

Dual attention fusion UNet for COVID-19 lesion segmentation from CT images

Yinjin Ma^{a,*}, Yajuan Zhang^b, Lin Chen^a, Qiang Jiang^c and Biao Wei^d

^a*School of Data Science, Tongren University, Tongren, China*

^b*Cangzhou Jiaotong College, Cangzhou, China*

^c*Tongren City People's Hospital, Tongren, China*

^d*Key Laboratory of Optoelectronic Technology and Systems, Ministry of Education, Chongqing University, Chongqing, China*

Received 3 January 2023

Revised 17 March 2023

Accepted 29 March 2023

Abstract.

BACKGROUND: Chest CT scan is an effective way to detect and diagnose COVID-19 infection. However, features of COVID-19 infection in chest CT images are very complex and heterogeneous, which make segmentation of COVID-19 lesions from CT images quite challenging.

OBJECTIVE: To overcome this challenge, this study proposes and tests an end-to-end deep learning method called dual attention fusion UNet (DAF-UNet).

METHODS: The proposed DAF-UNet improves the typical UNet into an advanced architecture. The dense-connected convolution is adopted to replace the convolution operation. The mixture of average-pooling and max-pooling acts as the down-sampling in the encoder. Bridge-connected layers, including convolution, batch normalization, and leaky rectified linear unit (leaky ReLU) activation, serve as the skip connections between the encoder and decoder to bridge the semantic gap differences. A multiscale pyramid pooling module acts as the bottleneck to fit the features of COVID-19 lesion with complexity. Furthermore, dual attention feature (DAF) fusion containing channel and position attentions followed the improved UNet to learn the long-dependency contextual features of COVID-19 and further enhance the capacity of the proposed DAF-UNet. The proposed model is first pre-trained on the pseudo label dataset (generated by Inf-Net) containing many samples, then fine-tuned on the standard annotation dataset (provided by the Italian Society of Medical and Interventional Radiology) with high-quality but limited samples to improve performance of COVID-19 lesion segmentation on chest CT images.

RESULTS: The Dice coefficient and Sensitivity are 0.778 and 0.798 respectively. The proposed DAF-UNet has higher scores than the popular models (Att-UNet, Dense-UNet, Inf-Net, COPLE-Net) tested using the same dataset as our model.

CONCLUSION: The study demonstrates that the proposed DAF-UNet achieves superior performance for precisely segmenting COVID-19 lesions from chest CT scans compared with the state-of-the-art approaches. Thus, the DAF-UNet has promising potential for assisting COVID-19 disease screening and detection.

Keywords: Coronavirus disease 2019 (COVID-19), computed tomography (CT), deep learning, dual attention, medical image segmentation

*Corresponding author: Yinjin Ma, School of Data Science, Tongren University, Tongren 554300, China. E-mail: dsjmyj@gztrc.edu.cn.

1. Introduction

The spread of coronavirus disease 2019 (COVID-19) has become a global pandemic since its outbreak at the end of 2019 [1]. The World Health Organization (WHO) regarded the severe epidemic of COVID-19 as a public health emergency, which greatly affected the daily life of humanity and seriously hindered the development of the economy in the world [2]. What is more, COVID-19 also took away thousands of people's lives every day in the world [3]. Hence, it is imperative to diagnose COVID-19 infections precisely and effectively.

In clinics, the reverse transcription-polymerase chain reaction (RT-PCR) is considered the standard gold test for COVID-19 diagnosis [4]. However, RT-PCR is time-consuming and encounters with high negative rate [5]. Chest computed tomography (CT) scans [6] can act as an assistant in evaluating and detecting COVID-19 infections due to the widespread availability of CT scan devices in general hospitals [7]. Previous reports have shown that chest CT scan owns highly sensitive and efficient for COVID-19 detection and facilitate COVID-19 lesion screening at the early stage [8]. Despite the great importance of thoracic CT scanning for COVID-19 diagnosing and screening, automatically segmenting COVID-19 infections from chest CT images is a challenging task [9]. Firstly, the appearances of COVID-19 lesions are complex and variant, such as ground glass opacity in the early phase and pulmonary consolidation in the late phase. Secondly, the position and size of COVID-19 infection areas in lung CT images vary greatly at different phases and among different patients. Thirdly, COVID-19 lesions are with irregular shapes, and the boundaries between lesions and normal tissues are ambiguous. The infection regions are usually with low contrast with surrounding areas [10]. Figure 1 shows two typical examples randomly selected from the dataset that includes CT images from patients with different degrees (severe, moderate, and mild symptoms) of COVID-19. These challenges make accurate segmentation of COVID-19 lesions difficult and obstruct the road to obtaining precise manual annotations with high quality for artificial intelligence (AI) based model training [11].

Recently, AI-based techniques, particularly deep learning methods, have been applied to medical applications [12]. Various methodologies with convolutional neural networks (CNN) that succeeded in computer vision have been proposed to address different problems of medical imaging analysis and achieve state-of-the-art performance [13]. Long et al. proposed a semantic segmentation model with fully convolutional neural networks showing superior advantages competing with traditional segmentation based on registration methods [14]. The UNet for medical image segmentation was proposed by Ronneberger et al. [15] and showed a decent performance, and it was used by Li et al. to segment the lung organs to distinguish COVID-19 infection from community-acquired pneumonia with CT scans [16]. The UNet++ was proposed by Zhou et al. [17] and adopted for segmenting and detecting COVID-19 lesions from thoracic CT images [18]. Based on ordinal regression, Guo et al. developed an ensemble learning approach for diagnosing COVID-19 disease from chest CT images,

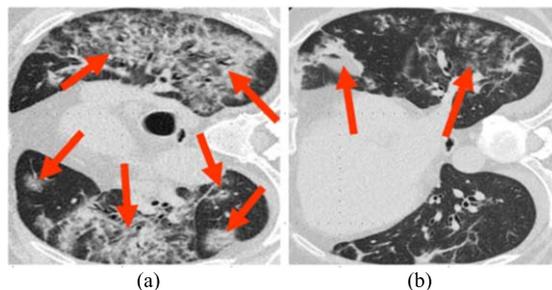


Fig. 1. COVID-19 CT scans with complex appearances of infection lesions. The two examples are from different COVID-19 patients. The areas pointed by red row is infected by COVID-19.

which attains good performance on classification for COVID-19 [19]. Fan et al. developed a network model called as Inf-Net, for infection segmentation of COVID-19 from chest CT to improve COVID-19 evaluation [20]. Wang et al. presented a noise-robust deep neural network architecture, referred to as COPLE-Net, for COVID-19 lesions automatic segmentation, and it achieves good performance [21]. By improving the UNet framework, Ma et al. proposed the pyramid pooling improved UNet, known as PPM-UNet, for segmenting COVID-19 infections from chest CT scans; the accuracy of lesion segmentation is also enhanced [22].

Different from general pulmonary infections, COVID-19 is a severe epidemic that spreads rapidly in the gathering crowd [23]. The radiologic features of COVID-19 lesions on chest lung CT images are complicated and varied. Segmenting COVID-19 lesions from chest CT scans is difficult for the above reasons. The tasks for segmentation of COVID-19 infections distributed on CT images still are challenging and remain room to be further improved [24]. Furthermore, available CT datasets of COVID-19 with high quality are in shortage for investigating and developing deep learning techniques for fighting COVID-19. Also, the CT image datasets of COVID-19 cases are expensive, and labeling is time-consuming [25]. Because dense connected convolution served as the unit of the UNet can extract accurate and precision represented feature of COVID-19 lesion in the CT images. Combined max-pooling and average-pooling can cover the shortage of single pooling (max or average pooling). Channel attention and position attention can focus on different channel and position of the feature map of COVID-19 CT images. We hypothesize that the integration of these technologies into UNet architecture can enhance the representation capacity and improve the performance for COVID-19 lesion segmentation from CT images.

This study proposed a dual attention feature fusion improved UNet framework with a loss function combining the intersection of Union (IoU) [26] and the binary cross entropy (BCE) [27], referred to as dual attention fusion UNet (DAF-UNet), for segmenting COVID-19 infections from chest CT scans. The motivation in this study is derived from the fact that, in the stage of screening COVID-19 cases rapidly, clinicians want to roughly locate the parts of infection and then extract the contour of the infection area accurately. Therefore, the area of lesions and its boundary are two critical factors that distinguish normal and infected tissues. The target for our study is to evaluate and screen overall lung infection of COVID-19, so we segmented the COVID-19 infection parts and these results can be adopted to quantitatively evaluate different types of lung lesions. In this study, the typical UNet architecture is also adopted but is expanded to a high-level network model. First, convolution in our UNet is replaced by the dense-connected convolution block [28] and a mixture of average-pooling and max-pooling (MX-pooling) acts as the downsampling, to retain and extract the detailed features of COVID-19 lesions on CT images. Second, the bridge-connected layers, including convolution, batch normalization [29], and Leaky ReLU activation [30], act as the skip connections between the encoder and decoder to bridge the semantic gap between the encoder and decoder. Third, A multiscale pyramid pooling module [31] acts as the bottleneck of the improved UNet architecture to fit the complex features of COVID-19 lesions. More importantly, the dual attention feature (DAF) fusion module containing channel attention and position attention [32], adopted to learn the long-dependency contextual [33] of COVID-19 infections distributed on chest CT images, follows after the improved UNet to enhance the capacity of deep neural network. The proposed DAF-UNet is first pre-trained on the COVID-19 CT dataset with the model-generated pseudo label, which consists of a great number of samples [20]. Then, the pre-trained DAF-UNet is refined on the public dataset of COVID-19 CT scans with standard high-quality annotations but including limited samples [34]. The evaluations on the testing set show that the proposed DAF-UNet can accurately segment COVID-19 lesions from chest CT images and achieve the highest performance competing with other state-of-the-art methods in this study qualitatively and quantitatively. The experimental results demonstrate that our DAF-UNet has the potential for COVID-19 evaluation and detection.

The rest of this article arranges as follows. Section 2 details the proposed DAF-UNet and loss functions. Section 3 describes the experiments and results, and section 4 provides the discussions. Section 5 summarizes the conclusions of this study.

2. Methods

This section illustrates the framework of dual attention feature fusion UNet (DAF-UNet). Then the key components, which includes improved UNet architecture, dual attention (the position attention module and the channel attention module), and the loss function are detailed.

2.1. Improved UNet architecture

Motivated by fully convolutional neural networks, Ronneberger proposed the original UNet, achieving stunning performance and becoming one of the most classic frameworks for medical image segmentation. The architecture of UNet consists of two symmetric components, (1) a contracting path to capture the contextual feature and compress it to a latent feature space, and (2) an expanding path to make precise locations and produce segmentation mask results corresponding to the input image. UNet architecture utilizes the skip connections between the contracting path and the extensive path for feature coupling. Also, a UNet framework can consider as an encoder-decoder. The encoder corresponds to the contracting path, and the decoder corresponds to the expanding path.

Figure 2 depicts the overall architecture of DAF-UNet. Because the slice thickness of chest CT scans of COVID-19 patients owns an extensive range (That is to say, the CT image is a single one and not the volume CT images.), 2D CNNs are adopted to segment COVID-19 lesions in this study. Because CT images are monochrome images and not color images, our deep learning model has one input channel to input COVID-19 CT images. Inspired by the initial UNet architecture and its variants which achieve high performance but differ from them, we extend the UNet to an advanced framework with several vital components. First, mixture-pooling(MX Pooling) combined with max-pooling and average-pooling replaces the only max-pooling utilized in the original UNet. The mixture of average-pooling and max-pooling serves as down-sampling, which retains more information than adopting single max-pooling. Second, to alleviate the semantic gap of the features between the encoder and decoder, the direct skip connections between the initial UNet are replaced by bridge connected layer, which contains a convolutional layer with 1×1 kernel size, a batch normalization layer and a Leaky ReLU layer. Bridge-connected layers are utilized to map the low-level features (which are extracted from the encoder) into the decoder, and the channel number is decreased to half. Thirdly, to precisely segment COVID-19 lesions with different scales, a multiscale pyramid pooling module (PPM) is adopted to act as the bottleneck of our UNet architecture. The PPM consists of four feature representations with different levels (the bin sizes are 1×1 , 2×2 , 3×3 , and 5×5), which can fit complex features of COVID-19 lesions in chest CT scans.

We replace each convolutional layer of the UNet architecture with a dense-connected convolution block. A dense connection block improves the information flow by introducing direct connections from the current layer to its former and latter layers. Utilization of dense connection blocks can facilitate the training process.

2.2. Dual attention feature fusion

Because of the challenge of segmenting complex lesions of COVID-19 infection from chest CT scans, dual attention fusion is integrated into the proposed network model to extract long-dependency contextual features. The dual attention feature (DAF) fusion module includes position and channel

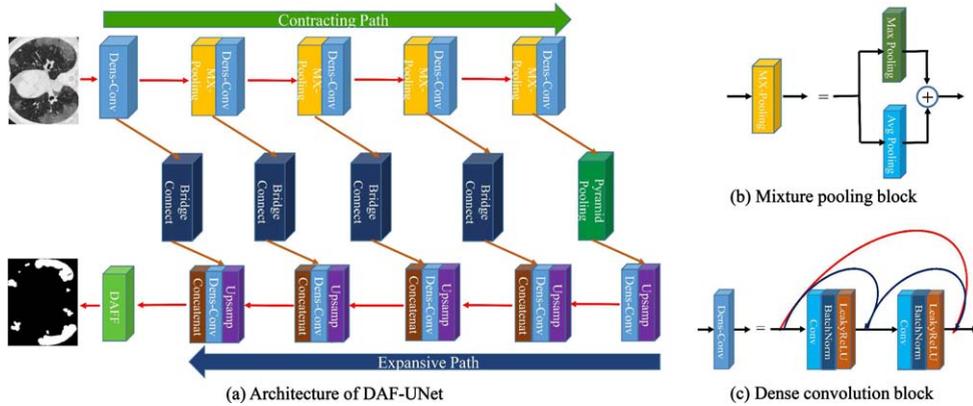


Fig. 2. Architecture of the DAF-UNet. It contains (a) UNet backbone, (b) Mixture pooling (MX-pooling), and (c) Dense convolution block.

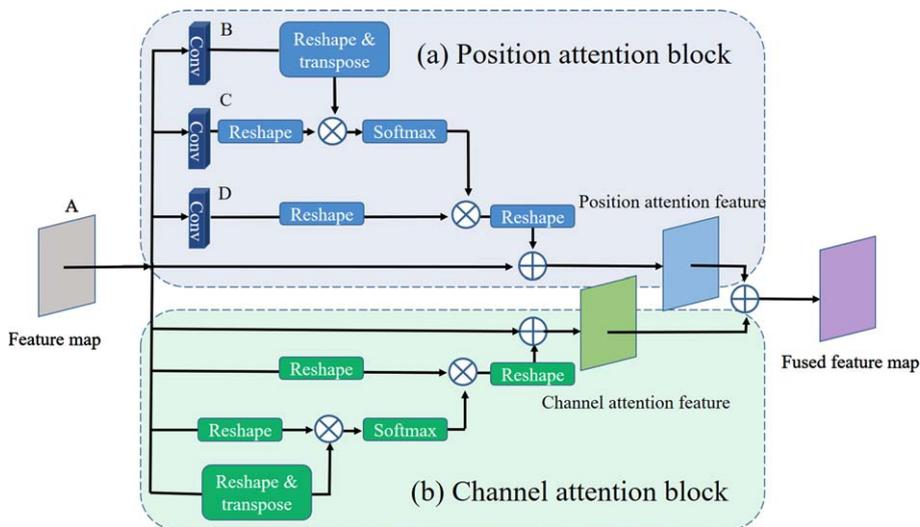


Fig. 3. Module of dual attention feature fusion (DAFF). It contains (a) Position attention block, and (b) Channel attention block.

attention blocks. The position attention block selectively aggregates the features relating to the position, and the channel attention block emphasizes the inter-dependent feature maps among all the channels. Then, feature maps obtained from the position attention and channel attention blocks are summarized to enhance the feature representation, which improves precise segmentation of COVID-19 lung lesions. Figure 3 presents the module of dual attention feature fusion.

2.2.1. Position attention block

Contextual information with a wide range contained in the COVID-19 lesions is encoded into the local feature maps by the position attention block, enhancing feature representation capability. The position attention block is illustrated in Fig. 3 (a).

Given a feature map $A \in \mathbb{R}^{C \times H \times W}$, where C is the number of channels, and H and W are the height and width of the feature maps, respectively. First, feature map A feeds into a convolution layer and obtains two new feature maps B and C . Just B and C reshape to $\mathbb{R}^{C \times N}$, where $N = H \times W$ is pixel

numbers of the new feature map are. After that, a matrix multiplication between the B' (transpose of B) and C is carried out. Finally, the spatial attention feature map $S \in \mathbb{R}^{N \times N}$ is calculated by utilizing an operation of softmax, which can be described as follows:

$$s_{ji} = \frac{\exp(B_i \cdot C_j)}{\sum_{i=1}^N \exp(B_i \cdot C_j)} \quad (1)$$

where s_{ji} denotes the measurement of the impact of i th position on j th position. The more similar the feature representation that the two positions, the more significant correlation they are.

Simultaneously, feature map A is also fed into another convolution layer to produce a new feature map $D \in \mathbb{R}^{C \times H \times W}$, and D is reshaped to $\mathbb{R}^{C \times N}$, and N denotes the pixels of the feature map (here, $N = H \times W$). Then, a matrix multiplication between D and s_{ji} is performed, and the result is reshaped to $\mathbb{R}^{C \times H \times W}$. Finally, the result is multiplied by a scale of α , and carried out an element-wise summary with the feature A . The final output of $E \in \mathbb{R}^{C \times H \times W}$ is obtained as follows:

$$E_j = \alpha \sum_{i=1}^N (s_{ji} D_i) + A_j \quad (2)$$

where E is the final feature of the weighted sum of each position feature and the initial feature A . α is a learned parameter to allocate the weight, which is initialized as zero, then inferred by Equation (2). Hence, position attention block can capture long-range dependency on global contextual information that it includes the global features of COVID-19.

2.2.2. Channel attention block

High-level feature maps of the CNN channel are taken into account as specific classifications, and they own the mutual association of semantics. In the tasks of lesion segmentation of COVID-19 infection, the high-level features are hypothesized to be ‘‘categorized’’ in the channels of the deep CNN. For example, similar semantic features (GGO, consolidation, reticulation) of contexts in the chest CT images of the COVID-19 case are more mutually correlated. Similar channels with semantical senses are categorized and aggregated by exploiting channel attention blocks.

The channel attention block is depicted in Fig. 3 (b). First, given the originated feature maps $A \in \mathbb{R}^{C \times H \times W}$ as the input, where C represents the number of network channels, and $H \times W$ denotes the size of the feature maps, the affinity matrix $X \in \mathbb{R}^{C \times C}$ can be calculated as follows:

$$x_{ji} = \frac{\exp(A_i \cdot A_j)}{\sum_{i=1}^C \exp(A_i \cdot A_j)}, i, j \in \{1, \dots, C\} \quad (3)$$

where x_{ji} depicts the effects of j th channel on i th channel. Second, a matrix multiplication of the X' (matrix transposition of X) and A is performed. Then the results of matrix multiplication are reshaped to $\mathbb{R}^{C \times H \times W}$. Finally, the multiplication between the result and a scale parameter β and the summation of all the elements with A are carried out to obtain the final feature maps of the channel attention block as follows:

$$E_j = \beta \sum_{i=1}^N (x_{ji} A_i) + A_j \quad (4)$$

where β is the learnable weight parameter. The final output feature maps are the full features of the originated features and all the channel features. In which the long-range inter-dependency among COVID-19 features in the chest CT scans is modeled so that it can boost the discriminability of the neural networks.

The long-range contextual information captured from two attention blocks is integrated to fully utilize inter-dependent global features of position and channel attention blocks. A sum fusion of the

position and channel attention output features is performed. At last, a convolutional layer followed by a soft max layer is exploited to generate the final segmented COVID-19 lesions.

2.3. Objective function

The loss function is also referred to as the objective function, which guides the training procedure of the deep learning method. The selection of loss function significantly impacts deep neural network models. Proper objective functions can accelerate the convergence of the network model during training and improve the performance of segmenting COVID-19 infections from chest CT scans. Our study uses a hybrid objective function of intersection over union (IoU) and binary cross entropy (BCE) losses to train the proposed DAF-UNet.

2.3.1. Intersection of union (IoU) loss function

In the task of COVID-19 infection segmentation from chest CT images, the targets is to do the mapping the relationship between the pixels of the COVID-19 CT scans and the segmented target labels. The metric of Intersection of Union (IoU) owns more advantages compared with the metrics of pixel-wise accuracy for image segmentation with imbalance data. IoU also is well known as Jaccard index.

IoU is generally adopted to measure the similarity between the segmented maps utilizing the network model and the standard targets. The metric of IoU is defined as follows:

$$IoU = \frac{A \cap B}{|A| + |B| - A \cap B} \quad (5)$$

where A denotes the segmentation of the network model, and B is the corresponding ground truth of segmentation. The symbol \cap denotes calculation of intersection between A and B . The loss function of IoU can be represented by Equation (6):

$$Loss_{IoU} = -\ln(IoU) \quad (6)$$

In image segmentation tasks of COVID-19 infection from CT scans, lesions and backgrounds in the CT image of the COVID-19 patients are primarily imbalanced. IoU loss can enhance the capability to segment COVID-19 infections from CT images.

2.3.2. Binary cross entropy (BCE) loss function

To obtain accurate segmentation of COVID-19 infection from chest CT images with high probability, the proposed DAF-UNet should train with a suitable objective function. For this purpose, binary cross entropy (BCE) loss is adopted in our study, which is described as follows:

$$Loss_{BCE} = -(A \log B - (1 - A) \log(1 - B)) \quad (7)$$

where A and B denote the same symbols as Equation (5). Usually, BCE loss is prone to suffer from issues of imbalanced data: the case that true negatives are more than the positives in segmentation or classification tasks. This issue can be addressed by combining it with IoU loss.

Eventually, the proposed DAF-UNet is trained in combination with IoU and BCE losses between segmented COVID-19 lesion maps and target labels of the ground truth by supervised learning. Hence, the total objective function for DAF-UNet learning represents as follows:

$$Loss_{overall} = \lambda \cdot Loss_{IoU} + (1 - \lambda) \cdot Loss_{BCE} \quad (8)$$

where the hyper-parameter λ is exploited to weight the balance between IoU loss and BCE loss. In our study, λ is set to 0.5 during the training.

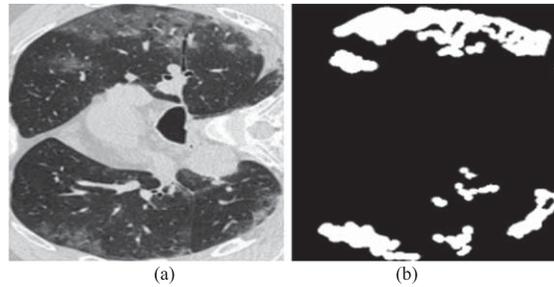


Fig. 4. Example of COVID-19 CT image and its annotated label. (a) a CT image of the dataset; (b) its corresponding COVID-19 lesion label annotated by a senior radiologist.

3. Experiments and Results

Details of the implementation of the proposed DAF-UNet are presented in this section. We first introduce the dataset of COVID-19 CT scans. Next, the experimental setting and implementation are described. Finally, the experimental results and ablation studies are unfolded

3.1. Dataset of COVID-19 CT scans

A publicly available dataset of chest CT images for COVID-19 lesion segmentation [34] is adopted to evaluate the proposed DAF-UNet. All of the chest CT scans in the dataset provided by more than 40 COVID-19 cases are gathered by the Italian Society of Medical and Interventional Radiology. This dataset of COVID-19 CT images collects 100 chest CT scans with different scales, and a senior radiologist labels the ground truth of lesion segmentations. The example of COVID-19 CT image and its annotated label are shown in Fig. 4. All of the COVID-19 chest CT scans in this dataset are split separately into 60, 20, and 20 for training, validation, and testing. Although these CT scans of COVID-19 lesions have high-quality labels, the dataset is small and prone to make the network model overfitted during the training.

Therefore, a COVID-19 CT dataset with pseudo labels released by Fan et al. [20] is introduced to our study. The Inf-Net trained using semi-supervised learning generates the pseudo labels of the COVID-19 CT scans. The pseudo labels aid the training of the proposed DAF-UNet. There are 1600 CT scans of COVID-19 and their pseudo-target labels in the dataset. An example of CT image and its pseudo label from pseudo label dataset are exhibited in Fig. 5. One can notice that the pseudo label is worse than the ground truth label which annotated by a radiologist. The COVID-19 CT dataset with pseudo labels can alleviate the shortage of COVID-19 CT data. Despite the pseudo labels generated from the semi-supervised model and owning more or fewer data quality problems, the pseudo-COVID-19 data can compensate for the lack of available COVID-19 CT scans. By utilizing the pseudo-COVID-19 dataset for network training, a coarse network model for COVID-19 segmentation is obtained, and it can be fine-tuned on the COVID-19 CT dataset with high-quality infection segmentation. The proposed DAF-UNet is first trained on the pseudo-COVID-19 CT data and then on the high-standard COVID-19 dataset. By this, the high performance of COVID-19 infection segmentation of DAF-UNet can be effectively enhanced, which is consistent with Fan et al. [20].

3.2. Implementation and parameter setting

The proposed DAF-UNet is implemented by convolution operation with two dimensions (2D CNN) since thicknesses of the chest CT scans in our dataset varies from patient to patient. The 2D DAF-UNet segments lesions of COVID-19 infection from CT scans slice by slice. The framework of DAF-UNet

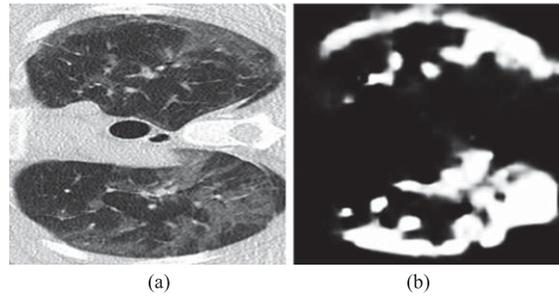


Fig. 5. An example of CT image and its pseudo label from the pseudo label dataset provided by Fan et al. (a) a CT image of the pseudo label dataset; (b) its corresponding pseudo label.

and all the training programs are implemented on a machine-learning platform with PyTorch based on Python [35]. The training procedure is accelerated by an NVIDIA RTX 2080 TI GPU configured with 11 GB VRAM graphic memory.

For precise COVID-19 lesion segmentation, all the chest CT scans in the dataset are pre-processed. First, pixels of the input CT images are normalized to the range from zero to one, and then the size is rescaled to 352×352 uniformly. In this study, the training procedure of DAF-UNet is optimized by the algorithm of Adam [36]. According to our experience, the learning rate is initialized to $1e-4$ and then is gradually reduced to $1e-5$. The parameter of Leaky ReLU is set as the default value ($1e-2$). The batch size is set as 6, and the number of the basic channel of the proposed DAF-UNet is set to 32 owing to the constraint on the graphic memory of the GPU. Two stages of the training strategy are adopted to train the network architecture. Second, The proposed DAF-UNet is first trained to utilize the CT dataset with pseudo labels for 100 epochs. By which a coarse model for COVID-19 lesion segmentation is produced. Third, fine-tuned training on the COVID-19 chest CT with high-quality target labels is performed to obtain an accurate model for COVID-19 segmentation. The time for whole training phase takes about 10 hours.

3.3. Metrics for evaluation

Many measurement metrics can be used to evaluate lesion segmentations of COVID-19 infections in chest CT scans. In our study, three metrics are adopted to evaluate the COVID-19 lesion segmentation performance of the proposed DAF-UNet and the compared segmentation methods. These evaluation metrics contain the Dice similarity coefficient (also referred to as Dice score, DIC), Sensitivity (also known as Recall and SEN.), and Specificity (SPEC). In addition, Enhance-alignment Measure (EM) and Mean Absolute Error (MAE) are introduced for further evaluating the performance of segmenting COVID-19 lesions from chest CT images.

3.3.1. Dice Similarity Coefficient (DIC)

Dice similarity coefficient (DIC) is used to evaluate the performance of results of COVID-19 lesion segmentation from CT images. Dice similarity coefficient between ground-truth label and the predicted segmented result is defined as follows:

$$\text{DIC}(A, B) = 2 \times \frac{|A \cap B|}{|A| + |B|} \quad (9)$$

where A and B are the ground-truth labels and predicted segmented results of COVID-19 lesions from CT images, respectively. DIC of one indicates the ground truth and the predicted results matching perfectly.

3.3.2. Enhanced-alignment measure (EM)

Enhanced-alignment measure (EM) [20] is the metric used to measure the similarity of two global and local binary maps. It defines as:

$$EM_{\Phi} = \frac{1}{w \cdot h} \sum_1^w \sum_1^h \Phi(A, B) \quad (10)$$

where w and h denote the width and height of the two compared binary maps, respectively. A and B separately represent the segmenting mask of COVID-19 CT scans and the ground truth of annotated labels. Φ is the representation of the enhanced-alignment matrix.

3.3.3. Mean absolute error (MAE)

Mean absolute error (MAE) [37] is adopted to evaluate the errors of pixels between the segmenting results and the ground truth. The performance evaluation of COVID-19 lesion segmentation of the proposed DAF-UNet can be measured by MAE, which is defined as:

$$MAE = \frac{1}{w \cdot h} \sum_1^w \sum_1^h |A - B| \quad (11)$$

where A is the predicted segmenting result, and B is the ground truth of the target label. w and h are the widths and the height of segmenting maps, respectively.

3.4. Segmentation results

This subsection evaluates the performance of COVID-19 lesion segmentation of DAF-UNet qualitatively and quantitatively. The aims of this study is to assist the detection of COVID-19 infection, this can fast screen the cases of COVID-19. So in our study, we segmented COVID-19 lesions from CT images and not to distinguish the three patterns of the infection. The segmenting results of the proposed DAF-UNet are compared with that of the state-of-the-art methods for medical image segmentation, including UNet [15], UNet++ [17], Dense-UNet [38], Gated-UNet [39], Inf-Net [20], and COPLE-Net [21]. We reproduce these models according to the paper, using the methods in our study to train these models, and then test the trained models by the same test set. Finally, we compared the results of COVID-19 lesions segmented from CT images.

3.4.1. Qualitative results

Figures 6 and 7 depicted two representative segmentation results of COVID-19 lesions from chest CT scans utilizing our network model and other state-of-the-art methods. From the two examples, the results of the proposed DAF-UNet outperform that of other comparing approaches in this study. As shown in the raw lung CT scans in Figs. 6 (a) and 7 (a), the characteristics of infection lesions of COVID-19 are varied in a highly complex range. Scales and locations of the infection lesion of COVID-19, such as reticulation, ground-glass opacity (GGO), and consolidation, are incredibly varied at different phases in different COVID-19 patients. In addition, infectious areas of the COVID-19 lesion with blurred boundaries, low contrast, and irregular shapes make infection lesions of COVID-19 in CT scans challenging to segment with high accuracy. The proposed DAF-UNet can obtain the segmenting results close to the corresponding ground truth.

In general, all schemes enable to segment COVID-19 lesions from chest CT images to some extent. One can observe that the proposed DAF-UNet can segment the most COVID-19 lesions from lung CT scans, including some details. The narrow strip lesions on the left of the CT scan shown in Fig. 6 can

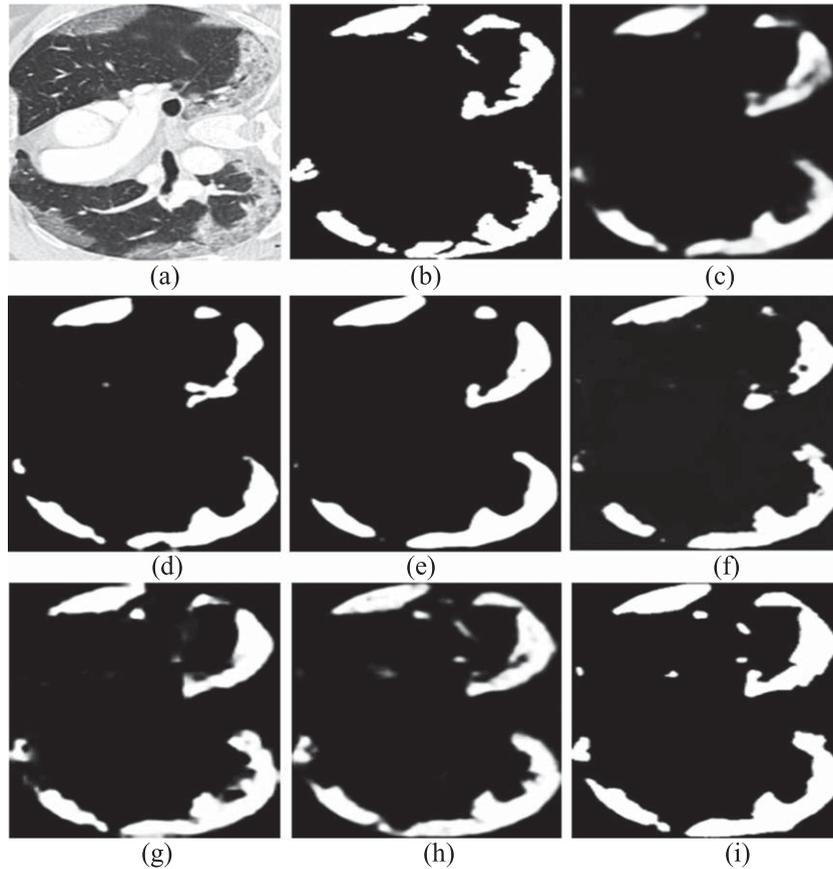


Fig. 6. Segmenting results of a COVID-19 CT scan from testing set using different methods. (a) Raw CT scan; (b) Ground truth; (c) UNet, (d) UNet++, (e) Att-UNet, (f) Dense-UNet, (g) Inf-Net, (h) COPLE-Net, and (i) DAF-UNet.

be segmented better by DAF-UNet. Most of the lesion regions distributed in Fig. 7 also are segmented by our model. Hence, the proposed DAF-UNet achieves the best visual performance compared with the other methods in this study.

3.4.2. Quantitative results

To evaluate the quantitative performance of our proposed DAF-UNet, we list the quantitative results of Figs. 6 and 7 in Table 1. The proposed DAF-UNet achieves the best scores regarding the Dice coefficient, sensitivity, EM, and MAE. Only specificity is slightly less than other methods in this study. In addition, the proposed DAF-UNet owns advantages on the MAE metric. In Fig. 6, although the EM of Att-UNet is as well as that of our DAF-UNet, the tiny COVID-19 lesions can be segmented by our method. The results with our method approximate the ground truth more than Att-UNet shown in Fig. 6.

The average values of the measured metric of the testing set are shown in Table 2. The average scores in the testing set hold a consistent trend, as in Figs. 6 and 7 and Table 1. One can observe that Table 2 further presents that the proposed DAF-UNet outperforms the other state-of-the-art segmenting models and obtains the best performance of COVID-19 lesion segmentation in our study. We also reported standard deviations (STD) of Dice similarity coefficient of the segmentation results shown in Table 3. Table 3 indicates robustness of our designed DAF-UNet. Also, the visual effects of the predicted COVID-19 segmentation shown in Figs. 6 and 7 support the quantitative results in this subsection.

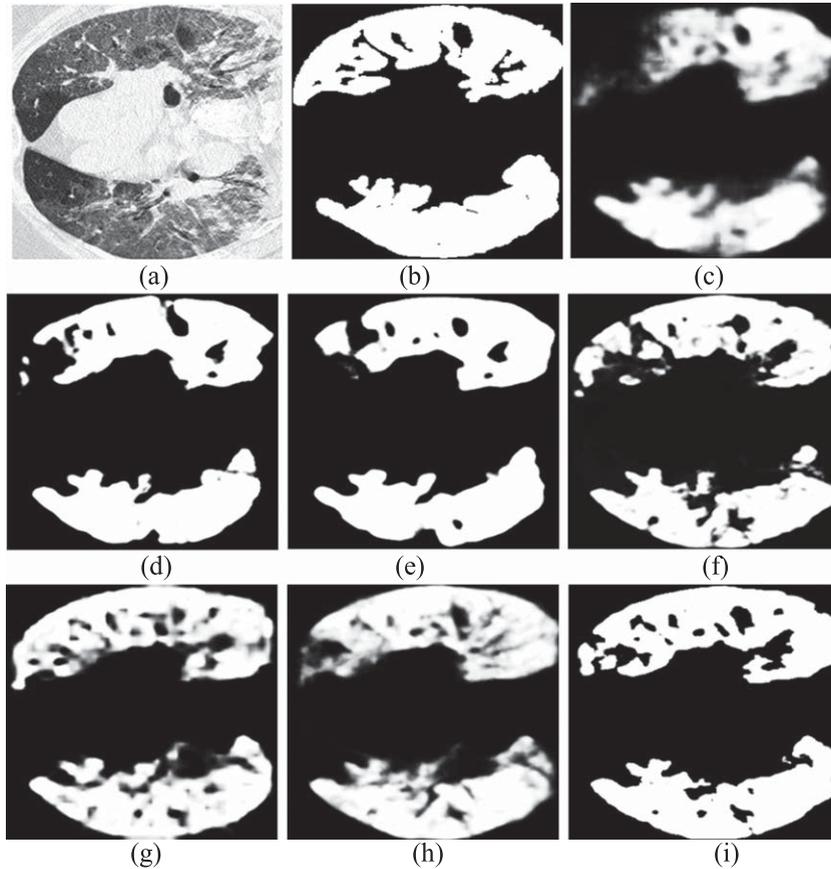


Fig. 7. Segmenting results of another COVID-19 CT scan from testing set using different methods. (a) Raw CT scan; (b) Ground truth; (c) UNet, (d) UNet++, (e) Att-UNet, (f) Dense-UNet, (g) Inf-Net, (h) COPLE-Net, and (i) DAF-UNet.

Table 1
Quantitative measurements of Figs. 6 and 7 by different methods

	Figure 6					Figure 7				
	Dic.	Sen.	Spec.	EM	MAE	Dic.	Sen.	Spec.	EM	MAE
UNet	0.773	0.743	0.968	0.909	0.063	0.772	0.688	0.945	0.798	0.161
UNet++	0.740	0.667	0.973	0.912	0.071	0.736	0.633	0.947	0.761	0.182
Att-UNet	0.812	0.781	0.972	0.945	0.054	0.843	0.776	0.953	0.867	0.119
Dense-UNet	0.804	0.764	0.967	0.928	0.061	0.842	0.770	0.957	0.859	0.119
Inf-Net	0.830	0.852	0.961	0.942	0.052	0.870	0.852	0.927	0.886	0.102
COPLE-Net	0.824	0.856	0.958	0.931	0.055	0.853	0.822	0.933	0.870	0.111
DAF-UNet	0.856	0.930	0.953	0.945	0.047	0.891	0.853	0.954	0.908	0.086

3.5. Trade-offs

In this subsection, several experiments are conducted to examine the effectiveness of the proposed DAF-UNet under varied circumstances, including components of different loss functions, with/without attention mechanisms, and with/without pre-training on the dataset with pseudo labels.

Table 2
Quantitative measurements (mean) associated with different segmentation methods in the testing set

	Dic.	Sen.	Spec.	EM	MAE
UNet	0.672	0.595	0.977	0.823	0.081
UNet++	0.725	0.642	0.979	0.868	0.070
Att-UNet	0.722	0.636	0.975	0.860	0.067
Dense-UNet	0.733	0.679	0.969	0.876	0.063
Inf-Net	0.759	0.750	0.963	0.912	0.061
COPLE-Net	0.742	0.723	0.968	0.880	0.064
DAF-UNet	0.778	0.789	0.969	0.923	0.057

Table 3
Standard Deviation (STD) of Dice Coefficient associated with different segmentation methods in the testing set

	UNet	UNet++	Att-UNet.	Dense-UNet	Inf-Net	COPLE-Net	DAF-UNet
STD	0.083	0.032	0.032	0.073	0.044	0.078	0.033

Table 4
Ablation studies with/without attention mechanism evaluating on testing set

	Dic.	Sen.	Spec.	EM	MAE
Without dual attention	0.749	0.793	0.962	0.891	0.069
With channel attention	0.758	0.803	0.959	0.901	0.067
With position attention	0.753	0.811	0.959	0.890	0.068
With dual attention	0.778	0.789	0.969	0.923	0.057

Note: The models trained to learn the overall loss function (8).

3.5.1. Effectiveness of attention mechanisms

To examine the lesion segmentations of COVID-19 infection from lung CT scans can be improved by introducing an attention mechanism. The efficient approach of channel attention is first searched; then, the spatial attention is validated; finally, the channel and spatial attention are combined. The experiments are summarized in Table 4. Both the channel and spatial attention improve the performance of COVID-19 lesion segmentation to some degree from Table 4. Utilizing a channel and spatial attention simultaneously can elevate the performance of the COVID-19 segmentation model.

3.5.2. Effectiveness of combined BCE and IoU losses

The proposed DAF-UNet are trained to learn BCE loss, IoU loss, and their combination to validate different components of loss functions. The measurements of experimental results are listed in Table 5. One can notice that the combination of BCE loss and IoU loss significantly improves the segmentation performance of the proposed DAF-UNet from Table 5.

3.5.3. Ablation on different datasets

The proposed DAF-UNet is trained on different datasets to evaluate the performance of segmentation of COVID-19 lesions. These datasets include ground-truth data labeled by a senior radiologist and pseudo-label data generated by Inf-Net. Our DAF-UNet is trained on the ground-truth dataset and pseudo-label dataset, respectively. Also, the DAF-UNet is first pre-trained on pseudo labels and

Table 5
Ablation studies for different components of loss function and their combination on testing set

Training on	Dic.	Sen.	Spec.	EM	MAE
With BCE loss	0.734	0.685	0.974	0.894	0.068
With IoU loss	0.735	0.853	0.947	0.860	0.072
With BCE and IoU losses	0.778	0.789	0.969	0.923	0.057

Table 6
Evaluation of DAF-UNet training on different dataset

Datasets	Dic.	Sen.	Spec.	EM	MAE
Pseudo labels	0.731	0.675	0.974	0.890	0.064
Ground truth	0.763	0.762	0.968	0.785	0.063
Pseudo labels + Ground-truth	0.778	0.789	0.969	0.923	0.057

then fine-tuned utilizing ground-truth labels. Table 6 lists the quantitative evaluations of COVID-19 segmentation with different datasets. From Table 6, one can observe that the proposed DAF-UNet first trained on pseudo-label data and then fine-tuned on ground-truth data can improve the overall segmentation performance significantly.

4. Discussion

Medical CT imaging is one of the most effective tools for COVID-19 detection and diagnosis, but CT image reading is tedious and time-consuming for radiologists. Machine learning-based auxiliary evaluations of CT imaging are crucial for screening and detecting COVID-19 lesions. This paper proposed a dual attention fusion UNet (DAF-UNet) for automatically segmenting COVID-19 lesions from chest CT images. The proposed method achieved better performance than this study's other state-of-the-art methods. We extended and improved the UNet architecture by replacing down-sampling with combinations of max-pooling and average-pooling, introducing dual attention fusion (containing position attention and channel attention), and exploiting dense convolution blocks as each convolution layer in the original UNet, and utilizing the bridge connected layers as the substitutes for alleviating the semantic gap between encoder and decoder of the UNet.

For achieving an excellent performance of COVID-19 lesion segmentation, deep learning-based methods not only want a suitable network architecture but also demand a significant number of high-quality training samples. Conducting deep learning approaches as a dominant way to combat the contagion of COVID-19 is of great importance. However, the CT data for research of COVID-19 applications is of acute lack and expensive, which may still be challenging for vast researchers. To handle this issue, the proposed DAF-UNet first pre-trained on a pseudo-label dataset to learn some radiological representations of COVID-19 lesions of medical CT images. Although the features hidden in the pseudo-label dataset need to be more accurate, a coarse model for COVID-19 segmentation was obtained. Then, the coarse model was further trained on a ground-truth dataset with high-quality labels. After fine-tuning, the DAF-UNet can extract the accurate features of COVID-19 infections from chest CT scans. By two stages of training, making full use of the tremendous amount of pseudo-label data and limited ground truth data annotated by radiologists, the proposed DAF-UNet outperforms the popular approaches on the quantitative and qualitative results.

Segmentation methods for COVID-19 lesions based on deep learning are representative end-to-end models, which input a chest CT image and output a segmented lesion mask. However, features of COVID-19 infections distributed on the lung CT scans are of great intricacy and profound changes, making segmenting COVID-19 challenging. Boundaries of ground-glass opacity are blurred and characterized by low contrast. Pulmonary consolidations are tinier than the whole CT images, which quickly leads to false-negative results. Hence, our proposed model extended the original UNet architecture to an advanced framework by adopting the combination of max-pooling and average-pooling to retain more information, introducing bridge-connected layers to replace the direct skip connection to alleviate the semantic gap between the encoder and decoder, and utilizing multi-scale pyramid pooling module to act as the bottleneck of UNet to fit complex features of COVID-19 infections. In addition, to get high accuracy in segmenting COVID-19 lesions from chest CT scans, a dual attention module containing position and channel attention blocks is fused to the proposed method to enhance the feature representations. Models with a strong capacity for representations and allocating proper loss functions can precisely segment complex COVID-19 infections from lung CT images.

Experiments on the testing set show that the proposed segmentation model achieves high performance competed quantitatively and qualitatively with state-of-the-art methods. Despite this, it must be noted that our model meanwhile owns its limitations. For example, in Figs. 6 and 7, the mask of segmenting results of COVID-19 lesions are not completely matched the corresponding ground truth. Even a few isolated points of COVID-19 lesions with tiny sizes cannot be segmented or segmented with errors. Hence, to precisely segment lesions from CT images, one model must integrate novel, power learning techniques, such as transformer, knowledge distillation, unsupervised learning, and even one-shot learning, and combine the complete CT data to enhance representation capacity, which will be our future research works.

In a nutshell, the contributions provided by this paper are as follows: An end-to-end deep learning method, DAF-UNet, is proposed for segmenting COVID-19 infections from chest CT scans. Bridge-connected layers are used to alleviate the semantic gap between the encoder and decoder. The multiscale pyramid pooling module acts as the bottleneck to fit the complex features of COVID-19 lesions in chest CT images. Dual attention feature fusion (DAFF) comprising channel attention and position attention is adopted to extract the long-dependency contextual features and enhance the learning capacity for the deep neural network model.

5. Conclusions

To sum up, by promoting the basic UNet to advanced architecture and then dual attention blocks being used to enhance the feature representation further, a dual attention fusion UNet referred to as DAF-UNet, is proposed for COVID-19 lesion segmentation from chest CT scans. The proposed DAF-UNet was first pre-trained on a pseudo annotation CT dataset containing lots of samples and then fine-tuned on an available COVID-19 CT dataset with high-quality but limited training samples. Quantitative and qualitative evaluations on the testing set demonstrate that the proposed DAF-UNet can accurately segment COVID-19 infections from CT images and attain high performance than other methods in our study. Our DAF-UNet has excellent potential to assist disease diagnoses and assessments in the fight against the COVID-19 epidemic.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported in part by the Doctoral Research Foundation Project of Tongren University under Grant trxyDH2222.

References

- [1] D. Benvenuto, M. Giovanetti, M. Salemi, et al., The global spread of 2019-nCoV: a molecular evolutionary analysis, *Pathog Glob Health* **114**(2) (2020), 64–67.
- [2] C. Wang, P. P. Horby, F. G. Hayden, et al., A novel coronavirus outbreak of global health concern, *The Lancet* **395**(10223) (2020), 470–473.
- [3] L. Yan, H. T. Zhang, Y. Xiao, et al., Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan, (2020) MedRxiv.
- [4] C. Huang, Y. Wang, X. Li, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *The Lancet* **395**(10223) (2020), 497–506.
- [5] T. Ai, Z. Yang, H. Hou, et al., Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, *Radiology* **296**(2) (2020), E32–E40.
- [6] G. Wang, H. Yu and B. De Man, An outlook on X-ray CT research and development, *Med Phys* **35**(3) (2008), 1051–1064.
- [7] D. Dong, Z. Tang, S. Wang, et al., The role of imaging in the detection and management of COVID-19: a review, *IEEE Reviews in Biomedical Engineering* **14** (2020), 16–29.
- [8] H. Shi, X. Han, N. Jiang, et al., Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study, *The Lancet Infectious Diseases* **20**(4) (2020), 425–434.
- [9] A. Oulefki, S. Agaian, T. Trongtirakul, et al., Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, *Pattern recogn* **114** (2021), 107747.
- [10] MY. Ng, EYP. Lee, J. Yang, et al., Imaging profile of the COVID-19 infection: radiologic findings and literature review, *Radiology* **2**(1) (2020), e200034.
- [11] A. Oulefki, S. Agaian, T. Trongtirakul, et al., Virtual Reality visualization for computerized COVID-19 lesion segmentation and interpretation, *Biomed Signal Proces* **73** (2022), 103371.
- [12] G. Litjens, T. Kooi, B.E. Bejnordi, et al., A survey on deep learning in medical image analysis, *Med Image Anal* **42** (2017), 60–88.
- [13] G. Wang, M. Kalra and CG. Orton, Machine learning will transform radiology significantly within the next 5 years, *Med Phys* **44**(6) (2017), 2041–2044.
- [14] J. Long, E. Shelhamer and T. Darrell, Fully convolutional networks for semantic segmentation, In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (2015), 3431–3440.
- [15] O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, In: *International Conference on Medical image computing and computer-assisted intervention, Springer Cham* (2015), 234–241.
- [16] L. Li, L. Qin, Z. Xu, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, *Radiology* (2020), 200905-0.
- [17] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, et al., A Nested U-Net Architecture for Medical Image Segmentation, arXiv preprint arXiv: 1807.10165.
- [18] S. Jin, B. Wang, H. Xu, et al., AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks, MedRxiv (2020).
- [19] X. Guo, Y. Lei, P. He, et al., An ensemble learning method based on ordinal regression for COVID-19 diagnosis from chest CT, *Phys Med Biol* **66**(24) (2021), 244001.
- [20] D. P. Fan, T. Zhou, G. P. Ji, et al., Inf-net: Automatic covid-19 lung infection segmentation from ct images, *IEEE Trans Med Imaging* **39**(8) (2020), 2626–2637.
- [21] G. Wang, X. Liu, C. Li, et al., A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images, *IEEE Trans Med Imaging* **39**(8) (2020), 2653–2663.
- [22] Y. J. Ma, Y. P. Feng, P. He, et al., Segmenting lung lesions of COVID-19 from CT images via pyramid pooling improved Unet, *Biomed Phys Eng Expr* **7**(4) (2021), 045008.
- [23] X. Ding, J. Xu, J. Zhou, et al., Chest CT findings of COVID-19 pneumonia by duration of symptoms, *Eur J Radiol* **127** (2020), 109009.
- [24] Y. Wang, Y. Zhang, Y. Liu, et al., Does non-COVID-19 lung lesion help? investigating transferability in COVID-19 CT image segmentation, *Comput Meth Prog Bio* **202** (2020), 106004.

- [25] J. Ma, Y. Wang, X. An, et al., Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation, *Med phys* **48**(3) (2021), 1197–1210.
- [26] L. Tychsen-Smith and L. Petersson, Improving object localization with fitness nms and bounded iou loss, In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), 6877–6885.
- [27] S. Bruch, X. Wang, M. Bendersky, et al., An analysis of the softmax cross entropy loss for learning-to-rank with binary relevance, In: *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval* (2019), 75–78.
- [28] G. Huang, Z. Liu, L. Van Der Maaten, et al., Densely connected convolutional networks, In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), 4700–4708.
- [29] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, In: *International conference on machine learning*, PMLR (2015), 448–456.
- [30] Y. Liu, X. Wang, L. Wang L, et al., A modified leaky ReLU scheme (MLRS) for topology optimization with multiple materials, *Appl Math Comput* **352** (2019), 188–204.
- [31] H. Zhao, J. Shi, X. Qi, et al., Pyramid scene parsing network, In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), 2881–2890.
- [32] J. Fu, J. Liu, H. Tian, et al., Dual attention network for scene segmentation, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), 3146–3154.
- [33] J. Chen, Y. Lu, Q. Yu, et al., Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306.
- [34] COVID-19 CT Segmentation Dataset. Accessed: (Apr 2020) [Online]. Available: <https://medicalsegmentation.com/covid19/>
- [35] A. Paszke, S. Gross, F. Massa, et al., Pytorch: An imperative style, high-performance deep learning library, *Advances in Neural Information Processing Systems* **32** (2019), 8026–8037.
- [36] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
- [37] R. P. Loce and E. R. Dougherty, Mean-absolute-error representation and optimization of computational-morphological filters, *Graphical Models and Image Processing* **57**(1) (1995), 27–37.
- [38] S. Cai, Y. Tian, H. Lui, et al., Dense-UNet: a novel multiphoton *in vivo* cellular image segmentation model based on a convolutional neural network, *Quant Imag Med Surg* **10**(6) (2020), 1275.
- [39] J. Schlemper, O. Oktay, M. Schaap, et al., Attention gated networks: Learning to leverage salient regions in medical images, *Med Image Anal* **53** (2019), 197–207.