ITC 4/53
Information Technology and Control
Vol. 53 / No. 4/ 2024
pp. 1188-1203
DOI 10.5755/j01.itc.53.4.35745

MS-UNet: A Novel Multi-scale U-shaped Network for COVID-19 CT Image Segmentation

Received 2023/11/28 | Accepted after revision 2024/04/17

HOW TO CITE: Liu, S., Si, F., Tang, X., Cai, T. (2024). MS-UNet: A Novel Multi-scale U-shaped Network for COVID-19 CT Image Segmentation. *Information Technology and Control*, 53(4), 1188-1203. https://doi.org/10.5755/j01.itc.53.4.35745

MS-UNet: A Novel Multi-scale U-shaped Network for COVID-19 CT Image Segmentation

Shangwang Liu, Feiyan Si, Xiufang Tang , Tongbo Cai

School of Computer and Information Engineering, Henan Normal University, Xinxiang, 453000, China;
e-mails: shwl2012@hotmail.com; 2208183020@stu.htu.edu.cn; tang15103739221@163.com;
caitongbo@outlook.com

Corresponding author: shwl2012@hotmail.com

The U-Net network has its own powerful capabilities in medical image segmentation tasks, yet still it is a challenging task to make U-Net accurately segment the infected lesions of COVID-19 CT images because these lesion areas are usually irregular in shape, various in size, and blurry in boundaries. In this paper, a novel multi-scale U-shaped network based on U-Net for accurate segmentation of lesion regions in COVID-19 CT images is proposed. First, we generate two auxiliary scale features($f_i^{0.5}$, $f_i^{1.5}$) based on the main scale feature ($f_i^{1.0}$) through zoom strategy. Secondly, we design the Scale Integration Module (SIM), which is capable of filtering and aggregating scale-specific features and can fully exploit multi-scale semantic information. Thirdly, the hierarchical mixed module (HMM) has successfully substituted for the down-up aggregation process of the U-Net network, which further enhances the mixed scale features. On the dataset COVID-19-CT829, compared with the recent COVID-19 segmentation model, hiformer, the Dice, Sen and F-measure of our network have increased by 2.24%, 2.83%, 3.14%, respectively; on the dataset COVID-19-CT100, the Dice, Sen and F-measure of our network have increased by 2.91%, 3.72%, 2.42%, respectively. Moreover, we have validated the generalizability and portability of our network on other medical datasets (Polyp segmentation dataset: CVC-612 and kvasir), and our network has also achieved superior results of COVID-19 CT image segmentation.

KEYWORDS: COVID-19 CT Image Segmentation, U-Net, zoom strategy, Scale Integration Module, hierarchical mixed module.

1. Introduction

Since the outbreak of COVID-19(corona virus disease 2019), it has spread rapidly around the world, causing millions of casalties and huge economic losses, and posing a serious threat to human life safety. RT-PCR[42](reverse transcription polymerase chain reaction) is now the main method for screening COVID-19 cases, but this method has some drawbacks, such as insufficient detection reagents, long detection time and low sensitivity [16, 46]. In order to further accelerate the detection speed and reduce the cost, the automatic screening of COVID-19 by using computer tomography (CT) is emerged [3, 44], because the CT images of the patient lungs with COVID-19 have obvious lesion characteristics indeed, which is expected to quickly and accurately segment the lesion sites from the CT images of the lungs of patients [2, 17, 49]. Hence, it is of great significance to rapidly diagnose and monitor COVID-19 patients with the assistance of CT images [9, 19, 22, 40]. In the study of COVID-19, it was observed that the lesions with consolidation characteristics would accumulate with infection. Therefore, measuring the change trend of the lesion area might help doctors make better treatment decisions.

However, relying solely on experts to evaluate the infection in CT scan images may be tiring and prone to interpretation errors. Quantifying lung injury caused by COVID-19 infection is a challenging task, especially for doctors who need to diagnose multiple CT images on the same day. Another important point is that experts must quantify the results before and after treatment to analyze whether the treatment is effective. This is undoubtedly a huge work. Therefore, it is a necessary task to use computer-aided diagnostic tools to automatically segment the CT image of COVID-19, both for the segmentation of the infectious findings of COVID-19 and for the volumetric quantification of the lesions.

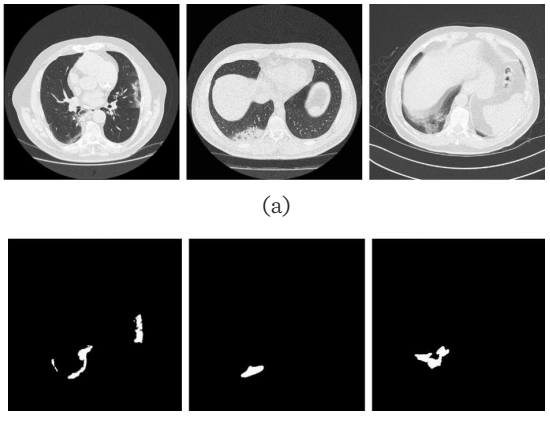
Nowadays, deep learning method is becoming a hot topic in in the field of medical imaging and has shown promising results [12, 29, 31, 47]. Fan et al. [15] proposed Inf-Net to segment the infected area of novel coronavirus, and proposed a semi supervised training method to solve the problem of insufficient labeled CT numbers and improved the segmentation performance. Yazdani et al. [48] introduced a model with residual connection and attention awareness unit, which could be used to discover the relationship of patients with and without COVID-19. Amyar et al. [4] put forward a multi-task deep learning model to jointly identify covid-19 patients and part of the lesion areas of covid-19 in their chest CT images. Gunraj et al. [18] built an enhanced deep neural network, which can detect COVID-19 from chest CT images by using various training strategies. Islam et al. [23] focused on the identification of brain tumors using MRI images using a federated learning (FL) and convolutional neural network (CNN) integrated architecture. The study explored the efficiency and accuracy of brain tumor detection through a distributed learning approach while maintaining data privacy. Połap et al. [34] proposed a Bilinear Pooling (BP) method that includes a toxicity detection module, by introducing toxicity detection, the model is able to identify and resist potential data contamination, protect the model from malicious attacks, and ensure the accuracy and reliability of data analysis.

U-Net is currently the most widespread image segmentation architecture, and it has been successful in many medical image domains due to its flexibility and optimized modular design [5]. Diniz et al. [13] improved the traditional U-Net, including batch normalization, leaky ReLU, dropout, and residual block techniques. Their method, which can automatically segment infections caused by COVID-19, is expected to be a tool to help medical professionals fight COVID-19. Lizzi et al. [30] quantified the accumulation of lung lesions in COVID-19 pneumonia by cascading two U-nets, where the first U-net was used to identify lung parenchyma, and the second U-net calculates the areas affected by COVID-19 lesions. In the initial stage of COVID-19 segmentation, the UNet network is usually unable to distinguish between healthy regions and diseased regions in healthy lungs. To cope with this, Shamim et al. [37] increase the set of weights that shrink and expand the UNet path, and add a modified convolutional module to create a connection between the encoder and decoder pipelines. But, the accuracy of semantic COVID-19 CT image segmentation of methods mentioned above is not good enough in practice to diagnose and treat patients with COVID-19 [43, 50], and it is still a challenging

task to accurately segment COVID-19 CT images, because COVID-19 infected lesions have the following characteristics (see Figure 1):

1 As far as the COVID-19 lesions in chest CT images are concerned, they often have irregular shape, various size, different location and blurry boundary, and it is hard to segment the entire lesion areas for the existed methods in computer vision fields.
2 There is no obvious difference between infected tissue and normal tissue because the target area is highly similar to the background area, which easily results in false negatives and false positives during CT Lung image segmentation.

Figure 1
Examples of the chest CT images (a) and its labeled lesion, (b) it can be seen that the lesion area in different CT Lung images changes greatly and the boundary is very unclear



As a human being, aiming at accurately finding the objects in a blurred scene, people often reference and compare the shape or appearance changes of the image at different scales by enlarging and reducing the size of the image. Inspired by this human behavior, we built and designed a novel multi-scale U-shaped network by scale-space strategies to identify more accurately COVID-19 lesion regions in CT Lung images. Scale space theory is an effective theoretical framework for promoting the understanding of image structure, and its ideas are widely used in the field of computer vision, such as image pyramids [1] and feature pyramids [26]. The existing inverted pyramid multi-scale network [45] often makes the feature representation lose texture and boundary details, which is not conducive to COVID-19 lung CT image segmentation. Therefore, we leverage the relationship between foregrounds and backgrounds at multiple scales, to fully perceive not only COVID-19 lesion regions but also normal tissues. Moreover, we also explore the fine-grained feature scale space among channels.

The multi-scale U-shaped network we designed is briefly introduced as follows:

1 We employ triple feature encoders to extract features at different scales and provide them to the scale merging layer for subsequent processing.
2 we put forward a scale integration module (SIM) based on the attention perception mechanism, which can screen out useful features in the auxiliary scale and integrate them into the main scale.
3 the hierarchical mixed module (HMM) we built increases the receptive field range and diversifies the feature representation within our module, gradually integrating multi-level features in a top-down manner.

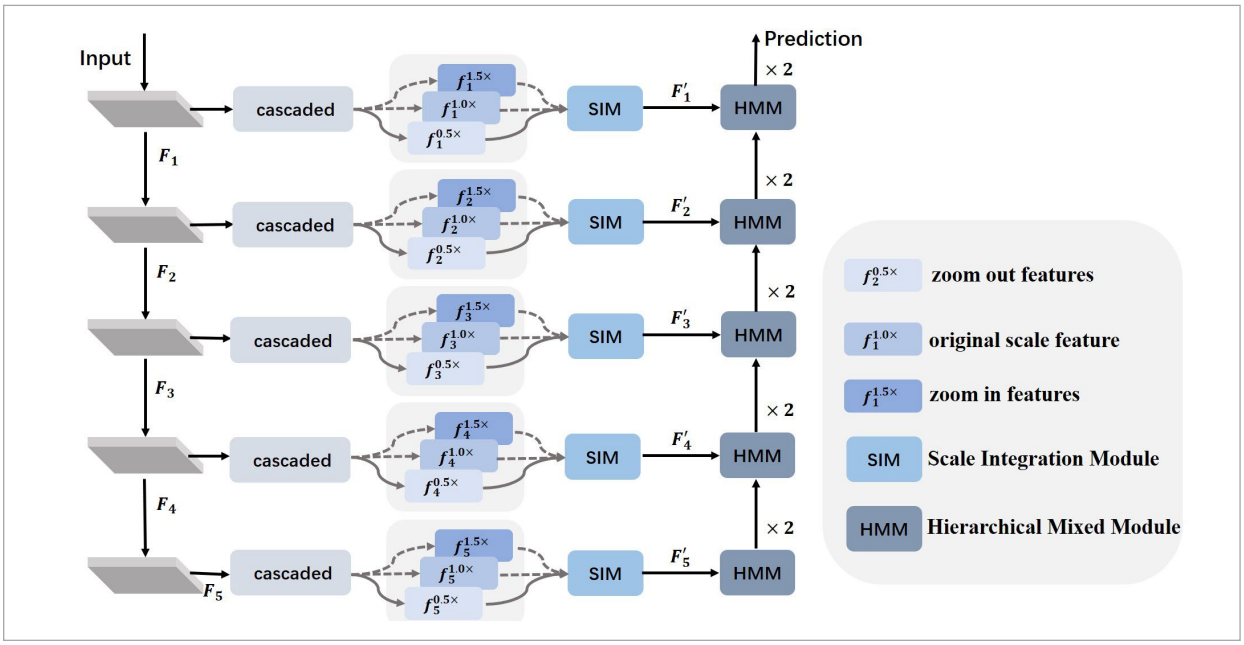
Thus, our method can capture not only fine-grained but also mixed-scale clues, and achieve more accurate COVID-19 CT Lung image segmentation.

2. Materials and Methods

The structure of our COVID-19 image segmentation network is shown in Figure 2.

In Figure 2, according to scale space theory [27, 33], a popular idea in the computer vision field, we build the U-shaped network with multi-scale structure for segmentation of COVID-19 lesions, and its network structure is as following: we adopt the first half structure of U-Net for extracting multi-level features $F_i(i = 1 \ldots 5)$ at first; Secondly, the multi-level features (F_i) are cascaded to obtain feature maps of three scales $f_i^{0.5}, f_i^{1.0}, f_i^{1.5}$. Subsequently, we filter the key semantic information in different scales via an attention-based scale integration module (SIM), which greatly enhances the detection effect. Finally, we propose a down-up hierarchical mixed module (HMM) to fuse multi-layer features. This down-up fusion structure, as well as the top-down structure in the feature extraction process, constitute our novel U-shaped

Figure 2
Overall framework



structure network, which is expected to capture both fine-grained and mixed-scale features for COVID-19 CT Lung image segmentation.

2.1. Scale Integration Module

The proposed scale integration module (SIM) can weight and combine specific information of different scales. Specifically, SIM can self adaptively highlight expressions of different scales via filtering and aggregation (see Figure 3).

As shown in Figure 3, before scale integration, it is necessary to adjust the size of the features $f_i^{1.5}$ and $f_i^{0.5}$ to make their resolution consistent with the main scale features $f_i^{1.0}$. That is to say, for $F_i^{1.5}$, we employ the "Max Pooling and Avg Pooling" hybrid structure for down sampling, in order to retain the effectiveness and diversity of COVID-19 infection region information in high-resolution features. For $f_i^{0.5}$, we adopt "Bilinear" between two "Conv+BN+ReLU" modules to directly up sampling. Thus, these features are input into the "generator", and three channel feature map is computed by Equation (1) through a series of convolution layers.

$$A_i^{0.5}, A_i^{1.0}, A_i^{1.5} = softmax\left(\tau\left(\left[\mu\left(f_i^{0.5}\right), f_i^{1.0}, \varphi\left(f_i^{1.5}\right)\right], \emptyset\right)\right), \quad (1)$$

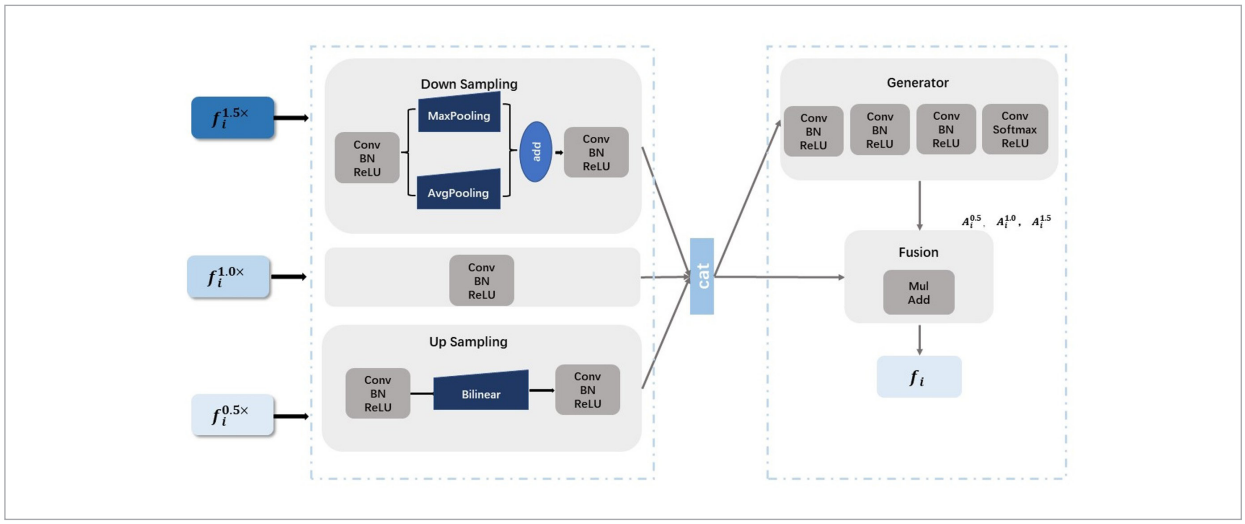
where μ and φ represent bilinear interpolation and mixed pooling (Max Pooling and Avg Pooling) operations, respectively;$[\mu(f_i^{0.5}), f_i^{1.0}, \varphi(f_i^{1.5})]$ implies the concatenation operation. τ(*, ∅) denotes the "Conv+BN+ReLU" layer stacked in the generator, and ∅ means the parameters of these layers. After softmax activation layer, the feature map of each scale $(A_i^{0.5}, A_i^{1.0}, A_i^{1.5})$ is finally calculated.

Let $A_i^{0.5}, A_i^{1.0}, A_i^{1.5}$ represent those weights to get the final output f_i, and it can be calculated by Equation (2).

$$f_i = A_i^{0.5} \cdot \mu\left(f_i^{0.5}\right) + A_i^{1.0} \cdot f_i^{1.0} + A_i^{1.5} \cdot \sigma\left(f_i^{1.5}\right) \quad (2)$$

Please note that some operations before and after the sampling operation are not shown in equations 1 and 2 for simplicity, but can be found them in Figure 2. Bilinear interpolation is a fine-grained image scaling technique that is achieved by computing a weighted average of the pixel values of the four immediately neighboring pixels. The goal of hybrid pooling is to meticulously extract image features to ensure that salient details are captured, while also giving due consideration to more subtle variations. In summary, by selectively aggregating scale-specific information and exploring subtle and critical semantic cues across scales, we aimed to construct a powerful feature rep-

Figure 3
Illustration of the Scale Integration Module



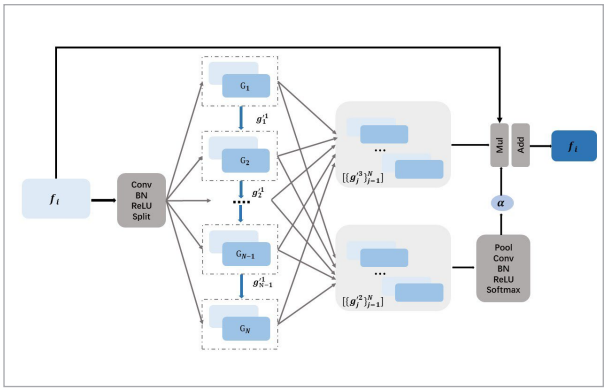
resentation that effectively responds to the challenges posed by irregular and diverse lesion shapes. This nuanced approach makes our model very sensitive to a large number of lesion features, thus improving diagnostic accuracy and reliability.

2.2. Hierarchical Mixed Module

After SIM, the auxiliary scale information ($f_i^{1.5}$ *and* $f_i^{0.5}$) is integrated with the main scale ($f_i^{1.0}$). During this procedure, different channels still contain different semantics, and it is necessary to mine these valuable clues contained in the different channels. To this end, we put forward the hierarchical mixed module (HMM) to carry out information interaction and feature refinement among channels. HMM further increases the range of receiving domain and diversifies the feature representation in the module. Obviously, capturing fine-grained features and mixing different scale information can enable our model to segment covid-19 infected areas more accurately. The structure of HMM is shown in Figure 4

We expand the number of channels of the feature map by 1*1 convolution, and divide the channels into N groups ($G_{j\ j=1}^{N}$)according to the channel dimension, then feature interaction of each group is implemented in an iterative manner. Specifically, the first group G_1 is split into three sets ($g_1^{'m^3}{}_{m=1}$)after the convolution block. Among them, $g_1^{'1}$ is utilized for information exchange with the next group of features, and $g_1^{'2}$

Figure 4
Illustration of the Hierarchical Mixed Module



and $g_1^{'3}$ are regarded as channel modulation. In group $j(1 < J < N)$, the feature G_j is connected with the previous group feature $G_{j-1}^{'1}$, and the feature group is thus divided into three feature sets like first group G_1. Such an iterative mixing strategy is good at learning the critical clues from different channels and obtaining a powerful feature representation.

In our proposed hierarchical hybrid module (HMM), we adopted an iterative hybrid strategy to enhance the integration of multi-scale features, which not only improves the ability of the model to identify fine textures and macro-anatomical structures, but also enhances the differentiation of features. Using this strategy, the model is able to complement each other

with key clues learned from different channels, so as to capture key information in the segmentation task. The implementation of this hybrid strategy ensures a comprehensive and accurate feature characterization and is essential to improve the accuracy and reliability of medical image segmentation.

2.3. Loss Function

Binary cross-entropy loss (BCEL) is widely used in image segmentation tasks. The binary cross-entropy loss (BCEL) function calculates the difference between actual and predicted labels. The calculation method is provided as follows:

$$L_{BCEL}^{i,j} = -G_{i,j}P_{i,j} - \left(1 - G_{i,j}\right)log\left(1 - P_{i,j}\right) \quad (3)$$

where G_{ij} and P_{ij} denote the ground truth and the predicted value at $position(i,j)$, respectively.

Because the COVID-19 infected lesion is complex and fuzzy, just BCEL function training cannot accurately identify the infected lesion. Hence, we also employ an uncertainty-aware loss (UAL) to assist with training. The key to the uncertainty-aware loss (UAL) is that it allows the model to self-adjust its predictions for insufficient data or more noise, reduce overconfident predictions, and may improve the robustness of the model in the face of unknown or uncertain situations. The formula is as follows:

$$L_{UAL}^{i,j} = 1 - \left|2P_{i,j} - 1\right|^2. \quad (4)$$

Thus, the total loss can be calculated by Equation (5).

$$L = L_{BCEL} + \gamma L_{UAL}, \quad (5)$$

where γ is the balance coefficient, and the adjustment strategy is an increasing cosine strategy.

3. Implementation Details

3.1. Dataset

We conducted experiments on two sub datasets of COVID19 CT segmentation [24]: COVID-19-CT100 and COVID-19-CT829. The former consists of 100 labeled CT images, where 50 images are randomly chosen for training and the rest of 50 images for testing. The latter is the first open-access COVID-19 dataset and contains 829 CT images, where 709 images are randomly chosen for training and the remained 120 images for testing. All the CT images are from more than 40 COVID-19 patients and collected by the Italian Society of Medical and Interventional Radiology. A radiologist segmented the CT images manually by using three labels for identifying lung infections [21]: ground-glass, consolidation and pleural effusion.

3.2. Training Parameters

Our model is implemented under the PyTorch frame on the operation system of Ubuntu V20.04 distribution. The hardware environment is as follows: CPU, Intel E5-2637; GPU, NVIDIA 3090ti(24G). We built our network by using the PyTorch framework, where adam optimization [28] was employed for training, and the weight value was attenuated to 1e-4. While, we employ the learning rate strategy poly, and the initial learning rate is 1e-3. Thus, we adopted the above strategy to train 100 epochs on the training set to get the final result, and set the batch size to 8. Moreover, we used the same strategy to train other segmentation networks to ensure the comparison fairly.

3.3. Evaluation Metrics

The evaluation metrics of Dice similarity coefficient (Dice), Sensitivity (Sen), Specificity (Spec), F-measure, mIOU, Mean Absolute Error (MAE) are employed. Specifically, Dice, Sen, Spec, F-measure and mIOU range from 0 to 1; and the larger these values, the better the model. By contrary, a lower value of MAE indicates better segmentation accuracy.

Dice, Sen, Spec and mIOU are usually adopted in image segmentation, and they can be calculated by Equations (6)-(9), respectively.

$$Dice = \frac{2TP}{FP+2TP+FN}, \quad (6)$$

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

$$Spec = \frac{TN}{FP+TN}, \quad (8)$$

$$mIOU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{FN+FP+TP}, \quad (9)$$

where TP represents the area that is predicted to be a positive sample, and it is actually a positive sample; FP denotes the part that is predicted to be a positive sample, but it is actually a negative sample; TN means

the area that is predicted to be a negative sample, and it is actually a negative sample; FN implicates a predicted negative sample, yet it is actually a positive sample; k refers to the category.

Hausdorff distance (HD) is a measure describing the degree of similarity between two sets of points, and it is a defined form of distance between the two sets of points. The smaller the value of Hausdorff distance, the higher the similarity of the two sets. Suppose there are two sets $A = a_1, ..., a_p$ and $B = b_1, ..., b_q$, then the Hausdorff distance between these two sets of points can be defined as follows.

$$H(A,B) = max\big(h(A,B), h(B,A)\big), \tag{10}$$

$$h(A,B) = max_{a\in A}\{min_{b\in B}\big||a-b|\big|\}, \tag{11}$$

$$h(B,A) = max_{b\in B}\{min_{a\in A}\big||b-a|\big|\}, \tag{12}$$

where $||*||$ is the distance normal form between point sets A and B.

Mean Absolute Error (MAE) can measure the error between prediction maps and ground truth maps at the pixel level, and it is computed from Equation (11).

$$MAE = \frac{1}{w\times h}\sum_x^w\sum_y^h\big|M_f(x,y), -(x,y)\big|. \tag{13}$$

In the evaluation of the neural network architecture, "FLOPs" (floating point operations per second) and "Params" (parameters) are the key indicators of the model performance and complexity. Table 1 presents a comparative analysis of these metrics in the different models. FLOPs represent computational workloads, representing the total number of floating-point operations required by an algorithm or model, thus providing a quantitative measure of algorithmic complexity and inference time requirements. The Parameter value quantifies the total number of trainable parameters of the model, measured in megabytes (M).

Table 1
Performance evaluation results of the partial

models	Flops	Params
U-Net	47.65 G	13.40 M
DeepLabV3	28.57 G	58.16 M
ResUNet	104.46 G	35.92 M
Attention-UNet	102.03 G	34.88 M
TransUNet	67.14 G	110.37 M
SegNet	30.73 G	29.44 M
MTUNet	44.78 G	75.07 M
Ours	23.83 G	13.40 M

4. Results and Discussion

4.1. Comparative Experiment

To fully evaluate our network in segmentation of COVID-19 CT images, we conduct experiments on two datasets (COVID-19-CT100 and COVID-19-CT8[26]), and compare it with the state-of-the-art segmentation networks such as PraNet [14], SwinUNet [8], SegNet [6], Deeplabv3 [11], PSPNet [51], TransUNet [10], U-Net [35], U-Net++ [52], Inf-Net [15], Attention-UNet [32], hiformer [20], DM2T-Net [41] and C2FVL [49], etc. In order to compare the performance of these networks fairly, all of them are trained with the same parameter optimization strategy. All networks use the same initial parameter settings for fair comparison, and the above optimization strategies are used to automatically update the weights during the training process of different networks. The numerical evaluation results of related methods are listed in Table 1.

From Table 2, we can see that our network consistently achieves the best or nearly the best performance in terms of all metrics on the two datasets. On dataset of COVID-19-CT829, our network is slightly worse than U-Net network in mIOU, but it, in contrast, confirm the strong performance of U-shaped network in image segmentation. This is exactly the reason that our multi-scale network based on the U-shaped network performs very well. On the dataset of COVID-19-CT829, compared with the recent COVID-19 segmentation model, hiformer, the Dice similarity coefficient of our network has increased by 2.24%, and the sensitivity (Sen) has raised by 2.83%, and the other evaluation indicator F-measure has improved by 3.14%; Meanwhile, the mean absolute error (MAE) has decreased by 0.25%. On the dataset of COVID-19-CT100, compared with hiformer, the Dice similarity coefficient has increased by 2.91%, the sensitivity (Sen) has raised by 3.72%, and F-measure has improved by 2.42%.

Table 2
Quantitative evaluation results of infection area segmentation on COVID-19-CT100 and COVID-19-CT829

Methods		Dice (%)	Sen (%)	Spec(%)	HD	mIOU(%)	MAE (%)
COVID-19-CT100	PraNet [14]	63.42	68.25	91.97	48.59	72.07	9.92
	SwinUNet [8]	70.86	71.76	94.69	45.09	76.98	7.53
	SegNet [6]	69.76	69.93	94.95	40.12	76.53	7.61
	Deeplabv3 [11]	67.59	71.38	93.35	40.78	73.79	8.98
	PSPNet [51]	68.19	69.77	94.57	44.92	75.70	7.98
	TransUNet [10]	75.16	78.10	95.58	38.68	80.31	6.42
	U-Net [35]	74.06	74.56	96.31	37.43	80.37	6.78
	U-Net++ [52]	75.58	75.97	96.26	34.66	80.15	6.16
	Inf-Net [15]	66.38	69.72	92.94	37.14	73.39	9.05
	Attention-Unet [32]	74.89	72.43	96.80	36.53	80.17	6.04
	Hiformer [20]	76.47	76.44	96.54	39.79	81.06	5.82
	DM2Tnet [41]	76.73	75.95	96.34	40.50	78.42	7.02
	C2FVL [49]	78.17	79.60	**97.03**	35.93	80.13	6.29
	Ours	**79.38**	**80.12**	96.62	**33.43**	**81.52**	**5.66**
COVID-19-CT829	PraNet [14]	78.84	83.34	99.04	18.96	83.09	1.41
	SwinUNet [8]	79.02	84.30	99.11	21.55	83.65	1.34
	SegNet [6]	80.53	86.19	99.10	19.97	84.58	1.27
	Deeplabv3 [11]	80.66	82.88	99.26	20.21	85.42	1.16
	PSPNet [51]	80.44	84.91	99.16	21.42	84.01	1.23
	TransUNet [10]	82.70	84.68	99.38	19.75	86.49	1.05
	U-Net [35]	83.73	86.68	99.34	18.31	87.83	0.96
	U-Net++ [52]	83.20	88.54	99.22	17.99	86.54	1.08
	Inf-Net [15]	83.06	85.19	99.30	15.71	86.60	1.06
	Attention-Unet [32]	84.01	86.83	99.38	18.65	87.45	0.97
	Hiformer [20]	83.84	86.65	99.37	16.24	87.17	1.00
	DM2Tnet [41]	83.91	89.55	99.32	17.22	86.95	0.92
	C2FVL [49]	85.06	**90.02**	99.15	15.37	**88.60**	0.88
	Ours	**86.25**	89.48	**99.70**	**14.22**	87.05	**0.75**

To further evaluate the superiority of our multi-scale U-shaped network (MS-UNet), some qualitative comparison (Visual comparison) experiments of related image segmentation methods were conducted, as shown in Figures 5-6.

In Figures 5-6, we can easily find out the superiority of our network. On the data set COVID-19-CT100, we can see that the image segmentation results of our network are more refined and its texture is also clearer, especially the results of these red boxes. Looking carefully at Figure 5 (Visual comparison of infection area segmentation result on dataset COVID-19-CT829), our network MS-UNet can perfectly segment some small infection areas (Rows 3, 4 and 5), which is almost impossible for other networks to achieve. Our image segmentation results have clearer, more complete object regions and

Figure 5
Visual comparison of infection area segmentation result on dataset COVID-19-CT100

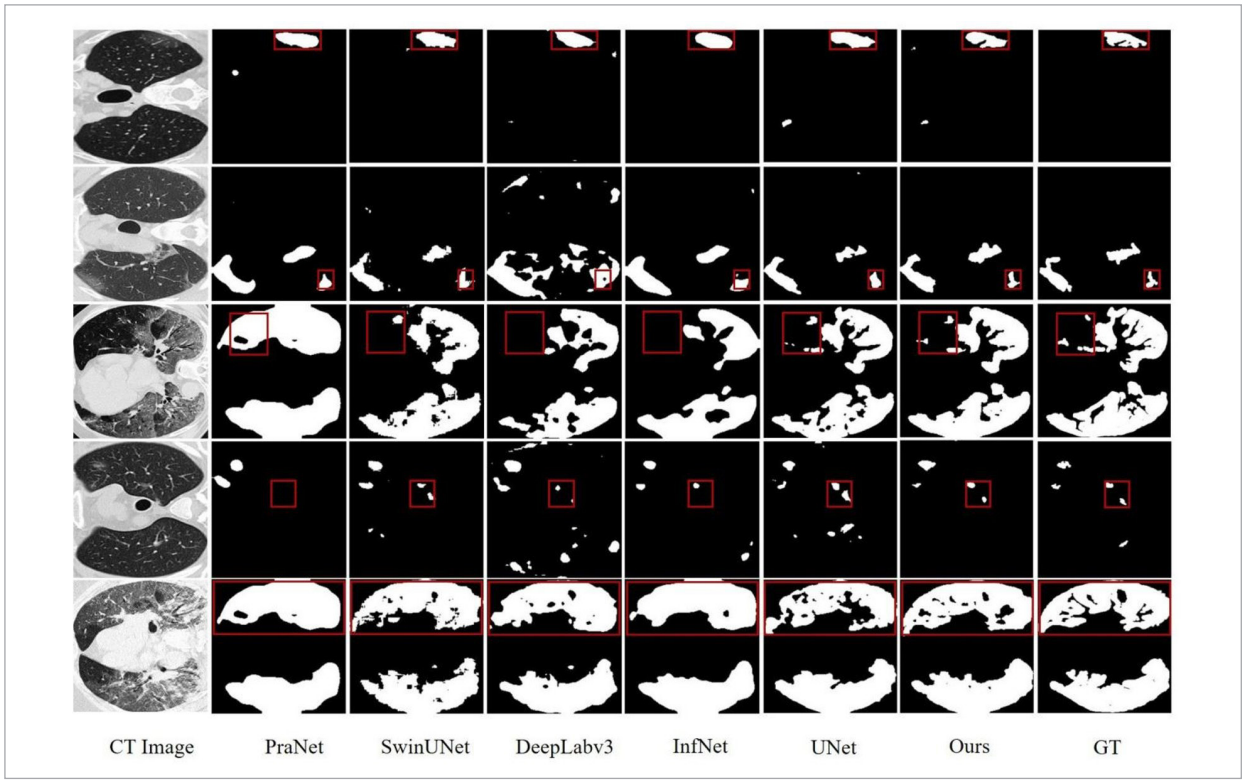


Figure 6
Visual comparison of infection area segmentation result on dataset COVID-19-CT829

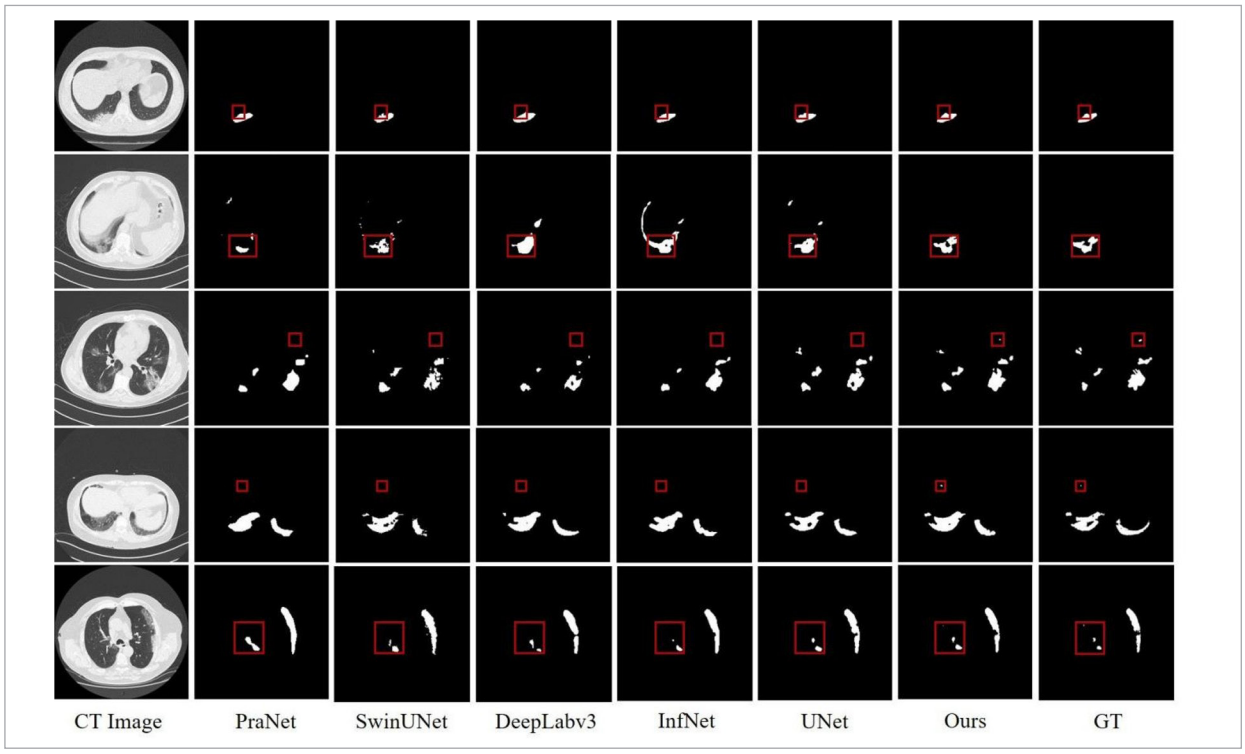
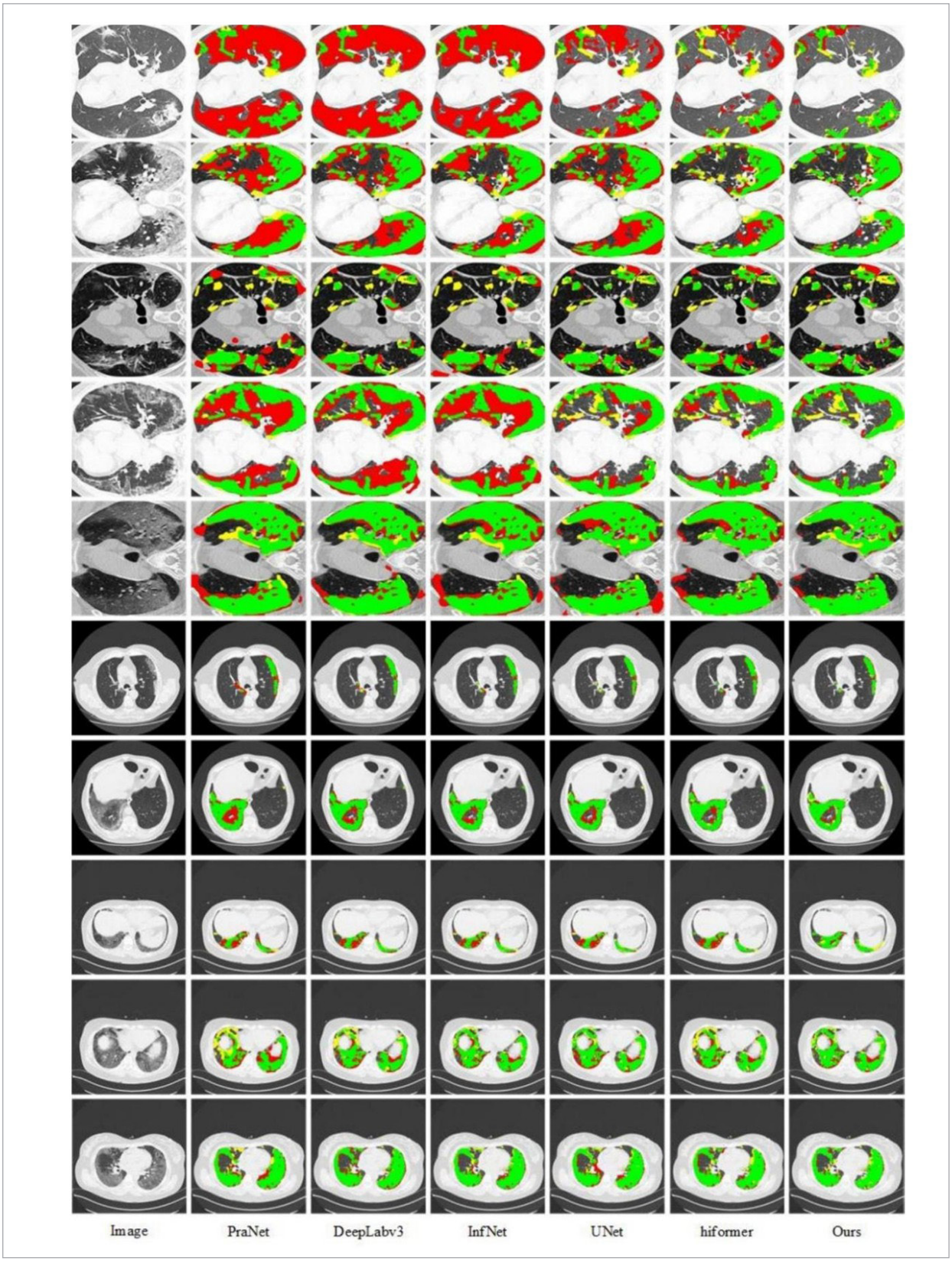


Figure 7

Visual comparison results: green represents true positive, red represents false negative and yellow represents false positive



sharper contours, all of which result from the ability of our network to capture fine-grained and mixed scale cues by using a zoom strategy. To more intuitively illustrate the difference among the segmentation results of each model, another kind of visual comparison experiments were conducted, as shown in Figure 7.

In Figure 7, green represents correctly predicted areas, red denotes missing areas, and yellow implies incorrectly predicted areas. We observed that the results of PraNet, DeepLabv3 and InfNet have more red areas, which indicates that the three models have a higher probability of omission; the results of UNet and hiformer have more yellow areas, which means that the two models have a higher error probability when predicting image pixels that do not belong to this category. For our MS-UNet network, it can obtain the maximum number of correct pixel prediction regions, and reduce the error prediction probability of adjacent regions as well, and predict correctly the negative samples to a great extent too. The reason is that we adopt the zoom strategy to extract more abundant features from the three scales, and the proposed scale integration module (SIM) can effectively fuse them together, both of which can lead to a sound COVID-19 CT image segmentation with exact background border area and pathological texture.

The PR (Precision-Recall) and F-measure curves of the related 12 networks are shown in Figures 8-9, where the red bold lines represent the results of our network. P means precision, and R represents recall. It is not difficult to see that the PR and F-measure curves of our network are all at the top, which illustrate that our network performs better than others for COVID-19 CT image segmentation. Please note that F-Measure is a weighted harmonic average of precision and recall, which can be calculated as follows.

$$F - Measure = \frac{\{(a)^2+1)*P*R}{a^2*P*R}, \quad (14)$$

where a is the weight, and P and R represent precision and recall, respectively.

4.2. Ablation Experiment

We also conducted an ablation experiment to prove the effectiveness of each module in our network. In the ablation experiment, the structure diagram of each part is shown in Figure 10, and the experimental results are listed in Table 2. The Backbone in Figure 10 is UNet,

Figure 8

Precision-Recall curve for datasets (a) COVID-19-CT829 and (b) COVID-19-CT100

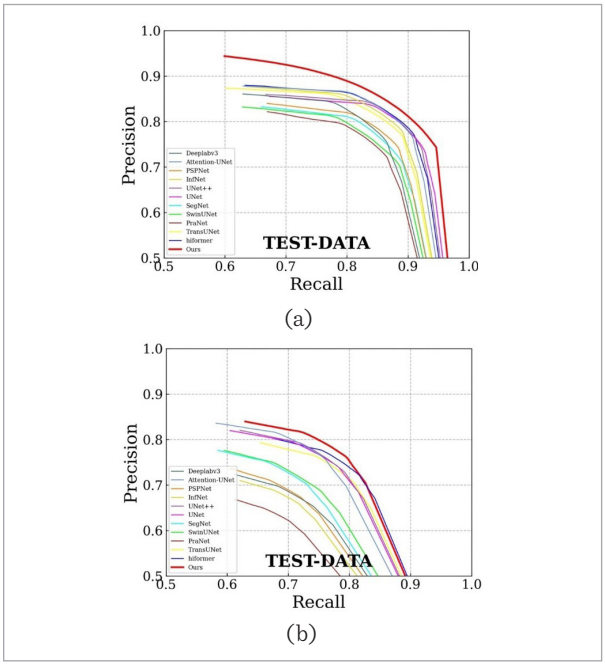


Figure 9

F-measure curve for datasets (a) COVID-19-CT829 and (b) COVID-19-CT100

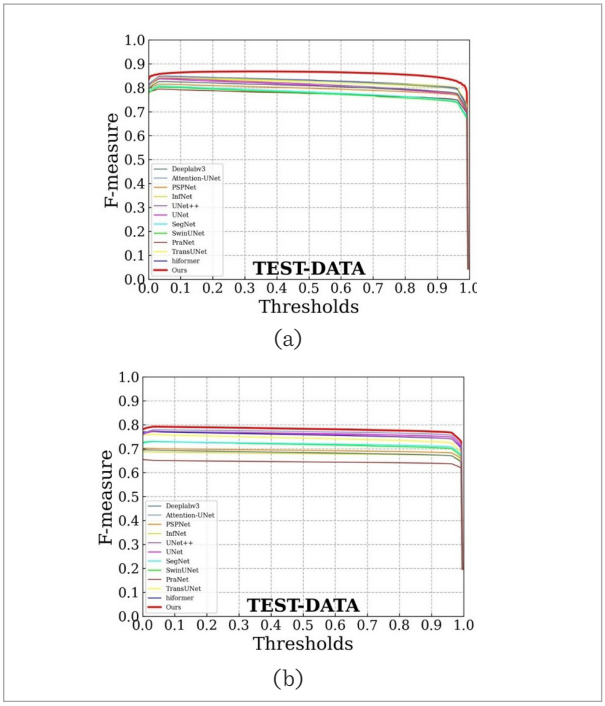
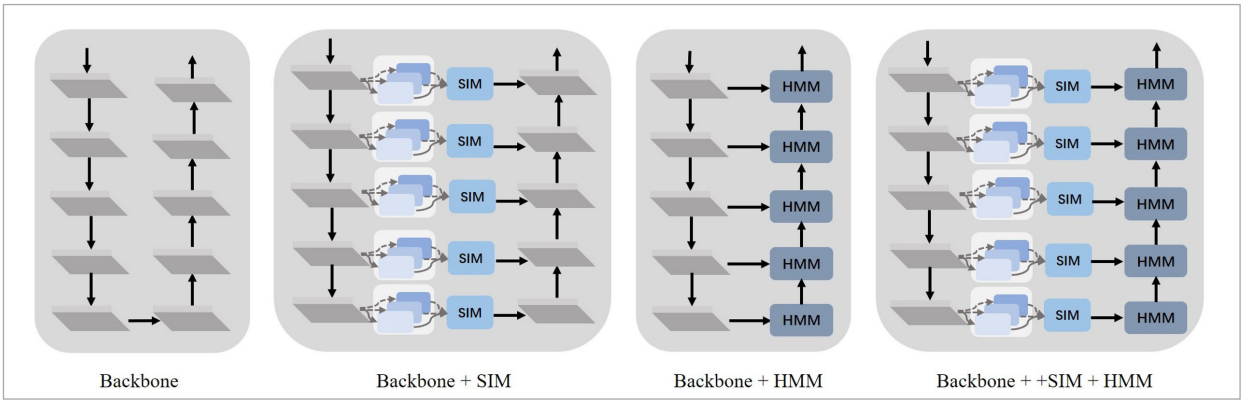


Figure 10
The structure diagram of each part in the ablation experiment



corresponding to row 1 in Table 2. The results of using the zoom strategy and fusing the scale integration module (HMM) in the UNet network correspond to the results in the second row of Table 2. After replacing the upward fusion process of the UNet network with the hierarchical mixed module (HMM) proposed, the results correspond to the data in row 3 of Table 2. The data in row 4 of Table 2 are the results of the complete MS-UNet model that we designed. From all the results in Table 2, it can be seen that after adding the SIM and HMM modules separately, the performance of the network has been improved, which can verify the effectiveness of these two modules.

4.3. Generalizability Study

The above experimental results have confirmed that the novel U-shaped network by adopting zoom strategy has stronger COVID-19 CT image segmentation ability. To further explore its generalizability and portability, we conducted experiments on two medical polyp segmentation datasets: CVC-612 and Kvasir, where dataset Kvasir is the current largest publicly available and challenging dataset. Kvasir-SEG dataset contains 1000 high-quality images of polyps along with the corresponding precise labels. The CVC-612 dataset includes 612 images from 31 colonoscopy sequences and their labels. We compare our model with 13 state-of-the-art segmentation methods fairly by using the same training parameter settings. The experimental results are shown in Table 3, where our network (MS-UNet) outperforms or comes close to other advanced models in all metrics in the two datasets mentioned above. This indicates that our model has a strong learning capability.

Table 3
Ablation experiment results

Methods		Dice (%)	Sen (%)	Spec (%)	F-measure (%)	mIOU (%)	MAE (%)
COVID-19-CT829	Backbone	83.73	86.68	99.34	80.45	**87.83**	0.96
	Backbone + SIM	85.42	88.12	9940	84.20	86.33	0.78
	Backbone + HMM	84.73	87.64	99.25	82.25	85.93	0.82
	Backbone + SIM + HMM(Ours)	**86.25**	**89.48**	**99.70**	**85.61**	87.05	**0.75**
COVID-19-CT100	Backbone	74.06	74.56	96.31	74.02	80.37	6.78
	Backbone + SIM	76.47	77.79	**96.69**	77.01	80.55	6.06
	Backbone + HMM	75.12	76.46	96.48	76.44	80.73	6.21
	Backbone + SIM + HMM(Ours)	**79.38**	**80.12**	96.62	**77.93**	**81.52**	**5.66**

Table 4
Experimental results on polyp segmentation dataset CVC-612 and Kvasir

	Methods	Dice (%)	Sen (%)	Spec (%)	HD	mIOU (%)	MAE (%)
CVC-612	PraNet [14]	84.03	85.52	91.36	10.11	92.50	1.07
	SwinUNet [8]	78.12	84.05	95.32	30.04	86.98	2.12
	SegNet [6]	85.16	87.11	97.72	15.97	91.73	1.28
	Deeplabv3 [11]	78.19	79.06	95.81	23.72	86.27	2.13
	PSPNet [51]	86.19	88.69	97.58	15.84	92.38	1.19
	TransUNet [10]	87.43	88.49	98.37	20.48	91.16	1.38
	U-Net [35]	86.37	88.23	96.18	12.62	92.24	1.19
	U-Net++ [52]	88.82	91.11	97.73	12.83	92.91	1.11
	Inf-Net [15]	86.34	87.69	97.71	15.97	90.43	1.49
	Attention-Unet [32]	88.09	89.48	97.84	12.75	96.23	1.14
	Hiformer [20]	90.75	92.74	97.86	9.94	95.20	0.73
	DM2Tnet [41]	91.02	91.79	97.89	9.42	96.67	0.75
	C2FVL [49]	91.23	**93.07**	98.54	8.97	**97.55**	0.69
	Ours	**91.64**	92.11	**99.37**	**8.45**	94.45	**0.62**
Kvasir	PraNet [14]	77.10	80.95	94.37	43.99	80.69	5.02
	SwinUNet [8]	64.70	69.67	93.06	63.39	73.53	8.17
	SegNet [6]	83.50	87.38	95.39	38.53	85.04	4.15
	Deeplabv3[11]	70.35	75.19	95.96	56.77	76.99	6.61
	PSPNet [51]	81.47	86.89	95.64	42.04	83.64	4.65
	TransUNet [10]	79.39	85.89	97.13	39.02	81.61	5.25
	U-Net [35]	85.65	89.43	98.00	41.08	85.94	3.84
	U-Net++ [52]	85.82	89.16	97.02	39.25	86.52	3.64
	Inf-Net [15]	76.31	83.56	93.53	43.73	80.02	5.15
	Attention-Unet [32]	86.84	89.13	**98.21**	38.46	87.60	3.35
	Hiformer [20]	87.85	91.15	97.90	31.64	88.62	3.10
	DM2Tnet [41]	86.73	90.41	96.96	28.84	89.33	2.93
	C2FVL [49]	89.86	**92.36**	98.05	23.78	90.75	2.12
	Ours	**90.55**	92.05	98.00	**18.14**	**93.55**	**1.21**

All of the above experiments can confirm that our model can accurately locate and segment the target region in various challenging scenes, and has strong generalization ability and portability too.

5. Conclusion

We have built and designed a multi-scale U-shaped network which combines scale integration modules (SIM) and hierarchical mixed module (HMM) for COVID-19 CT image segmentation. The SIM and HMM are good at learning to distinguishing the effective semantic information in mixed scale, and fully exploring the imperceptible clues between the candidate object and the background environment. Both quantitative and qualitative experimental results confirmed that our network (MS-UNet) is superior to the existed COVID-19 CT image segmentation models. Future research will aim to develop federated learning methods conduct-

ed across multiple healthcare settings to enhance diagnostic accuracy and efficiency without sacrificing patient privacy. Further, we will explore toxicant detection techniques to identify and resist data contamination and ensure data quality and model robustness in a distributed learning environment.

The powerful learning and generalization capability of MS-UNet demonstration means that its application is not limited to COVID-19 CT images, but can also be extended to other medical image analysis tasks, such as tumor recognition and disease markers, which will further enhance the utility and value of the model. Future work will aim to simplify the inference structure of the network and design more lightweight models to facilitate the integration and application of this technology, especially in resource-constrained environments.

Acknowledgement

This work was supported by the key scientific research project of higher school of Henan Province under Grant 21A520022.

References

1. Adelson, E., Anderson, C., Bergen, J., Burt, P., Ogden, J. Pyramid Methods in Image Processing. RCA Engineer, 1984, 29(6), 33-41.
2. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology, 2020 Aug; 296(2), E32-E40. https://doi.org/10.1148/radiol.2020200642
3. Amin, J., Anjum, M., Sharif, M., Rehman, A., Saba, T., Zahra, R. Microscopic Segmentation and Classification of COVID-19 Infection with Ensemble Convolutional Neural Network. Microscopy Research and Technique, 2022, 85(1), 385-397. https://doi.org/10.1002/jemt.23913
4. Amyar, A., Modzelewski, R., Li, H., Ruan, S. Multi-Task Deep Learning-Based CT Imaging Analysis for COVID-19 Pneumonia: Classification and Segmentation. Computers in Biology and Medicine, 2020, 126, 104037. https://doi.org/10.1016/j.compbiomed.2020.104037
5. Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., Merhof, D. Medical Image Segmentation Review: The Success of U-Net. ArXiv, abs/2211.14830, 2022.
6. Badrinarayanan, V., Kendall, A., Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12), 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615
7. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Miguel, C.R., Vilariño, F. WM-DOVA Maps for Accurate Polyp Highlighting in Colonoscopy: Validation vs. Saliency Maps from Physicians. Computerized Medical Imaging and Graphics, 2015, 43, 99-111. https://doi.org/10.1016/j.compmedimag.2015.02.007
8. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M. Swin-Unet: UNet-Like Pure Transformer for Medical Image Segmentation. ECCV 2022 Workshops. Lecture Notes in Computer Science, 2022, vol 13803. Springer, Cham. https://doi.org/10.1007/978-3-031-25066-8_9
9. Chen, C., Zhou, K., Zha, M., Qu, X., Xiao, R. An Effective Deep Neural Network for Lung Lesions Segmentation from COVID-19 CT Images. IEEE Transactions on Industrial Informatics, 2021, 17(9), 6528-6538. https://doi.org/10.1109/TII.2021.3059023
10. Chen, J., Lu, Y., Yu, Q., Luo, X., Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv preprint arXiv:2102.04306, 2021.
11. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Proceedings of the European Conference on Computer Vision (ECCV), 2018, 801-818. https://doi.org/10.1007/978-3-030-01234-2_49
12. Cong, R., Yang, H., Jiang, Q., Gao, W., Li, H., Wang, C.C., Zhao, Y., Kwong, S. T. BCS-Net: Boundary, Context, and Semantic for Automatic COVID-19 Lung Infection Segmentation from CT Images. IEEE Transactions on Instrumentation and Measurement, 2022, 71, 1-11. https://doi.org/10.1109/TIM.2022.3196430
13. Diniz, J., Quintanilha, D., Santos Neto, A., da Silva, G., Ferreira, J., Netto, S., Araújo, J., Da Cruz, L., Silva, T., da S. Martins, C., Ferreira, M., Rego, V., Boaro, J., Cipriano, C., Silva, A., de Paiva, A., Junior, G., de Almeida, J., Nunes, R., Mogami, R., Gattass, M. Segmentation and Quantification of COVID-19 Infections in CT Using Pulmonary Vessels Extraction and Deep Learning. Multimedia Tools and Applications, 2021, 80(19), 29367-29399. https://doi.org/10.1007/s11042-021-11153-y

14. Fan, D., Ji, G., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2020, 263-273. https://doi.org/10.1007/978-3-030-59725-2_26

15. Fan, D., Zhou, T., Ji, G., Zhou, Y., Shao, L. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. IEEE Transactions on Medical Imaging, 2020, 39(8), 2626-2637. https://doi.org/10.1109/TMI.2020.2996645

16. Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., Ji, W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology, 2020, 296(2), E115-E117. https://doi.org/10.1148/radiol.2020200432

17. Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Zhong, N.S. China Medical Treatment Expert Group for COVID-19: Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine, 2020, 382(18), 1708-1720. https://doi.org/10.1056/NEJMoa2002032

18. Gunraj, H., Sabri, A., Koff, D., Wong, A. COVID-Net CT-2: Enhanced Deep Neural Networks for Detection of COVID-19 From Chest CT Images Through Bigger, More Diverse Learning. Frontiers in Medicine, 2021, 8, 729287. https://doi.org/10.3389/fmed.2021.729287

19. He, K., Zhao, W., Xie, X., Ji, W., Liu, M., Tang, Z., Shi, Y., Shi, F., Gao, Y., Liu, J., Zhang, J., Shen, D. Synergistic Learning of Lung Lobe Segmentation and Hierarchical Multi-Instance Classification for Automated Severity Assessment of COVID-19 in CT Images. Pattern Recognition, 2021, 113, 107828. https://doi.org/10.1016/j.patcog.2021.107828

20. Heidari, M., Kazerouni, A., Kadarvish, M. S., Azad, R., Aghdam, E. K., Cohen-Adad, J., Merhof, D. HiFormer: Hierarchical Multi-Scale Representations Using Transformers for Medical Image Segmentation. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, 6191-6201. https://doi.org/10.1109/WACV56688.2023.00614

21. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G. Automatic Lung Segmentation in Routine Imaging is Primarily a Data Diversity Problem, Not a Methodology Problem. European Radiology Experimental, 2020, 4(1), 50. https://doi.org/10.1186/s41747-020-00173-2

22. Houssein, E. H., Helmy, B. E., Oliva, D., Jangir, P., Manoharan, P., Elngar, A., Shaban, H. An Efficient Multi-Thresholding-Based COVID-19 CT Images Segmentation Approach Using an Improved Equilibrium Optimizer. Biomedical Signal Processing and Control, 2022, 73, 103401. https://doi.org/10.1016/j.bspc.2021.103401

23. Islam, M., Reza, T., Kaosar, M., Parvez, M. Z. Effectiveness of Federated Learning and CNN Ensemble Architectures for Identifying Brain Tumors Using MRI Images. Neural Processing Letters, 2023, 55(4), 3779-3809. Springer. https://doi.org/10.1007/s11063-022-11014-1

24. Jenssen, H. B. COVID-19 CT Segmentation Dataset. http://medicalsegmentation.com/covid19/. Accessed 04 10, 2020.

25. Jha, D., Smedsrud, P.H., Riegler, M., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D. Kvasir-Seg: A Segmented Polyp Dataset. International Conference on Multimedia Modeling. Springer, Cham, 2020, 451-462. https://doi.org/10.1007/978-3-030-37734-2_37

26. Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. Feature Pyramid Networks for Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 2117-2125. https://doi.org/10.1109/CVPR.2017.106

27. Lindeberg, T. Scale-Space Theory in Computer Vision. Springer Science & Business Media, 2013.

28. Liu, Z., Shen, Z., Li, S., Helwegen, K., Huang, D., Cheng, K. How Do Adam and Training Strategies Help BNNs Optimization. International Conference on Machine Learning, PMLR, 2021, 6936-6946.

29. Liu, Y., Wu, Y.H., Ban, Y., Wang, H., Cheng, M. Rethinking Computer-Aided Tuberculosis Diagnosis, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, 2643-2652. https://doi.org/10.1109/CVPR42600.2020.00272

30. Lizzi, F., Agosti, A., Brero, F., Cabini, R. F., Fantacci, M. E., Figini, S., Lascialfari, A., Laruina, F., Oliva, P., Piffer, S., Postuma, I., Rinaldi, L., Talamonti, C., Retico, A. Quantification of Pulmonary Involvement in COVID-19 Pneumonia by Means of a Cascade of Two U-Nets: Training and Assessment on Multiple Datasets Using Different Annotation Criteria. International Journal of Computer Assisted Radiology and Surgery, 2022, 17(2), 229-237. https://doi.org/10.1007/s11548-021-02501-2

31. Narin, A., Kaya, C., Pamuk, Z. Automatic Detection of Coronavirus Disease (COVID-19) Using X-Ray Images and Deep Convolutional Neural Networks. Pattern Analysis and Applications, 2021, 24(3), 1207-1220. https://doi.org/10.1007/s10044-021-00984-y

32. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M. J., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S. G., Hammerla, N. Y., Kainz, B., Glocker, B., Rueckert, D. Atten-

tion U-Net: Learning Where to Look for the Pancreas. 2018. ArXiv, abs/1804.03999.

33. Pang, Y., Zhao, X., Xiang, T., Lihe, Z., Lu, H. Zoom In and Out: A Mixed-Scale Triplet Network for Camouflaged Object Detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, 2150-2160. https://doi.org/10.1109/CVPR52688.2022.00220
34. Połap, D., Jaszcz, A., Wawrzyniak, N., Zaniewicz, G. Bilinear Pooling with Poisoning Detection Module for Automatic Side Scan Sonar Data Analysis, IEEE Access,2023, 11, 72477-72484. https://doi.org/10.1109/ACCESS.2023.3295693
35. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
36. Salehi, S., Abedi, A., Balakrishnan, S., Gholamrezanezhad, A. Coronavirus Disease 2019 (COVID-19) Imaging Reporting and Data System (COVID-RADS) and Common Lexicon: A Proposal Based on the Imaging Data of 37 Studies. European Radiology, 2020, 30(9), 4930-4942. https://doi.org/10.1007/s00330-020-06863-0
37. Shamim, S., Awan, M., Mohd, Z., Naseem, U., Mohammed, M., Garcia-Zapirain, B. Automatic COVID-19 Lung Infection Segmentation Through Modified U-Net Model. Journal of Healthcare Engineering, 2022, 6566982. https://doi.org/10.1155/2022/6566982
38. Shan, D., Li, Z., Chen, W., Li, Q., Tian, J., Hong, Q. Coarse-to-Fine COVID-19 Segmentation via Vision-Language Alignment.2023. arXiv preprint arXiv:2303.00279. https://doi.org/10.1109/ICASSP49357.2023.10096683
39. Tobergte, D., Curtis, S. Scale-Space Theory in Computer Vision. 2013.
40. Wang, C., Horby, P., Hayden, F., Gao, G. A Novel Coronavirus Outbreak of Global Health Concern. The Lancet. 2020 Feb 15;395(10223):470-473. https://doi.org/10.1016/S0140-6736(20)30185-9
41. Wang, L., Wang, J., Zhu, L., Fu, H., Li, P., Cheng, G., Feng, Z., Li, S., Heng, P. Dual Multiscale Mean Teacher Network for Semi-Supervised Infection Segmentation in Chest CT Volume for COVID-19. IEEE Transactions on Cybernetics, 2023, 53(10), 6363-6375. Epub 2023 Sep 15. PMID: 37015538. https://doi.org/10.1109/TCYB.2022.3223528
42. Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., Tan, W. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA, 2020, 323(18), 1843-1844. https://doi.org/10.1001/jama.2020.3786
43. Wang, L., Lin, Z.Q., Wong, A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. Scientific Reports, 2020, 10, 19549. https://doi.org/10.1038/s41598-020-76550-z
44. Wang, S., Kang, B., Ma, J., Zeng, X., Xu, B. A Deep Learning Algorithm Using CT Images to Screen for Coronavirus Disease (COVID-19). European Radiology, 2021, 31(8), 6096-6104. https://doi.org/10.1007/s00330-021-07715-1
45. Wang, W., Zhao, S., Shen, J., Hoi, S.C.H., Borji, A. Salient Object Detection with Pyramid Attention and Salient Edges. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019,1448-1457. https://doi.org/10.1109/CVPR.2019.00154
46. West, C., Montori, V., Sampathkumar, P. COVID-19 Testing: The Threat of False-Negative Results. Mayo Clinic Proceedings, 2020, 95(6). https://doi.org/10.1016/j.mayocp.2020.04.004
47. Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Ni, Q., Chen, Y., Su, J., Lang, G., Li, Y., Zhao, H., Liu, J., Xu, K., Ruan, L., Sheng, J., Qiu, Y., Wu, W., Liang, T., Li, L. A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia. Engineering (Beijing), 2020, 6(10), 1122-1129. https://doi.org/10.1016/j.eng.2020.04.010
48. Yazdani, S., Minaee, S., Kafieh, R., Saeedizadeh, N., Sonka, M. COVID CT-Net: Predicting COVID-19 From Chest CT Images Using Attentional Convolutional Network. 2020. arXiv preprint arXiv:2009.05096.
49. Ye, Z., Zhang, Y., Wang, Y., Huang, Z., Song, B. Chest CT Manifestations of New Coronavirus Disease 2019 (COVID-19): A Pictorial Review. European Radiology, 2020, 30(8), 4381-4389. https://doi.org/10.1007/s00330-020-06801-0
50. Zhang, J., Xie, Y., Li, Y., Shen, C., Xia, Y. COVID-19 Screening on Chest X-Ray Images Using Deep Learning Based Anomaly Detection. 2020. arXiv, abs/2003.12338.
51. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. Pyramid Scene Parsing Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 2881-2890. https://doi.org/10.1109/CVPR.2017.660
52. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2018, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1