ABSTRACT The new coronavirus, which has become a global pandemic, has confirmed more than 88 million cases worldwide since the first case was recorded in December 2019, causing over 1.9 million deaths. Since COIVD-19 lesions have clear imaging features on CT images, it is suitable for the auxiliary diagnosis and treatment of COVID-19. Deep learning can be used to segment the lesions areas of COVID-19 in CT images to help monitor the epidemic situation. In this paper, we propose a multi-point supervision network (MPS-Net) for segmentation of COVID-19 lung infection CT image lesions to solve the problem of a variety of lesion shapes and areas. A multi-scale feature extraction structure, a sieve connection structure (SC), a multi-scale input structure and a multi-point supervised training structure were implemented into MPS-Net. In order to increase the ability to segment various lesion areas of different sizes, the multi-scale feature extraction structure and the sieve connection structure will use different sizes of receptive fields to extract feature maps of various scales. The multi-scale input structure is used to minimize the edge loss caused by the convolution process. In order to improve the accuracy of segmentation, we propose a multi-point supervision training structure to extract supervision signals from different up-sampling points on the network. Experimental results showed that the dice similarity coefficient (DSC), sensitivity, specificity and IOU of the segmentation results of our model are 0.8325, 0.8406, 0.9988 and 0.742, respectively. The experimental results demonstrated that the network proposed in this paper can effectively segment COVID-19 infection on CT images. It can be used to assist the diagnosis and treatment of new coronary pneumonia.

INDEX TERMS COVID-19, CT, U-Nnet, MPS-Net.
the late stage [10], as shown in Fig.1 In clinical applications, segmentation of different organs and lesions from chest CT slices can provide important information for doctors to diagnose and quantify pulmonary diseases [15]. However, manual segmentation of lesions on CT images requires extensive experience of the radiologist. Due to the increasing number of patients, using CT frequently has placed a heavy burden on the radiology department. Therefore, there is an urgent need to find a low-cost, rapid tool to effectively detect and diagnose COVID-19.

Deep learning has been widely applied in autonomous driving, facial recognition and medical image processing in recent years [12]–[14]. Convolutional Neural Networks (CNNs) is one of the most representative deep learning technologies, which has been used to process medical images in many researches. At present, many common deep learning networks are used for image segmentation including FCN [16], SegNet [17], U-Net [18], etc. U-Net and the modified U-Net (U-Net++ [19]) have been widely used in medical image segmentation. In this paper, a multi-point supervised training network (MPS-Net) is proposed, which is based on U-Net network structure with multi-scale feature extraction structure, sieve connection structure and multi-point supervised structure, and it is used for automatic and accurate segmentation of COVID-19 CT images. The experimental results show that this method has better performance and certain competitiveness than the existing classical methods. The main contributions of this paper are as follows:

1. We propose a network model with a sieve connection (SC) structure, a multi-scale input structure and a multi-point supervised training structure for COVID-19 lesion segmentation.
2. Due to the geometrical features of the lesions are diverse, we use the sieve connection module to screen out the characteristic information of lesions of different sizes.
3. A multi-point supervised training structure is proposed to improve the segmentation accuracy of the model by using different scales ground truth at different up-sampling points.

The rest of this paper is organized as follows: Section II presents the related works; Section III presents the improved method; Section IV analyzes and discusses the experimental results; Section V summarizes the paper and draws our conclusions.

II. RELATED WORKS

With the advances in artificial intelligence, many deep learning network models have been proposed for medical image processing. For example, Milletari et al. [20] proposed a U-Net-based V-Net for prostate MRI image segmentation network. V-Net combined with different MRI images to achieve end-to-end prostate segmentation. Wu et al. [21] introduced a COVID-19 classification and segmentation system, that was trained on a dataset containing 144,167 CT scans, collected from 400 COVID-19 patients and 350 uninfected cases. Their JCS model achieved a 78.3% Dice Coefficient on the segmentation test set, and a sensitivity of 95.0%, and a specificity of 93.0% on the classification test set. Jin et al. [22] created that DUNet network introduced the idea of transformable convolution based on U-Net. The algorithm utilizes the local characteristics of retinal vessels to achieve the end-to-end segmentation task. DUNet can adaptively adjust the size of the convolution kernel according to the thickness and shape of the segmented vessels, and obtain accurate vascular segmentation results based on multi-scale convolution. Recently, with the COVID-19 outbreak, some researchers have proposed segmentation models of COVID-19 lesions and achieved good performance. For example, Fan et al. [23] have proposed a new COVID-19 lung infection segmentation network (Inf-Net) that automatically segment the infected area. The Inf-Net uses a parallel partial decoder to generate a global map and aggregate high-level features. Then use explicit edge attention and implicit reverse attention to enhance the representation. In addition, the author also uses weak supervision to train the network. The dataset consists of 50 CT images with ground truth labels and 1600 CT images with pseudo labels. The DSC, sensitivity and specificity of the model were 0.739, 0.725 and 0.960 respectively. Zheng et al. [24] designed a multi-scale discriminant network (MSD-Net) for multi-class segmentation of COVID-19 lung infections on CT images. The network increases the receptive field through pyramid convolution block (PCB) to enhance the segmentation ability of infected areas of different sizes, and then uses channel attention block (CAB) structure and residual refinement block (RRB) structure to fuse and improve the feature map. The dice similarity coefficient (DSC), sensitivity and specificity of the model for three different types of infection were (0.7422, 0.8593, 0.9742), (0.7388, 0.8268, 0.9869) and (0.8769, 0.8645, 0.9889). Wang et al. [25] proposed a COPL-E-Net automatic segmentation network and a new loss function (NR-Dice). It is characterized by noise robustness. The DSC of experimental results was 80.72±9.96. Xie et al. [26] proposed the RTSU network based on U-Net. A new non-local neural network (NNN) module is introduced to make use of the structured relationship. The proposed module learns the visual and geometric relations between all convolution...
features in order to generate self-attention weights and use transfer learning to conduct secondary training of the network. The results showed that RTSU-Net performed better than the other three models (3D-UNet, FRV-Net, PDV-Net) in severe pulmonary infection caused by COVID-19. Song et al. [27] used ResNet50 network to process all sections in each 3D chest CT image, and then formed the final prediction of each CT image. Wang et al. [28] combined U-Net and 3D-CNN for the diagnosis and assessment of COVID-19. The sensitivity and specificity of the DSC obtained by the model are 0.754, 0.972 and 0.922, respectively. Zhou et al [29] proposed a deep learning algorithm for solving the problem of large scenes and small objects. The algorithm decomposes the 3D segmentation problem into three 2D segmentation problems, thus reducing the complexity of the model by an order of magnitude. After several times of removing the poor quality sample images, the DSC of the algorithm segmentation results reached 0.903.

This paper presents an MPS-Net for CT image segmentation of COVID-19 lesion areas. We used the image enhancement method of rotation, flip, translation and cropping to expand 300 CT images to 9000 for network training, and 68 CT images were used to test the model after training. Dice similarity coefficient (DSC), Sensitivity (Sens), Specificity (Spec) and IOU were used to evaluate the segmentation results of the model. The Datasets from: https://medicalsegmentation.com/covid19.

III. METHOD

U-Net is a convolutional neural network for biomedical image segmentation proposed by Ronneberger et al. in 2015 [18]. The network structure is composed of symmetrical encoder part on the left and decoder part on the right. The encoder part is used to extract features from the image, and the structure follows the typical structure of convolutional network. Each convolution layer contains two 3 × 3 convolution operations, and each convolution operation is followed by an activation function (ReLU) and a max-pooling operation with a pooling size of 2 × 2 and step size of 2 for down-sampling. The number of repetitions is four. In each down-sampling step, we double the number of feature channels. Decoders, on the other hand, are used to construct segmentation map based on features obtained from encoders. It consists of a 2 × 2 transpose convolution of for up-sampling (feature channel reduced by half), which is connected to the encoder’s feature map and two 3 × 3 convolutions, each followed similarly by a activation function ReLU. In the last layer, the classifier, 1 × 1 convolution and a sigmoid activation function are used to generate the probability map of the final segmentation.

A. NETWORK ARCHITECTURE

The classical U-Net model realizes the fusion of low-level image features and high-level image features through the multi-jump structure, so as to extract more features. In this paper, a new network based on U-Net is proposed for lesion segmentation of COVID-19 CT images. Multi-scale feature extraction structure is used to replace the 3 × 3 convolution operation in U-Net. The multi-scale feature extraction module based on Inception structure adopts convolution of different kernel sizes to improve the generalization and expression ability of the network. Multi-scale input structure can compensate for lost edge position information in convolution and pooling operations. A sieve connection structure is designed to replace the original connection structure in U-Net and reduce the number of channels by half. At the same time, in order to preserve the location information of the target feature accurately, the maximum pool index is used in the upsampling process of the feature fusion decoder. The structure can retain the information of different granularity and ensure the comprehensiveness of the extracted information. In the decoder phase, We add second stage upsampling to extract more feature information. At first, we use the attention gate module to make the model pay more attention to the really important information in the training. Then, multi-point supervised training is introduced to improve the training accuracy of the model. The overall structure is shown in Fig. 2. 

B. MULTI-SCALE FEATURE EXTRACTION MODULES

The Inception module [30] is the core structure of the GoogleNet network model that achieved the best results in ILSVRC 2014. The structure of Inception network increases the depth and width of the network and improves the utilization of computing resources in the network while keeping the calculation budget unchanged. Multiple scale feature information is processed at the same layer using filters of different sizes, and then aggregated at the next layer in order to extract multiple scale fusion features in the next Inception module [31]. As shown in Fig. 3, filters of sizes 1 × 1, 3 × 3, and 5 × 5 are used for base Inception. The output of convolution and the max pool is concatenated, forming the input for the next stage. Rather than using a properly sized filter at one level, using multiple sized filters makes the network wider and deeper, so it can recognize different scale features. In order to further refine the feature, 1 × 1 convolution is applied to reduce the feature dimension before the convolution of 3 × 3 and 5 × 5.

In this work, we use two 3 × 3 filters in series instead of 5 × 5 convolution filters, because they have an equivalent receiving field. This reduces the cost of calculation, resulting in fewer parameters and less training time. The multi-scale feature extraction method based on inception structure is shown in Fig. 4. Three sets of convolution operations (1 × 1 convolution, 3 × 3 convolution and two 3 × 3 convolutions) with different scales of step size of 2 and padding of 0 are adopted in the previous layer feature map (H × W × C). After connecting the three sets of obtained feature maps, an activation function (ReLU) and a batch normalization (BN) are adopted.

The problem of overfitting can be alleviated by using BN layer. The convolution operation of multi-scale can make the network model have good ability to learn the target features of different scales.
C. SIEVE CONNECTION MODULE
The sieve connection module (SCM) is used to replace the original skip connection layer in U-Net network. Similar to the multi-scale feature extraction structure, the sieve connection structure also adopts three different size convolution kernels. When the convolution operation with reduced receptive field was used, the input feature map is elements-wise multiplied by the feature graph with larger receptive field as input. Then, the extracted feature maps with three different scales are combined with the input feature maps, and the convolution of $1 \times 1$ is used to alleviate the semantic gap between the low-level features and the high-level features, and the number of feature channels is reduced by half. The sieve skip connection structure is shown in Fig.5.

D. MULTI-SCALE INPUT MODEL
Convolution and pooling inevitably lead to the loss of pixel position information, which can be alleviated by using multi-scale input (image pyramid). An image pyramid is a
The bottom of the pyramid is a high-resolution representation of the image to be processed, while the top is a low-resolution approximation.

In this paper, the input image is downsampled with a size of 2 and a step size of 2 according to the steps of the encoder, followed by the convolution operation of 3 × 3. Then, the feature map obtained from the multi-scale input and the output feature map obtained from the lower sampling of the upper layer are concatenated as the input feature of this layer, as shown in Fig. 6.

**E. THE DECODING MODULE**

The decoding module is composed of attention module and multi-point supervision training module. In the existing methods, the single signal supervised training model is mostly used. To improve the accuracy of segmentation results, we designed a multi-point supervision (MPS) training module. On the basis of U-Net structure, the module introduces the second stage of up-sampling to further extract the feature information. After each up-sampling, the attention gates (AGs) is used to make the network pay more attention to the important information. AGs are commonly used for natural image analysis, knowledge graphs, and language processing (NLP), for image subtitle [32], machine translation [33], [34], and classification [35]–[37] tasks. The initial work is to explore the attention map by interpreting the gradient of the output category score relative to the input image.

It has also been used for image classification and image segmentation in recent years [38]. The schematic diagram of the structure is shown in the Fig. 7. Input features (Xl) are Scaled with attention coefficients (α) computed in AG. Spatial regions are selected by analysing both the activations and contextual information provided by the gating signal (g) which is collected from a coarser scale. Grid resampling of attention coefficients is done using trilinear interpolation. At the same time, the information extracted in the first stage of up-sampling is used in the attention module to eliminate the ambiguity of irrelevant and noise response in the second sieve skip connection and highlight the significant characteristics of sieve skip connection transmission. The convolution operation of size 1 × 1 is used to reduce the channel number of the feature graph obtained by AG to 2. The obtained feature image and the downsampling gold standard image are used for multi-point training. The experimental results show that the multi-point supervised network module can effectively improve the segmentation accuracy of the network. The structure diagram is shown in Fig. 2.

**F. LOSS FUNCTION**

Early pneumonia lesions often occupy a small area of the image. If cross-entropy is used for training, the predicted results may deviate significantly from the background. The Dice loss function proposed by Milletari et al. [20] can effectively avoid this problem, and the formula is as follows:

\[
\text{Loss}_{\text{Dice}} = 1 - \frac{2 \sum_i N p_i g_i}{\sum_i N p_i^2 + \sum_i N g_i^2} = \frac{\sum_i N (p_i - g_i)^2}{\sum_i N p_i^2 + \sum_i N g_i^2}
\]  

(1)

where \(N\) is the number of pixels in the image, \(p_i\) is the probability predicted by the neural network, and \(g_i\) is the pixel value corresponding to the gold standard. In addition, according to the formula, when the Dice coefficient is close to 1, the prediction result will be close to the ground true value. However, since the use of the Dice loss function will lead to a lower recall rate of the segmentation results, according to the Tversky index [39], we use the Tversky loss function on the basis of Dice. The definition of the loss function is as follows:

\[
\text{Loss}_{\text{Tversky}} = \frac{\sum_i N p_{0i} g_{0i}}{\sum_i N p_{0i} g_{0i} + \alpha \sum_i N p_{0i} g_{1i} + \beta \sum_i N p_{1i} g_{0i}}
\]  

(2)

where, \(p_{0i}\) is the probability that the output is predicted to be pathological, and \(p_{1i}\) is the probability that the output is predicted to be non-pathological. Similarly, \(g_{0i}\) is 1 and \(g_{1i}\) is 0 for the lesion area, and \(g_{0i}\) is 0 and \(g_{1i}\) is 1 for the lesion area. By adjusting the hyperparameters \(\alpha\) and \(\beta\), we can control the tradeoff between false positives and false negatives. When \(\alpha = \beta = 0.5\), this function becomes the Dice loss function. When \(\alpha = \beta = 1\), the function becomes the Jaccard loss function. In this paper, we set \(\alpha\) to 0.4 and \(\beta\) to 0.6.
Binary cross entropy (bce) is a special case of cross-entropy Loss function and is used only for dichotomous problems, and the formula is as follows:

$$\text{Loss}_{\text{bce}} = -\sum [y \cdot \log(p) + (1-y) \cdot \log(1-p)]$$  

(3)

where, $y$ and $p$ were ground truth and prediction, $y$ was 0 or 1, and $p$ was between 0 and 1. Therefore, the loss function adopted by the network is defined as:

$$\text{Loss} = \lambda \text{Loss}_{\text{bce}} + (1 - \lambda) \text{Loss}_{\text{Tversky}}$$  

(4)

where, $\lambda$ is the penalty coefficient used to balance the two functions and takes the value of 0.6.

Since the multi-point supervision model is adopted in this paper and 5 supervisory signals are extracted in the decoding stage, the overall loss function of this model is defined as follows:

$$\text{Loss}_{\text{overall}} = \alpha \text{Loss}_s + \sum_i \beta_i \text{Loss}_i$$  

(5)

where, $\text{Loss}_s$ and $\text{Loss}_i$ respectively represent loss functions $\text{Loss}_s$, $\text{Loss}_1$, $\text{Loss}_2$, $\text{Loss}_3$ and $\text{Loss}_4$ in FIG. 2, and their formulae are uniformly defined as (4). $\alpha$ is 0.6, $\beta$ is 0.4.

G. EXPERIMENT SETUP
We used a PC equipped with an Intel Core i7-10700k, 3.8 GHz CPU with 16 GB RAM and 8 GB of RTX2070S GPU for MSFFU-Net training. The operating system of
TABLE 1. Comparison with other methods.

| Method       | MPS Model | Dice.    | Sens.    | Spec.    | IOU    |
|--------------|-----------|----------|----------|----------|--------|
| U-Net        | 0.7525    | 0.8061   | 0.9977   | 0.646    |        |
| H-DenseUNet  | 0.7910    | 0.7841   | 0.9987   | 0.689    |        |
| U-Net++      | 0.8216    | 0.8366   | 0.9987   | 0.729    |        |
| MPS-Net      | 0.8159    | 0.8205   | 0.9987   | 0.725    |        |

In the process of model training, the stochastic gradient descent optimization algorithm was used to iteratively solve the parameters of the COVID-19 image segmentation network. The initial learning rate was set to 0.0001, which was changed to 0.1 times the current value every 20 epochs. The batch size was set to 2, total training epochs to 50, and the order of the data set was randomly scrambled after each training epoch. It took about 80h to train the network on this platform.

### H. EVALUATION METRICS

In order to quantitatively evaluate the performance of the segmentation results of the proposed algorithm, four evaluation indicators were used in this paper named: Sensitivity (Sens), Specificity (Spec), IOU and DSC to evaluate the experimental results. Sensitivity is the ratio of the number of correctly detected COVID-19 infection pixels to the total number of detected pixels. Specificity is the ratio of the number of correctly detected non-COVID-19 infection pixels to the total number of non-COVID-19 infection pixels. Dice is the ratio of twice the number of correctly detected COVID-19 infection to the sum of the total number of detected pixels and the total number of ground truth pixels. The expressions of Sens, Spec, IOU and DSC are defined as follows:

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)
\]

\[
\text{Specificity} = \frac{TN}{FP + TN} \quad (7)
\]

\[
\text{DSC}(G, S) = \frac{2 |G \cap S|}{|G| + |S|} \quad (8)
\]

\[
\text{IOU}(G, S) = \frac{|G \cap S|}{|G| \cup |S|} \quad (9)
\]

where TP, TN, FP, FN, G, and S denote true positive, true negative, false positive, false negative, ground truth and segmentation, respectively. In this model, positive refers to COVID-19 infection and negative refers to background. Therefore, they are the four kinds of COVID-19 infection segmentation results based on the fact that each pixel can be segmented correctly or incorrectly. TP represents infection pixels are correctly detected as COVID-19 infection; TN represents background pixels are correctly detected as background; FP represents background pixels are incorrectly detected as COVID-19 infection; FN represents infection pixels are incorrectly detected as background.

### IV. RESULTS AND DISCUSSION

Segmentation results of COVID-19 images are shown in Fig.8. In order to verify the performance of the model, we compared three models with our model and used the same parameters, configuration environment and training sets. Table 1 presents DSC, Sens., Spec and IOU results for different models. Compared with U-Net, H-DenseUNet, and Unet++, DSC scores increased by 8%, 4.1% and 1.1%, respectively. Sens scores increased by 3.6%, 5.6% and 0.4%, respectively; IOU scores improved by 9.4%, 6% and 1.3%, respectively. Compared to our own model, after adding multi-point monitoring, Dice scores, Sens scores, Spec scores and IOU scores increased by 1.7%, 2% and 1.7%, respectively.

Fig.8 shows a qualitative comparison between our model and the other four segmentation models. COVID-19 lesions showed rich diversity in morphology and area. This property leads to segmentation difficulties. Remarkably, the MSP-Net proposed in this paper can effectively solve this problem, indicate that our MSP-Net outperform the baseline.
methods. Specifically, they produce segmentation results that are close to the basic facts, and far fewer organizations are missegmented. In Fig.8, column (a) represents the original CT image; Column (b) represents the segmentation results of V-Net; (c) Column represents the segmentation results of the H-DenseUNet; Column (d) represents the segmentation result of U-Net++; Columns (e) represents the segmentation result of MSP-Net without multi-point supervision; Columns (f) and (g) represent the segmentation results of MSP-Net and the ground truth, respectively.

Fig.9 shows the detailed information of segmentation results more clearly. It can be seen that the method proposed in this paper can accurately segment the pathological areas of COVID-19 CT images. It has low over-segmentation rate and missing segmentation rate.

In order to further verify the segmentation effect of our model, we divided the test samples into six parts according to the size of lesion area (0-1K, 1K-2K, 2K-3K, 3K-4K, 4K-5K, >5K, the unit is the number of pixels) and evaluate the segmentation results after using different models to segment these parts. Fig.10 shows the DSC scores after the segmentation of the six groups of samples using different models. As shown in the figure, the method proposed in this paper is superior to other methods in all the six groups of samples and significantly improves the segmentation accuracy of samples of different sizes. Fig.11 shows the IOU scores after the segmentation of the six groups of samples using different models.
models. The data distribution is shown in Table 2. And the best results are highlighted.

Fig.12 shows the Sens scores after the segmentation of these six groups of samples using different models. Fig.13 shows the Spec scores after the segmentation of these six groups of samples using different models. As the figure shows, our model is overall superior to the others. The data distribution is shown in Table 3. And the best results are highlighted.

Through qualitative and quantitative analysis results, it can be seen that our MSD-Net is superior to other segmentation models and can produce more accurate segmentation results, which are closer to the real situation and less misorganization of segmentation.

V. CONCLUSION

Under the circumstance that novel coronavirus has not been effectively controlled in the world and medical resources are effective, it is of great significance to use deep learning to assist doctors in clinical diagnosis. In clinical diagnosis, the geometrical shape of the lesion plays an important role in the correct diagnosis. Therefore, the correct segmentation of lesions can effectively help doctors to make a correct diagnosis. In this paper, we propose a multi-point supervised network for segmentation of COVID-19 lesions. We introduced the multi-scale feature extraction module on the basis of U-net network, replaced the original skip connection structure of U-net with the sieve connection, reduced the boundary missing caused by convolution operation through the multi-scale input model, and introduced the multi-point supervision structure to improve the segmentation accuracy. Compared with other advanced network models, our network has better segmentation results. DSC index, Sens index, Spec index and IOU index are 0.8325, 0.8406, 0.9988 and 0.742, respectively. The experimental results show that the proposed MPS-Net can effectively segment the COVID-19 CT infected lesions.

ACKNOWLEDGMENT

(Hong-Yang Pei and Dan Yang are co-first authors.)

REFERENCES

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, D. Phil, W. Tan, and P. Niu, “A novel coronavirus from patients with pneumonia in China, 2019,” New Eng. J. Med., vol. 382, no. 8, pp. 727–733, Feb. 2020.

[2] D. Benvenuto, M. Giovannetti, M. Salemi, M. Prosperi, C. De Flora, L. C. Alcantara, S. Angeletti, and M. Ciccozzi, “The global spread of 2019-nCoV: A molecular evolutionary analysis,” Pathogens Global Health, vol. 114, no. 2, pp. 64–67, Feb. 2020.

[3] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, “Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19,” IEEE Rev. Biomed. Eng., vol. 14, pp. 4–15, 2020, doi: 10.1109/RBME.2020.2987975.

[4] J. M. Sharfstein, S. J. Becker, and M. M. Mello, “Diagnostic testing for the novel coronavirus,” J. Amer. Med. Assoc., vol. 323, no. 15, pp. 1437–1438, 2020.

[5] WHO. (Jan. 12, 2021). COVID-19 Weekly Epidemiological. [Online]. Available: https://www.who.int/publications/m/item/weekly-epidemiological-update—12-january-2021

[6] Z. Y. Zu, M. D. Jiang, P. P. Xu, W. Chen, Q. Q. Ni, G. M. Lu, and L. J. Zhang, “Coronavirus disease 2019 (COVID-19): A perspective from China,” Radiology, vol. 296, Feb. 2020, Art. no. 200490.

[7] H. Liu, F. Liu, J. Li, T. Zhang, D. Wang, and W. Lan, “Clinical and CT imaging features of the COVID-19 pneumonia: Focus on pregnant women and children,” J. Infection, vol. 80, no. 5, pp. e7–e13, May 2020.

[8] C. S. Guan, Z. B. Lv, S. Yan, Y. N. Du, H. Chen, L. G. Wei, R. M. Xie, and B. D. Chen, “Imaging features of coronavirus disease 2019 (COVID-19): Evaluation on thin-section CT,” Acad. Radiol., vol. 27, no. 5, pp. 609–613, May 2020.

[9] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, “Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases,” Radiology, vol. 2019, Feb. 2020, Art. no. 200642.

[10] M.-Y. Ng, E. Y. Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M.-S. Lui, C.-S.-Y. Lo, B. Leung, P.-L. Khong, C. K.-M. Hui, K.-Y. Yuen, and M. D. Kuo, “Imaging profile of the COVID-19 infection: Radiologic findings and literature review,” Radiol., Cardiothoracic Imag., vol. 2, no. 1, Feb. 2020, Art. no. e200034.

[11] COVID-19 CT Segmentation Dataset. Accessed: Apr. 11, 2020. [Online]. Available: https://medicalsegmentation.com/covid19/

[12] X. Qiu, “Transfer learning with edge attention for prostate MRI segmentation,” 2019, arXiv:1912.09847. [Online]. Available: http://arxiv.org/abs/1912.09847

[13] J. Zheng, D. Yang, Y. Zhu, W. Gu, B. Zheng, C. Bai, L. Zhao, H. Shi, J. Hu, S. Lu, W. Shi, and N. Wang, “Pulmonary nodule risk classification in adenocarcinoma from CT images using deep CNN with scale transfer module,” IET Image Process., vol. 14, no. 8, pp. 1481–1489, Jun. 2020, doi: 10.1049/iet-ipr.2019.0248.

[14] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, C. Zhang, and D. Shen, “Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning,” IEEE Trans. Med. Imag., vol. 39, no. 8, pp. 2606–2614, Aug. 2020, doi: 10.1109/TMI.2020.2992546.
[15] P. M. Gordaliza, A. Muñoz-Barrutia, M. Abella, M. Desco, S. Sharpe, and J. J. Vaquero, “Unsupervised CT lung image segmentation of a mycobacterium tuberculosis infection model,” Sci. Rep., vol. 8, no. 1, pp. 1–10, Dec. 2018.

[16] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3431–3440.

[17] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[18] L. Kaiser, and I. Polosukhin, “Attention is all you need,” in Proc. NIPS, Dec. 2017. Available: http://arxiv.org/abs/1706.03762

[19] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “U-Net++: A nested U-Net architecture for medical image segmentation,” in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, 2018, pp. 3–11. [Online]. Available: https://arxiv.org/abs/1904.00357 and https://arxiv.org/abs/1912.05074

[20] F. Millietari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in Proc. 4th Int. Conf. 3D Vis. (3DV), Oct. 2016, pp. 565–571.

[21] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, R.-G. Zhang, and M.-M. Cheng, “JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation,” 2020, arXiv:2004.07054. [Online]. Available: http://arxiv.org/abs/2004.07054

[22] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, “DUNet: A deformable network for retinal vessel segmentation,” Knowl.-Based Syst., vol. 178, pp. 149–162, Aug. 2019.

[23] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, “Inf-Net: Automatic COVID-19 lung infection segmentation from CT images,” IEEE Trans. Med. Imag., vol. 39, no. 8, pp. 2626–2637, Aug. 2020.

[24] B. Zheng, Y. Liu, Y. Zhu, F. Yu, T. Jiang, D. Yang, and T. Xu, “MSDNet: Multi-scale discriminative network for COVID-19 lung infection segmentation on CT,” IEEE Access, vol. 8, pp. 185786–185795, 2020.

[25] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, “A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images,” IEEE Trans. Med. Imag., vol. 38, no. 8, pp. 2653–2663, Aug. 2020.

[26] W. Xie, C. Jacobs, J.-P. Carbonnier, and B. van Ginneken, “Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans,” IEEE Trans. Med. Imag., vol. 39, no. 8, pp. 2664–2675, Aug. 2020.

[27] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, H. Huang, J. Chen, R. Wang, H. Zhao, Y. Zha, J. Shen, Y. Chong, and Y. Yang, “Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images,” Med. Imag., Feb. 2020, doi:10.1101/2020.02.23.20026930.

[28] B. Wang, S. Jin, Q. Yan, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, and Z. Zheng, “AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system,” Appl. Soft Comput., vol. 98, Jan. 2021, Art. no. 106897.

[29] L. Zhou, Z. Li, J. Zhou, H. Li, Y. Chen, Y. Huang, D. Xie, L. Zhao, M. Fan, S. Hashmi, and F. Abdelkareem, “A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis,” IEEE Trans. Med. Imag., vol. 39, no. 8, pp. 2626–2654, Aug. 2020.

[30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1–9.

[31] R. Koshy and A. Mahmood, “Optimizing deep CNN architectures for face liveness detection,” Entropy, vol. 21, no. 4, p. 423, Apr. 2019.

[32] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, “Bottom-up and top-down attention for image captioning and visual question answering,” 2017, arXiv:1707.07998. [Online]. Available: http://arxiv.org/abs/1707.07998

[33] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2014, arXiv:1409.0473. [Online]. Available: http://arxiv.org/abs/1409.0473

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in Proc. NIPS, 2017, pp. 6000–6010.

[35] S. Jetley, N. A. Lord, N. Lee, and P. Torr, “Learn to pay attention,” in Proc. Int. Conf. Learn. Represent., Feb. 2018, pp. 1–14. [Online]. Available: https://openreview.net/forum?id=HyzbhWRW

HONG-YANG PEI received the B.S. degree from the Dalian University of Science and Technology, Dalian, China, 2019. He is currently pursuing the M.S. degree with the College of Information Science and Engineering, Northeastern University, Shenyang.

His current research interests include computer vision and deep learning.

DAN YANG received the B.S. and M.S. degrees in biomedical engineering and the Ph.D. degree in detection and automatic control engineering from Northeastern University (NEU), Shenyang, China, in 2002, 2005, and 2009, respectively. From 2009 to 2012, she was a Postdoctoral Researcher of computer and science engineering with NEU. From 2013 to 2014, she was a Visiting Scholar with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA, sponsored by the China Scholarship Council. She is currently an Associate Professor with the School of Information Science and Engineering, NEU. She is involved in research projects with the Fundamental Research Funds for the Central Universities of China, and the National Natural Science Foundation of China. Her current research interests include intelligent computing, physiological signal detection, and biological electromagnetic systems.

GUO-RU LIU received the B.S. degree from the School of Information and Communication Engineering, North University of China, Taiyuan, China, 2019. He is currently pursuing the M.S. degree with the College of Information Science and Engineering, Northeastern University, Shenyang.

His current research interests include electrical impedance tomography, magnetic induction tomography, and other related hybrid imaging technique.

TIAN LU received the B.S. degree from the School of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, China, 2018. She is currently pursuing the M.S. degree with the College of Information Science and Engineering, Northeastern University, Shenyang.

Her main research interests include electrical impedance tomography, especially monitoring of lung ventilation, and magnetic induction tomography.

* * *

VOLUME 9, 2021 47153