang Y, et al. STCNet: Alternating CNN and improved transformer network for COVID-19 CT image segmentation. Biomedical Signal Processing and Control. 2024; 93: 106205. doi: 10.1016/j.bspc.2024.106205
2. Niitsu H, Mizumoto M, Li Y, et al. Tumor Response on Diagnostic Imaging after Proton Beam Therapy for Hepatocellular Carcinoma. Cancers. 2024; 16(2): 357. doi: 10.3390/cancers16020357
3. Shimizu S, Nakai K, Li Y, et al. Boron Neutron Capture Therapy for Recurrent Glioblastoma Multiforme: Imaging Evaluation of a Case with Long-Term Local Control and Survival. Cureus. 2023. doi: 10.7759/cureus.33898
4. Li S, Mo Y, Li Z. Automated Pneumonia Detection in Chest X-Ray Images Using Deep Learning Model. Innovations in Applied Engineering and Technology. Published online December 12, 2022: 1–6. doi: 10.62836/iaet.vli1.002
5. Bueno C, Barker MD, Orphan VJ. X-Ray Detector Physics and Applications II. Society of Photo Optical; 1993. doi: 10.1117/12.164737
6. Zheng T, Lin F, Li X, et al. Deep learning-enabled fully automated pipeline system for segmentation and classification of single-mass breast lesions using contrast-enhanced mammography: a prospective, multicentre study. eClinicalMedicine. 2023; 58: 101913. doi: 10.1016/j.eclinm.2023.101913
7. Zhang J, Chen D, Ma D, et al. CdcSegNet: Automatic COVID-19 Infection Segmentation from CT Images. IEEE Transactions on Instrumentation and Measurement. 2023; 72: 1–13. doi: 10.1109/tim.2023.3267355
8. Shi F, Wang J, Shi J, et al. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. IEEE Reviews in Biomedical Engineering. 2021; 14: 4–15. doi: 10.1109/rbme.2020.2987975
9. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV). 2016. doi: 10.1109/3dv.2016.79
10. Wang G, Liu X, Li C, et al. A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions from CT Images. IEEE Transactions on Medical Imaging. 2020; 39(8): 2653–2663. doi: 10.1109/tmi.2020.3000314
11. Yu L, Hu Z, Zhang F, et al. Unmanned aerial vehicle image biological soil crust recognition based on UNet++. International Journal of Remote Sensing. 2022; 43(7): 2660–2676. doi: 10.1080/01431161.2022.2066486
12. Wang B, Jin S, Yan Q, et al. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system. Applied Soft Computing. 2021; 98: 106897. doi: 10.1016/j.asoc.2020.106897
13. Cong R, Zhang Y, Yang N, et al. Boundary Guided Semantic Learning for Real-Time COVID-19 Lung Infection Segmentation System. IEEE Transactions on Consumer Electronics. 2022; 68(4): 376–386. doi: 10.1109/tce.2022.3205376
14. Ibtehaz N, Kihara D. Acc-unet: a completely convolutional unet model for the 2020s. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. 2023: 692-702. doi: 10.48550/arXiv.2308.13680

15. Zhou T, Canu S, Ruan S. An automatic COVID-19 CT segmentation network using spatial and channel attention mechanism. ARXIV preprint arXiv:2004.06673, 2020.
16. Li CF, Xu YD, Ding XH, et al. MultiR-Net: A Novel Joint Learning Network for COVID-19 segmentation and classification. Computers in Biology and Medicine. 2022; 144: 105340. doi: 10.1016/j.compbiomed.2022.105340
17. Xiao H, Ran Z, Mabu S, et al. SAUNet++: an automatic segmentation model of COVID-19 lesion from CT slices. The Visual Computer. 2022; 39(6): 2291–2304. doi: 10.1007/s00371-022-02414-4
18. Zhao S, Li Z, Chen Y, et al. SCOAT-Net: A novel network for segmenting COVID-19 lung opacification from CT images. Pattern Recognition. 2021; 119: 108109. doi: 10.1016/j.patcog.2021.108109
19. Jia W, Ma S, Geng P, et al. DT-Net: Joint Dual-Input Transformer and CNN for Retinal Vessel Segmentation. Computers, Materials & Continua. 2023; 76(3): 3393–3411. doi: 10.32604/cmc.2023.040091
20. Karlinsky L, Michaeli T, Nishino K, et al. Computer Vision – ECCV 2022 Workshops. Springer Nature Switzerland; 2023. doi: 10.1007/978-3-031-25066-8
21. Li K, Wang Y, Gao P, et al. Uniformer: unified transformer for efficient spatiotemporal representation learning. ARXIV preprint arXiv:2201.04676. 2022.
22. Bello IM, Zhang K, Su Y, et al. Densely multiscale framework for segmentation of high resolution remote sensing imagery. Computers & Geosciences. 2022; 167: 105196. doi: 10.1016/j.cageo.2022.105196
23. Azad R, Heidari M, Shariatnia M, et al. Transdeeplab: convolution-free transformer-based deeplab v3+ for medical image segmentation. Proceeding of the International Workshop on Predictive Intelligence in Medicine. 2022: 91-102. doi: 10.48550/arXiv.2208.00713
24. Tang F, Wang L, Ning C, et al. CMU-Net: A Strong ConvMixer-based Medical Ultrasound Image Segmentation Network. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). 2023. doi: 10.1109/isbi53787.2023.10230609
25. Zhang K, Liu X, Shen J, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. Cell. 2020; 181(6): 1423–1433.e11. doi: 10.1016/j.cell.2020.04.045
26. Ma J, Wang Y, An X, et al. Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. Medical Physics. 2021; 48(3): 1197–1210. doi: 10.1002/mp.14676
27. Liu J, Zhao D, Shen J, et al. HRD-Net: High resolution segmentation network with adaptive learning ability of retinal vessel features. Computers in Biology and Medicine. 2024; 173: 108295. doi: 10.1016/j.compbiomed.2024.108295
28. Geng P, Lu J, Zhang Y, et al. TC-Fuse: A Transformers Fusing CNNs Network for Medical Image Segmentation. Computer Modeling in Engineering & Sciences. 2023; 137(2): 2001–2023. doi: 10.32604/cmes.2023.027127
29. Chen L, Bentley P, Mori K, et al. DRINet for Medical Image Segmentation. IEEE Transactions on Medical Imaging. 2018; 37(11): 2453–2462. doi: 10.1109/tmi.2018.2835303
30. Wang R, Lei T, Cui R, et al. Medical image segmentation using deep learning: A survey. IET Image Processing. 2022; 16(5): 1243–1267. doi: 10.1049/ipr2.12419
31. Fan DP, Zhou T, Ji GP, et al. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. IEEE Transactions on Medical Imaging. 2020; 39(8): 2626–2637. doi: 10.1109/tmi.2020.2996645
32. Cao H, Wang Y, Chen J, et al. Swin-unet: unet-like pure transformer for medical image segmentation. In: Proceedings of the European Conference on Computer Vision; 2022. doi: 10.48550/arXiv.2105.05537