**ORIGINAL PAPER**

# DBF-Net: a semi-supervised dual-task balanced fusion network for segmenting infected regions from lung CT images

Xiaoyan Lu[1] · Yang Xu[1,2] · Wenhao Yuan[1]

## Abstract
Accurate segmentation of infected regions in lung computed tomography (CT) images is essential to improve the timeliness and effectiveness of treatment for coronavirus disease 2019 (COVID-19). However, the main difficulties in developing of lung lesion segmentation in COVID-19 are still the fuzzy boundary of the lung-infected region, the low contrast between the infected region and the normal trend region, and the difficulty in obtaining labeled data. To this end, we propose a novel dual-task consistent network framework that uses multiple inputs to continuously learn and extract lung infection region features, which is used to generate reliable label images (pseudo-labels) and expand the dataset. Specifically, we periodically feed multiple sets of raw and data-enhanced images into two trunk branches of the network; the characteristics of the lung infection region are extracted by a lightweight double convolution (LDC) module and fusiform equilibrium fusion pyramid (FEFP) convolution in the backbone. According to the learned features, the infected regions are segmented, and pseudo-labels are made based on the semi-supervised learning strategy, which effectively alleviates the semi-supervised problem of unlabeled data. Our proposed semi-supervised dual-task balanced fusion network (DBF-Net) creates pseudo-labels on the COVID-SemiSeg dataset and the COVID-19 CT segmentation dataset. Furthermore, we perform lung infection segmentation on the DBF-Net model, with a segmentation sensitivity of 70.6% and specificity of 92.8%. The results of the investigation indicate that the proposed network greatly enhances the segmentation ability of COVID-19 infection.

**Keywords** COVID-19 · Image segmentation · DBF-Net · Semi-supervised learning

## 1 Introduction

In December 2019, the COVID-19 outbreak emerged and then spread locally around the world. It not only causes a drain on the world economy but also poses a threat to the lives of human beings all over the world (Nicola et al. 2020). Therefore, early detection, early diagnosis, and early treatment are important methods for improving patient's survival rates (Yan et al. 2020). However, it is difficult to confirm the severity of infection by direct judgmental analysis of patients (Munusamy et al. 2021); thus, doctors diagnose COVID-19 lung involvement by CT images. CT technology can show some distinctive features, including ground-glass opacity (GGO), pulmonary fibrosis (PF), pleural effusion (PE), and pulmonary consolidation (PC), which has important research value and practical significance for the early diagnosis of lung lesions (Shi et al. 2020; Kanne 2019). In diagnosis, doctors need to rely on imagination to convert the 2D CT patient images into 3D images to obtain the location and size of pathological tissues. Nevertheless, with the large increase in the number of confirmed and suspected cases of COVID-19, doctors need to spend considerable time and effort manually labeling the CT lesion area. Therefore, computer-aided systems can be used to help doctors diagnose lung infection and quantitatively evaluate the effect before and after treatment. It not only improves the medical image interpretation efficiency of doctors but also strengthens their clinical diagnosis ability and improves the patient cure rate.

✉ Yang Xu
  xuy@gzu.edu.cn

  Xiaoyan Lu
  luxiaoyan6169@163.com

  Wenhao Yuan
  ywh12123@163.com

1   College of Big Data and Information Engineering, Guizhou University, Guiyang, Guizhou, People's Republic of China

2   Guiyang Aluminum Magnesium Design and Research Institute Co., Ltd, Guiyang, Guizhou, People's Republic of China

Ⓢ Springer

Recently, deep learning-based computerized imaging diagnostic systems have been used to help examine infected patients, which utilize models to obtain features to identify areas of lung infection. For example, Kumar Singh et al. (2021) proposed a model for segmenting the COVID-19 infections of lung CT images based on a receptive-field-aware (RFA) module, called LungINFseg. The RFA module can enlarge the receptive field of the segmentation models and learn context information. Wang et al. (2020) proposed a deep convolutional neural network (COVID-Net), which aims to screen patients with suspected infection by identifying obvious signs of COVID-19 from chest X-rays. To relieve the diagnostic pressure caused by the lack of labeled data, Zhou et al. (2019) adopted a self-supervised learning strategy to effectively improve the utilization rate of a mass of unlabeled images. Moreover, they used model genesis to achieve 3D transfer learning of medical images. Alhudhaif et al. (2021) designed a generalized convolutional neural network capable of identifying COVID-19 through feature extraction from chest X-ray images. Wang et al. (2021) proposed a DeepSC-COVID model for 3D lesion classification and segmentation COVID-19 and realized assisted diagnosis of COVID-19 through multitask learning.

Despite the emergence of intelligent diagnostic systems for COVID-19 and the active exploration of lung infection regions, there are still many challenges. First, feature analysis and information extraction are affected due to the large morphological difference and variable location of infected regions in lung CT images. Second, compared with natural images, CT images have low contrast and are susceptible to noise, resulting in blurred edges between different tissues or between tissues and lesions, which increases the segmentation difficulty. In addition, data collection and labeling for the study are difficult. Therefore, making reliable pseudo-labels is essential to assisting doctors in diagnosing patients.

To deal with the above challenges, we propose a novel semi-supervised dual-task balanced fusion network (DBF-Net) to produce high-quality pseudo-labels based on lung infection areas in CT images. Inspired by the method of the radiologist in detecting the infected region, we first roughly locate the infected region and then further determine the outline of the infected area according to the local characteristics. In our view, relatively clear regions and boundaries are key features in determining whether the lung is infected. Our deep learning model extracts boundary information layer by layer through a fusiform equilibrium fusion pyramid. The original CT image and the enhanced image are fed into both network branches for training and learning, thus extracting more complete image information. In addition, we design a semi-supervised learning framework to combine unlabeled and labeled data for training, and effectively make pseudo-labels to expand the infected region segmentation training

dataset. In summary, our research mainly includes the following threefold:

(1) We design a novel deep learning network (DBF-Net). The dual-task learning of lung CT image segmentation is realized by adopting a unique dual-branch training method. By using a lightweight double convolution module for down-sampling, this module is a simpler and more effective down-sampling method than the ordinary down-sampling module.

(2) We propose the fusiform equilibrium fusion pyramid in the down-sampling of our model for feature extraction layer by layer. First, our pyramid convolution is divided into different levels, and each level corresponds to a convolution kernel of different sizes. Then, the convolution kernel at the top of the pyramid is sequentially divided. The convolution kernel at the bottom of another pyramid realizes feature fusion in sequence. After that, the aggregated features can communicate the context information, reduce the number of convolution parameters and achieve the favorable effect of balanced fusion features.

(3) We combine the image enhancement approach with a semi-supervised learning strategy in our model to generate pseudo-labels, which improves the utilization of unlabeled data by selecting a specific quantity of unlabeled data and labeled data for mixed training each time.

## 2 Related work

### 2.1 Lung CT image segmentation

The diagnostic results of lung CT images can be used as the evaluation basis for patients with COVID-19 (Sluimer et al. 2006; Kamble et al. 2020). Radiologists segmented the lung lesion area by viewing the lung CT images and combining them with clinical information to diagnose COVID-19 patients. After the global outbreak of COVID-19, many researchers have carried out research on COVID-19 lung CT image segmentation based on deep learning. Based on U-Net, Chen et al. (2020) utilized aggregated residual transformations and soft attention mechanism to learn robust and expressive feature representations, thereby improving the model's ability to distinguish various symptoms of COVID-19. Rajamani et al. (2021) proposed a dynamic deformable attention network (DDANet) for COVID-19 lesion semantic segmentation. The model is based on a deformable criss-cross attention block, which continuously learn sparse attention filter offsets to capture sufficient context information and improve segmentation performance. To solve the problem of insufficient training samples, Shan et al. Shan

et al. (2021) proposed a VB-Net model based on "bottleneck structure" to segment COVID-19 CT images and proposed a semi-supervised training strategy of "human-in-the-loop (HITL)" participated by professional doctors to reduce network training time and improve segmentation efficiency. In addition, some studies combine classification and segmentation. Wang et al. (2020) developed a weakly-supervised deep learning framework using 3D CT volumes, which can accurately predict the probability of COVID-19 infection and find lesion regions in chest CT without labeling the lesions for training. The easily-trained and high-performance deep learning algorithm provides a method for quickly identifying COVID-19 patients, which is conducive to controlling the outbreak of SARS-CoV-2. Li et al. (2020) created a fully automated framework for detecting COVID-19 through lung CT, distinguishing community-acquired pneumonia from other non-pneumonic lung diseases.

## 2.2 Semi-supervised learning

Semi-supervised learning (SSL) has been extensively studied in various computer vision tasks. Recently, to reduce the labeling burden, an increasing number of scholars have devoted themselves to the study of deep learning models for semi-supervised medical image segmentation. Existing semi-supervised methods are mainly divided into two categories. The first category is based on pseudo-labels (Fan et al. 2020; Bai et al. 2017), which jointly improves the segmentation model by training labeled images so that unlabeled images are tested on the network to obtain pseudo-labels. Fan et al. (2020) incrementally augmented the training dataset with unlabeled data and then generated pseudo-labels for training. Bai et al. Bai et al. (2017) redefined pseudo-labels by improving pseudo-segment labels, adjusting network parameters, or using conditional random field (CRF). However, this method ignores the pseudo-labels properties, which may not improve the network learning performance. The second class of methods learns from both labeled and unlabeled images, and they usually consist of a supervised loss function for labeled images and an unsupervised regularization loss function for all images. Cui et al. (2019) proposed a consistency loss to exploit unlabeled data and added an exponential moving average to prevent overfitting. Li et al. (2020) introduced more data perturbations and model perturbations on the teacher-student model to construct the consistency of the same input under different perturbations. Chen et al. (2019) simultaneously optimized the supervised segmentation and unsupervised segmentation reconstruction targets, and the reconstruction targets adopted an attention mechanism to separate the image reconstruction regions corresponding to different categories. Nie et al. (2018) proposed a novel deep adversarial network to facilitate partial unlabeled images to approximate labeled images for biological-image segmentation.

## 3 Methods

This section introduces our proposed balanced fusion network (DBF-Net) based on dual-task consistency, as well as corresponding key modules. Our model combines image enhancement and semi-supervised learning framework (Shan et al. 2021) to augment the training dataset with limited labeled data, thereby improving the segmentation accuracy of lung CT images. In addition, we extend DBF-Net to utilize pseudo-labels for segmentation tasks. Experimental comparison with mainstream segmentation algorithms shows the superiority of the proposed algorithm.

### 3.1 Dual-task balanced fusion network (DBF-Net)

The architecture of our DBF-Net is shown in Fig. 1. It adopts the commonly used and efficient encoder-decoder structure for medical image segmentation (Zhou et al. 2020; Liu et al. 2017; Wang et al. 2019). In the encoder, we first simultaneously feed the original image and the enhanced image into the network branch and perform the dimension-raising operation through a 1×1 standard convolution. Then, the lightweight double convolution (LDC) module is used to perform the down-sampling operation to extract image information layer by layer. Meanwhile, the fusiform equilibrium fusion pyramid (FEFP) is embedded behind the LDC module of the two branches. Furthermore, in each FEFP operation, feature fusion of different feature layers is realized with the other branch to reduce information loss. Second, we utilize the attentional feature fusion (AFF) module (Dai et al. 2021) to optimize the results of the double branches, and then use the atrous spatial pyramid pooling (ASPP) module (Chen et al. 2017) to increase the convolutional receptive field to effectively learn the edge features of the lung-infected regions. Finally, we use a decoding structure similar to U-Net to complete up-sampling and obtain the final mask of the lung-infected regions.

### 3.2 Lightweight double convolution (LDC)

The lightweight module reduces the time and space complexity when extracting image features (Li et al. 2019). To improve the rate of feature extraction and transfer, the number of parameters is reduced to some extent. We extract features from the input image through a lightweight double convolution module, the structure is shown in Fig. 2.

Specifically, we take the convolution operation, batch normalization, and activation function as a basic processing unit and integrate the max-pooling operation in the middle layer
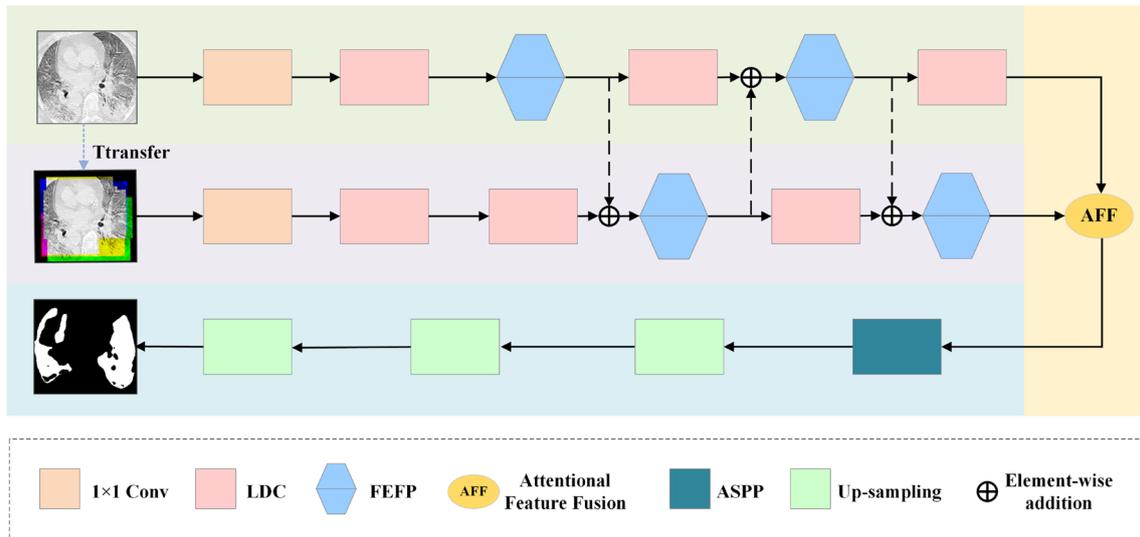
**Fig. 1** The architecture of our proposed DBF-Net, which consists of a lightweight double convolution (LDC) module connected to the fusiform equilibrium fusion pyramid (FEFP) convolution
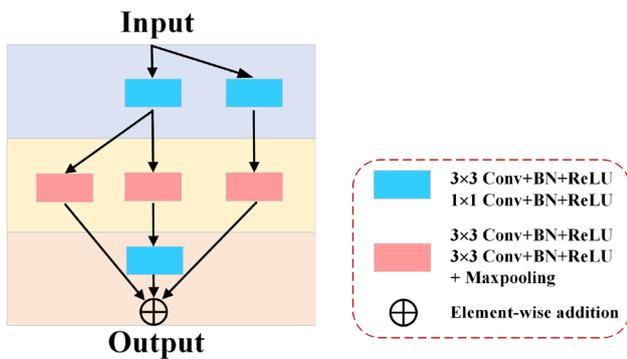


**Fig. 2** The structure of lightweight double convolution module (LDC), which is used as the down-sampling operation module

to achieve down-sampling. From the global perspective, the residual structure we propose preserves as much of the boundary information lost due to ordinary down-sampling operations as possible while featuring extraction and transmission, which is critical for accurate medical image segmentation.

The final output feature F can be formulated as follows:

$$F = f_2\big(f_1(x)\big) + f_1\big[f_2\big(f_1(x)\big)\big] + f_2\big(f_1(x)\big)$$
$$= 2f_2\big(f_1(x)\big) + f_1\big[f_2\big(f_1(x)\big)\big]$$
(1)

In this formula, $f_1$ and $f_2$ can be expressed as:

$$f_1 = \mathrm{ReLU}\big[\mathrm{BN}\big(\mathrm{Conv}_3(x)\big)\big] + \mathrm{ReLU}\big[\mathrm{BN}\big(\mathrm{Conv}_1(x)\big)\big] \quad (2)$$

$$f_2 = \mathrm{MP}\big[2\mathrm{ReLU}\big(\mathrm{BN}\big(\mathrm{Conv}_3(x)\big)\big)\big] \quad (3)$$

where x is the input of the LDC module, $\mathrm{Conv}_3(.)$ denotes the 3×3 convolutional layer, $\mathrm{Conv}_1(.)$ represents the 1×1 convolutional layer, and + is the element-wise addition, MP(.) denotes max-pooling.

### 3.3 Fusiform equilibrium fusion pyramid (FEFP)

Different from ordinary pyramidal convolution (Duta et al. 2020) (PyConv), our fusiform equilibrium fusion pyramid (FEFP) module not only contains different levels of kernels with different sizes and depths, but also balances the features extracted from large- and small-scale kernels in symmetric form. Therefore, in addition to expanding the convolution receptive field, FEFP can also capture richer multiscale details than PyConv.

Our FEFP is shown in Fig. 3, which is composed of two symmetrical pyramid splices. To be able to utilize different depth kernels at each level on FEFP, we divide the input feature maps into different groups by grouping convolution and apply the kernel independently for each input feature map group. The input feature map $FM_i$ is divided into two feature blocks $FM_{i_1}$ and $FM_{i_2}$. Each level of $\{1, 2, 3, 4\}$ of the FEFP convolution corresponds to a different space size kernel $\{K_1^2, K_2^2, K_3^2, K_4^2\}$ and $\{K_{1*}^2, K_{2*}^2, K_{3*}^2, K_{4*}^2\}$. The kernels of different depths are obtained by grouping in Fig. 3:

$$\left\{ FM_{i1}, \frac{FM_{i1}}{\left(\frac{K_2^2}{K_1^2}\right)}, \frac{FM_{i1}}{\left(\frac{K_3^2}{K_1^2}\right)}, \frac{FM_{i1}}{\left(\frac{K_4^2}{K_1^2}\right)} \right\}, \left\{ FM_{i2}, \frac{FM_{i2}}{\left(\frac{K_{2*}^2}{K_{1*}^2}\right)}, \frac{FM_{i2}}{\left(\frac{K_{3*}^2}{K_{1*}^2}\right)}, \frac{FM_{i2}}{\left(\frac{K_{4*}^2}{K_{1*}^2}\right)} \right\}.$$

The output feature maps of the two pyramids are $\{FM_{o11}, FM_{o12}, FM_{o13}, FM_{o14}\}$ and $\{FM_{o21}, FM_{o22}, FM_{o23}, FM_{o24}\}$, as well as $\big(FM_{o14} + FM_{o21}\big) + \big(FM_{o13} + FM_{o22}\big) +$
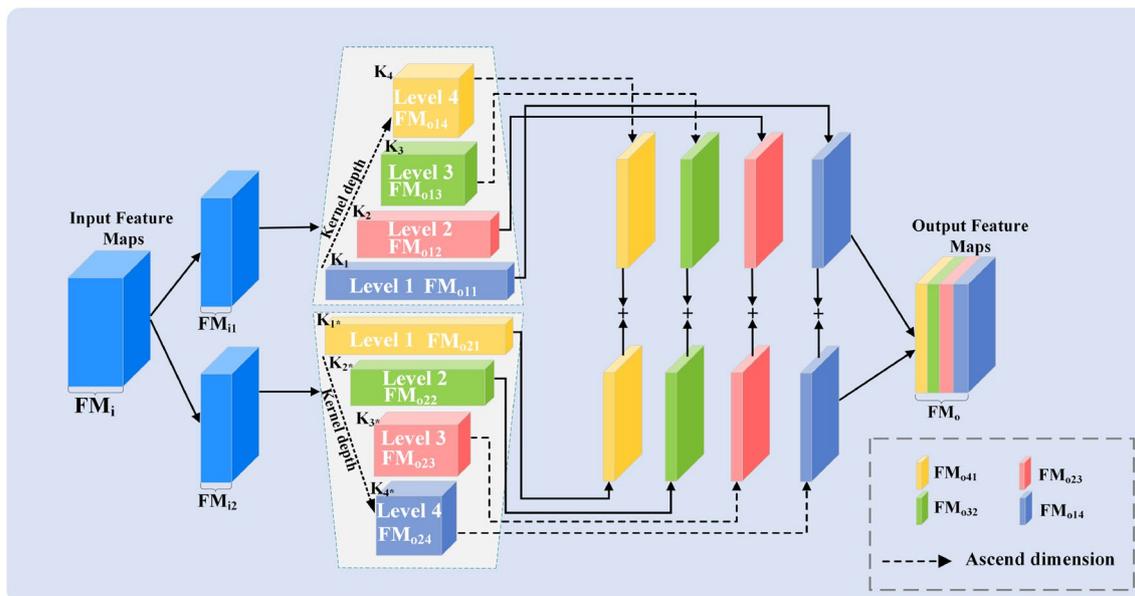
**Fig. 3** The structure of fusiform equilibrium fusion pyramid convolution (FEFP), which is utilized to achieve balanced feature fusion

$\left(FM_{o12} + FM_{o23}\right) + \left(FM_{o11} + FM_{o24}\right) = FM_o$. The convolution kernel at the top of the pyramid and at the bottom of the other pyramid are sequentially combined to achieve feature fusion. Finally, the output feature map $FM_o$ is obtained by connecting the feature maps at each level according to the number of channels.

The FEFP module kernel type is a symmetric pyramid. With the increase in kernel size, the kernel depth decreases from level 1 to level n and vice versa. Kernels of different sizes communicate information to maximize feature complementarity. Through the interconnection of receptive fields of different sizes, the feature fusion of different scale kernels is realized, and infected areas the recognition in CT images is improved.

## 3.4 Semi-supervised learning strategy

Manual labeling of the infected regions in lung CT images is time-consuming and labor-intensive, resulting in very little labeled data. To augment the dataset, we adopt the combination of image enhancement and semi-supervised learning strategy to improve DBF-Net.

The training set is augmented by a small quantity of labeled data to help unlabeled data generate pseudo-labels. First, we send the training set composed of labeled images and enhanced images to pretrain the DBF-Net model. Then, N unlabeled images are randomly fed for prediction to obtain N corresponding pseudo-labels. The training set is mixed with N pseudo-label images and then fed to the network for

training again so that the weight is continuously updated. Repeating the above operations, we periodically feed the training set of labeled images and N unlabeled images and then complete the network training after 200 epochs, thereby generating the desired high-quality pseudo-labels.

Specifically, in the dataset we used, there are 1600 unlabeled images and 100 labeled images. During the experiment, 60 labeled images and corresponding enhanced images were used as the training set, 10 labeled images were used as the verification set, and 30 labeled images were used as the test set. Then, 8 unlabeled images were randomly sent to predict each time, i.e., $N_i = 8$. The semi-supervised learning framework is shown in Fig. 4.

## 3.5 Image enhancement for medical images

Taking different image enhancement methods may affect the network model performance. In general, the fundamental purpose of image processing is to learn the critical information of the image. Because of our needs and the characteristics of medical image processing, we carried out four transformations on lung CT images (Zhou et al. 2019), as shown in Fig. 5.

(1) Nonlinear transformation The pixel value in the CT image is the corresponding value of the X-ray attenuation coefficient of each tissue, also known as the Hounsfield Unit (HU) value. Different HU values correspond to different tissues. The nonlinear function is used to perform the nonlinear transformation on the

**Fig. 4** Semi-supervised DBF-Net architecture diagram, where blue refers to labeled images and enhanced images, yellow refers to unlabeled images, and red refers to our proposed DBF-Net
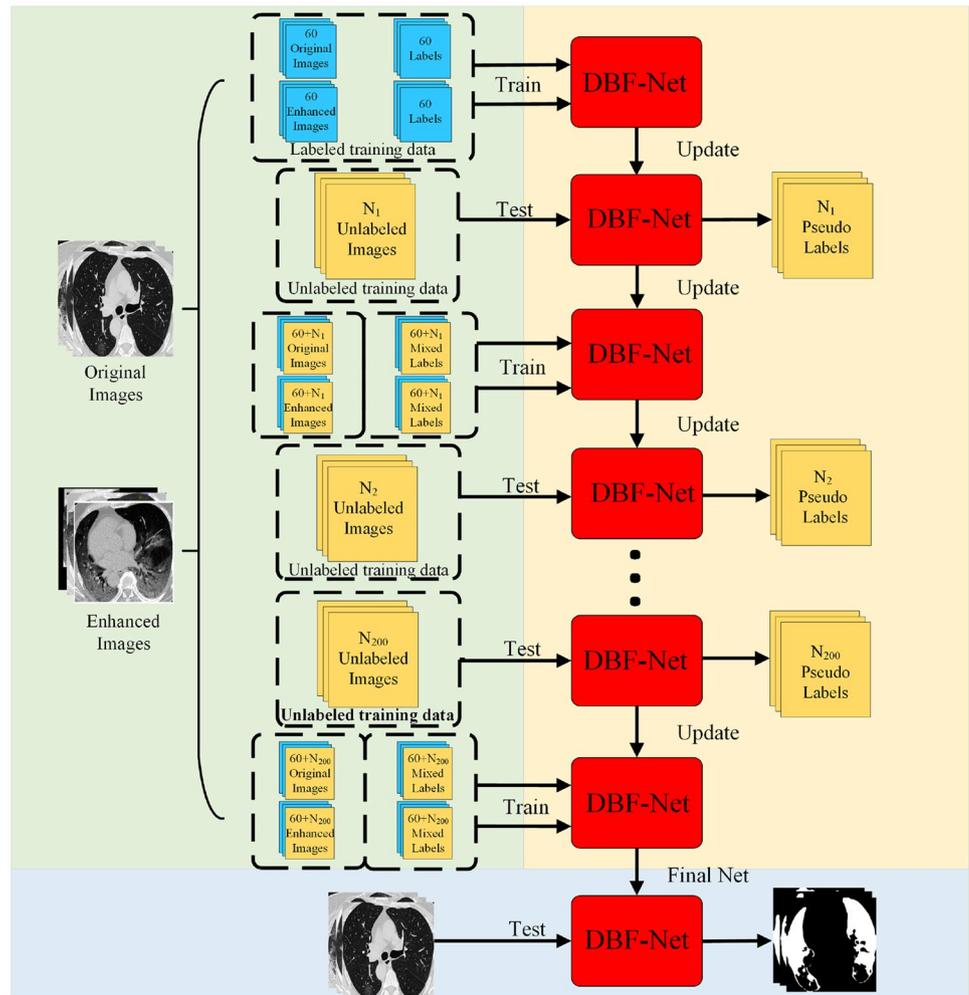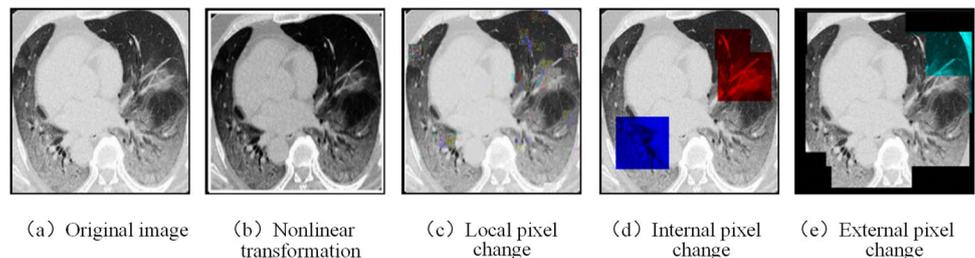


**Fig. 5** Methods of lung CT image enhancement. **a** Original image. **b** Nonlinear transformation. **c** Local pixel change. **d** Internal pixel change. **e** External pixel change



(a) Original image   (b) Nonlinear transformation   (c) Local pixel change   (d) Internal pixel change   (e) External pixel change

input image HU. The global contrast enhancement of the image is realized by adjusting the transformation parameters to identify different tissues.

(2) Local pixel change In CT image A, a small cube c is randomly determined. The pixel position in cube c is randomly scrambled to obtain $c'$, and then c is replaced by $c'$. This process is repeated several times to obtain the transformed CT image $\tilde{A}$. On the premise that the

overall image shape does not change greatly, the model can learn the local structure and texture features.

(3) Internal pixel change In CT image A, two cubes $c_1$ and $c_2$ are randomly selected, and $c_1 \cap c_2 = \emptyset$ is satisfied. The pixel values of two cubes are exchanged, i.e., $c'_1 = c_2$, $c'_2 = c_1$. Then, this process is repeated several times to obtain the transformed image $\tilde{A}$.

(4) External pixel change Irregular masking on the outer edge of the original image prompts the model to analyze the internal structural information to infer the external structure and extract more critical visual features. The overall algorithm flow is shown in Algorithm 1.

---

**Algorithm 1** Semi-supervised DBF-Net

---
**Require:** Labeled data and unlabeled data
**Ensure:** The trained CT segmentation model DBF-Net
1: Unlabeled images are divided into 200 groups, each group is represented as $N_i$, where i=1, 2, . . . 200
2: Send labeled images and enhanced images as training dataset D to train for the pretraining model
3: **repeat**
4:     Send Group i unlabeled images $N_i$ for training to get pseudo-labels $P_i$, and then enhance $N_i$ to get $\bar{N}_i$
5:     Mix the dataset D and $\bar{N}_i$ and send it back to the model for training
6: **until** i=200, all unlabeled images are sent into the model
7: **return** Trained model

---

# 4 Experiments

## 4.1 Lung datasets

In this paper, the COVID-SemiSeg dataset (Fan et al. 2020) and COVID-19 CT segmentation dataset (Milletari et al. 2016) are used to perform the experimentation and comparison with mainstream approaches.

(1) The COVID-SemiSeg dataset is aimed at semi-supervised COVID-19 infection segmentation and 3D CT images from more than 20 COVID-19 patients, and the dataset is extended with the help of many unlabeled CT images.
(2) The COVID-19 CT segmentation dataset consists of 100 labeled axial CT images from over 40 COVID-19 patients. The CT images were all collected by the Italian Society of Medical and Interventional Radiology, and radiologists segmented the CT images based on three labels of ground-glass opacity (GGO), consolidation, and pleural effusion to determine regions of lung infection.

We strictly split the COVID-19 CT segmentation dataset containing 100 labeled axial CT images. Specifically, 60 are used for training, 10 for validation, and 30 for testing. The COVID-SemiSeg dataset contains 1600 unlabeled CT images. We perform training on this dataset and the training set in the COVID-19 CT segmentation dataset following the semi-supervised learning strategy in Sect. 3.4.

## 4.2 Experimental settings

All the experiments of the proposed method DBF-Net are conducted on Intel I7-11700K with NVIDIA RTX3080TI GPU. The development environment is based on the Ubuntu 20.04 operating system, CUDA11.4 + Pytorch1.9, and the programming language is Python 3.8.

Since resizing images will have an impact on image quality, we first resample the original image slices and then crop all slices uniformly to $384 \times 384$. The training is performed with an Adam optimizer with a momentum of 0.9 and a weight decay of 0.0005. The initial learning rate is 0.01, the batch size is set to 4, and a total of 200 epochs are trained.

For image segmentation, the cross-entropy loss function is widely used as the main function. To solve the problem of CT image category imbalance and difficult-to-classify samples, this paper trains the DBF-Net model by combining the Dice loss function and the Focal loss function. Then, the final loss function is:

$$L = \gamma L_{Dice} + (1 - \gamma)L_{Focal}$$
$$= \gamma \left( C - \sum_{c=0}^{C-1} \frac{TP_n(c)}{TP_p(c) + \alpha FN_p(c) + \beta FP_p(c)} \right) \quad (4)$$
$$-(1-\gamma)\frac{1}{N}\sum_{c=0}^{C-1}\sum_{n=1}^{N} g_n(c)\big(1 - P_n(c)\big)^2 log\big(P_n(c)\big)$$

In the equation, $c$ is a specific class; $TP_P(c)$, $FN_P(c)$, $FP_P(c)$ are the true positive rate, false-negative rate, and false-positive rate of the class; $P_n(c)$ refers to when the pixel $n$ is of class $c$; $g_n(c)$ refers to the real situation that pixel $n$ is class $c$; $C$ is the total number of classes; $N$ is the sum of the number of pixels; $\alpha$ and $\beta$ are the penalty weights of false negative and false positive, respectively, set to 0.5; $\gamma$ and $1 - \gamma$ is the weight of Dice loss and Focal loss, while $\gamma$ is set to 0.3.

## 4.3 Evaluation metrics

To evaluate the performance of our proposed method in the lung lesion segmentation task, we used several evaluation metrics. Sensitivity, Specificity, Dice, and Precision were defined as follows:

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$Dice = \frac{2TP}{2TP + FP + FN} \qquad (7)$$

$$Precision = \frac{TP}{TP + FP} \qquad (8)$$

where *TP* refers to the number of infected and accurately predicted regions; *TN* refers to the number of uninfected and accurately predicted regions; *FP* refers to the number of uninfected regions that are wrongly predicted to be infected; *FN* refers to the number of infected regions that are wrongly predicted to be uninfected.

Note that the increment of evaluation metrics in the experimental subjects are computed by following mathematical expression:

$$I = \frac{ES_a - ES_b}{ES_b} \times 100\% \qquad (9)$$

where *I* denotes the incremental ratio of an evaluation metric, $ES_a$ and $ES_b$ denote experimental subjects respectively.

## 4.4 Comparison of segmentation performance for different algorithms

In order to verify the segmentation performance of the proposed DBF-Net, we utilized DBF-Net as the lung CT image segmentation model, and compared it with the existing classical algorithms. The segmentation results are shown in Fig. 6. The first row is the original CT images, the second row is the result of manual marking by radiologists as the evaluation standard; the third row is the segmentation result of FCN-8s (Long et al. 2015); the fourth row is the segmentation result of U-Net++ (Zhou et al. 2019); the fifth row is the segmentation result of the combination of the ResNet50 encoder (He et al. 2016) and the U-Net decoder (ResUNet). The sixth row is the segmentation result of our proposed algorithm DBF-Net.

The experimental results show that compared with the segmentation results of U-Net++ and other algorithms, the segmentation effect of the proposed DBF-Net has the best performance and high image quality. For the same CT

**Fig. 6** From top to bottom, there are several exemplar results in 2D views (**a**) obtained by the corresponding ground truth (**b**) on the COVID-19 CT segmentation dataset, FCN-8s (**c**), U-Net++ (**d**), ResUNet (**e**), and our DBF-Net Model (**f**), where the red and green labels indicate the GGO and consolidation, respectively
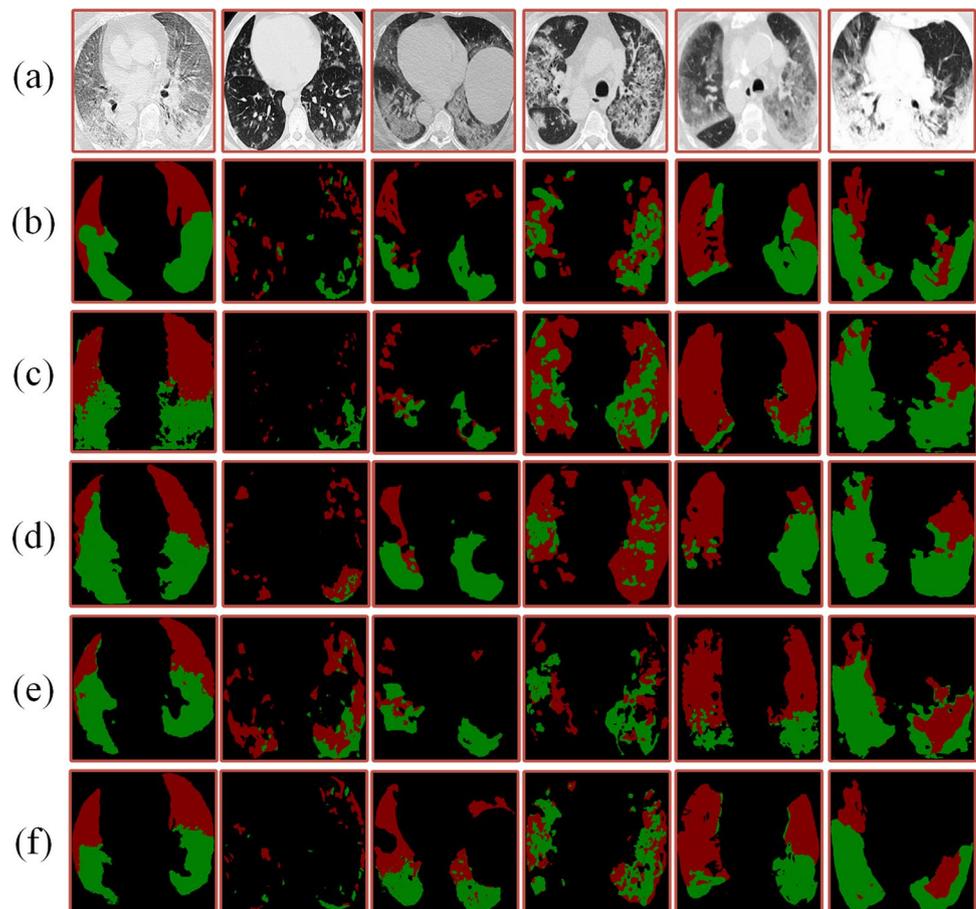
**Table 1** Comparison of quantitative results of different CT segmentation algorithms

| Algorithm | Sensitivity (%) | Specificity (%) | Dice (%) | Precision (%) |
|---|---|---|---|---|
| FCN-8s (Long et al. 2015) | 58.7 | 83.2 | 45.1 | 43.9 |
| U-Net++ (Zhou et al. 2019) | 65.9 | 88.7 | 59.5 | 58.6 |
| ResUNet | 68.5 | 90.6 | 62.9 | 62.3 |
| DBF-Net (ours) | **70.6** | **92.8** | **68.7** | **67.5** |

Bold indicates the best result

**Table 2** Comparison between DBF-Net and other semi-supervised models

| Algorithm | Sensitivity (%) | Specificity (%) | Dice (%) | Precision (%) |
|---|---|---|---|---|
| COPLE-Net (Wang et al. 2020) | 67.8 | 85.2 | 65.4 | 63.9 |
| Semi-Inf-Net (Fan et al. 2020) | 69.7 | 92.3 | **70.8** | 67.1 |
| DBF-Net (ours) | **70.6** | **92.8** | 68.7 | **67.5** |

Bold indicates the best result

images, the FCN-8s segmentation effect is the worst, and the boundary segmentation is relatively rough; U-Net++ and ResUNet have different degrees of over-segmentation of the image. However, our proposed algorithm performs significantly in segmenting the boundary contour of the lung lesion region, with fewer incorrectly segmented regions, and the segmentation effect is relatively close to the image manually labeled by the radiologist.

Although the subjective evaluation is simple and direct, it is susceptible to subjective factors, so it is still necessary to quantitatively evaluate the segmentation results. The quantitative results of different segmentation algorithms are shown in Table 1. It can be seen that the DBF-Net performance is better than that of the other models in four evaluation metrics. Its sensitivity reaches 70.6%, specificity reaches 92.8%, the Dice coefficient reaches 68.7%, precision reaches 67.5%, and the segmentation effect is relatively better. To sum up, under the premise of the semi-supervised learning strategy, the proposed DBF-Net model has a great improvement in CT image segmentation performance.

### 4.5 Performance comparison of different semi-supervised models

In this section, we compare the performance of DBF-Net to recent semi-supervised models. This concludes COPLE-Net (Wang et al. 2020) and Semi-Inf-Net (Fan et al. 2020). Table 2 presents the segmentation performance of the competing approaches across different measures by implementing them under the same experimental settings. COPLE-Net performed relatively poorly on different measures. Compared with COPLE-Net, Semi-Inf-Net showed better
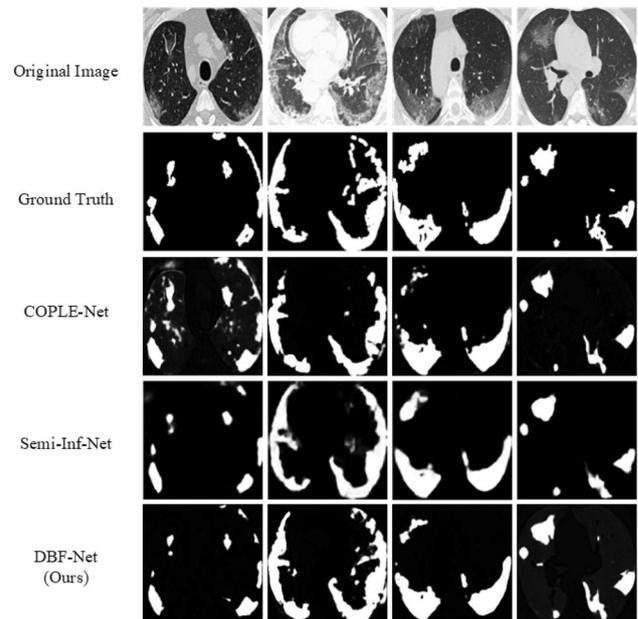


**Fig. 7** Visual comparison of segmentation performance for different semi-supervised models

performance improvements. However, the multistage training of Semi-Inf-Net limited the realization of the optimal performance. In contrast with COPLE-Net, DBF-Net improved sensitivity by 4.1%, specificity by 8.9%, the Dice coefficient by 5.0% and precision by 5.6%. It attained robust segmentation performance with performance improvements over competing models. Fig. 7 shows a graphical comparison of the segmentation results of different real-world COVID-19 axial slices.

**Table 3** Performance of the network with different blocks

| Base Backbone | LDC | FEFP | Image Enhancement | Sensitivity(%) | Specificity(%) | Dice(%) | Precision(%) |
|---|---|---|---|---|---|---|---|
| √ | | | | 57.2 | 84.5 | 54.1 | 52.7 |
| √ | | √ | | 62.8 | 86.2 | 62.3 | 59.4 |
| √ | | √ | √ | 65.7 | 88.6 | 64.5 | 63.9 |
| | √ | √ | | 68.4 | 90.5 | 67.8 | 65.6 |
| | √ | √ | √ | **70.6** | **92.8** | **68.7** | **67.5** |

Bold indicates the best result
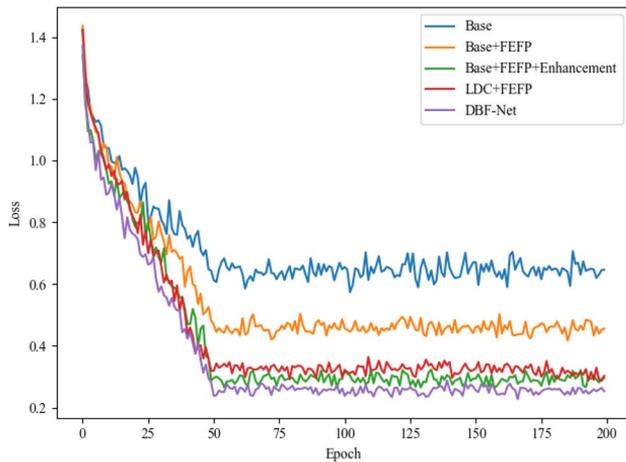


**Fig. 8** Training progress of different module combinations

## 4.6 Ablation experiments

In this section, for a deeper analysis of the DBF-Net performance, ablation experiments were performed to enable understanding of the model behavior under different settings. In these experiments, we select U-Net as the base backbone for our DBF-Net.

(1) Impact of different modules on model performance

Table 3 lists the impact of different modules on model performance. Based on U-Net, our FEFP module and image enhancement are added one after another, thus demonstrating the effectiveness of the FEFP module's multiscale extraction features and image enhancement methods for lung image segmentation. Then, combined with the semi-supervised learning strategy, we replace the down-sampling structure of U-Net with LDC and FEFP modules to form DBF-Net, while adding image enhancement methods to improve the network effect. Finally, pseudo-labels are generated, and DBF-Net is used to achieve lung lesion segmentation.

As seen in the first and second rows of the table, the performance of the FEFP module is improved by up to 15.2% based on U-Net, which makes the model more accurate in segmenting the infected region. The second and fourth rows of the table show that the LDC module has a maximum performance improvement of 10.4% compared to the original down-sampling module in U-Net, which is necessary for performance improvement. The fourth and fifth rows of the table show that the image enhancement approach for medical images has a maximum performance improvement of 3.2% over DBF-Net, which can effectively improve the segmentation effect of the proposed model.

To further visualize the segmentation model performance, we plot the loss curves when different modules are combined for training, as shown in Fig. 8. During the training process, with the superposition of modules, the convergence speed of the network in this paper gradually accelerates, and the accuracy after convergence is relatively high. Therefore, our proposed DBF-Net is easier to train and can better localize the lung lesion region.

(2) Impact of image enhancement on model performance

To reflect the important role of image enhancement in our model, we first combine two original images, one original image, and an enhanced image, and then send them to DBF-Net for performance comparison. The comparison results

**Table 4** Performance comparison of different image combination methods

| Image combination | Sensitivity (%) | Specificity (%) | Dice (%) | Precision (%) |
|---|---|---|---|---|
| Two original images | 68.2 | 90.3 | 66.4 | 65.9 |
| Original and enhanced images | **70.6** | **92.8** | **68.7** | **67.5** |

Bold indicates the best result

**Fig. 9** Visual comparison of segmentation effects for different image combinations. **a** Original image. **b** Ground truth. **c** Segmentation effect of two original image combinations. **d** Segmentation effect of a combination of the original image and enhanced image
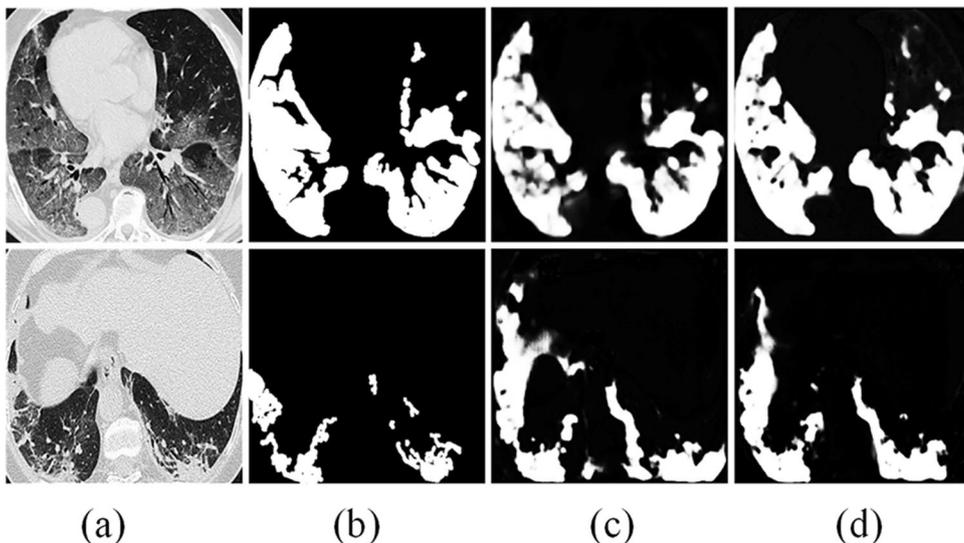


(a)　　　　(b)　　　　(c)　　　　(d)

**Table 5** Comparison between FEFP and PyConv

| Pyramid type | Sensitivity(%) | Specificity(%) | Dice(%) | Precision(%) |
|---|---|---|---|---|
| PyConv | 67.2 | 88.5 | 64.9 | 63.7 |
| FEFP (ours) | **70.6** | **92.8** | **68.7** | **67.5** |

Bold indicates the best result

are shown in Table 4. It can be seen that the original image is transformed by image enhancement dedicated to medical image segmentation, and then sent to DBF-Net, and the network segmentation effect is better enhanced (sensitivity 3.5%; specificity 2.8%; Dice 3.5%; precision 2.4%). Due to the uniqueness of the image enhancement method, the

structure and texture features of lung CT images are well highlighted in model segmentation, and the definition and accuracy of pseudo-labels are improved. The corresponding visualization is shown in Fig. 9.

(3)　Impact of different pyramids on model performance

The main purpose of this experiment was to investigate the impact of different pyramids on CT image segmentation performance. We conducted a comparison experiment between the classical PyConv (Duta et al. 2020) and the proposed FEFP module for the segmentation task, and the experimental results are shown in Table 5. Each evaluation metric of the FEFP module is better than pyramid convolution. Given the complex texture of a CT image and its susceptibility to noise, when a single PyConv is used for

**Fig. 10** Visual comparison of the impact of semi-supervised learning strategy on the model. **a** Original image. **b** Ground truth. **c** Pseudo-labels generated by training with 60 labeled images. **d** Pseudo-labels generated by training with mixed images that include 60 labeled images and 1600 unlabeled images



(a)　　　　(b)　　　　(c)　　　　(d)
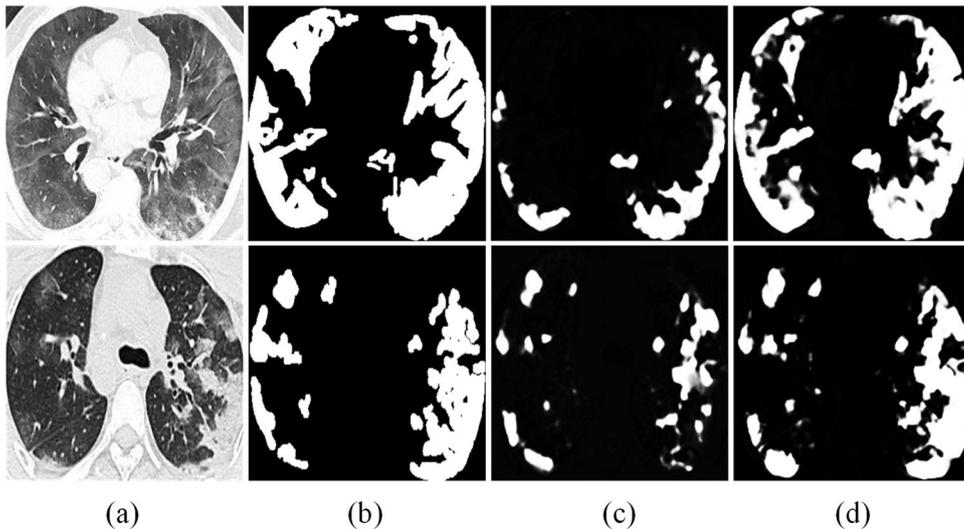
down-sampling, the feature extraction of the infected region of the CT image is insufficient, and it is difficult for the model to better learn the edge features, which reduces the training accuracy. Our FEFP in this paper communicates pixel information by fusing the features of different feature layers. Compared with PyConv, the performance has a minimum improvement of 4.9% in all metrics.

(4) Impact of semi-supervised learning strategy on model performance

The model combined with the semi-supervised learning strategy uses unlabeled images to produce pseudo-labels, thus helping to significantly reduce manual labeling costs. To investigate the impact of the semi-supervised learning strategy on model performance, we put 60 labeled images and mixed images (including labeled images and unlabeled images) into the model for fully-supervised and semi-supervised training, respectively, under the assumption of insufficient manual labeling. In the network input, the original image is enhanced by the image enhancement method to ensure the consistency of the experimental conditions. The experimental results are shown in Fig. 10.

Under the premise of a small number of manually labeled data, using only labeled data will lead to overfitting of the model. Thus, incomplete segmentation occurs when predicting in the testing dataset. In contrast, we use a semi-supervised learning strategy to generate pseudo-labels, which makes up for the insufficient data problem to avoid the overfitting phenomenon caused by model training. Therefore, in the same training epochs, DBF-Net combined with semi-supervised learning strategy can obtain more complete segmentation and higher quality pseudo-labels.

(5) Impact of different training scales on model performance

In this experiment, we use different training scales for DBF-Net to compare the quality of pseudo-labels. The qualitative results are shown in Fig. 11. Initially, we selected 32 unlabeled CT images and 60 labeled images to train together

**Table 6** Comparison of Segmentation performance of different training scales

| Training scale | Sensitivity (%) | Specificity (%) | Dice (%) | Precision (%) |
|---|---|---|---|---|
| 60/8 | 70.6 | 92.8 | 68.7 | 67.5 |
| 60/16 | 67.2 | 88.9 | 63.4 | 62.1 |
| 60/32 | 61.5 | 81.3 | 56.9 | 55.7 |

and found that the edge information of the infected area was less extracted, and the images were blurred; then, we tested 16 unlabeled and 8 unlabeled images with labeled images. After experiments, it was found that the training effect of the combination of 8 unlabeled images and 60 labeled images is better than that of 16 unlabeled and 60 labeled images, it has more accurate boundaries, and the segmentation effect is significantly better. The quantitative results are shown in Table 6. From the perspective of various metrics, the combination of 8 unlabeled images and 60 labeled images has the best effect, which can accurately segment the GGO and consolidation infections.

## 5 Conclusion

In this paper, we propose a novel semi-supervised dual-task balanced fusion network model (DBF-Net), which can help doctors identify infected regions in CT images of COVID-19 patients and reduce the variability of manual diagnosis. The model utilizes a lightweight double convolution module and a fusiform equilibrium pyramid convolution for down-sampling to maximize the localization of infected regions and combines a semi-supervised learning strategy to alleviate the shortage of labeled data. Additionally, we adopt an image enhancement method specifically for medical images to extract more critical visual features and obtain richer pixel information. A series of experimental results on the test set show that the DBF-Net model is superior to other segmentation models with three evaluation metrics Sensitivity, Specificity, and Precision. The proposed algorithm

**Fig. 11** Comparison of visual effects of segmentation results of different training scales



CT Images    Ground Truth    60/8    60/16    60/32
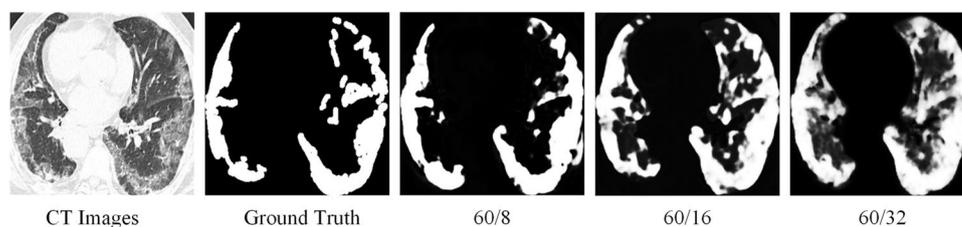
is highly competitive in segmenting the COVID-19 lung infected regions. In future work, we will continue to improve the DBF-Net segmentation model, such as combining segmentation with vision transformer, thus solving the problems of little data and inaccurate lesion localization. This can not only assist doctors in clinical diagnosis but also have important implications for medical research in the big data era.

## Declarations

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

## References

Alhudhaif A, Polat K, Karaman O (2021) Determination of covid-19 pneumonia based on generalized convolutional neural network model from chest x-ray images. Expert Syst Appl 180:115141

Bai W, Oktay O, Sinclair M, Suzuki H, Rajchl M, Tarroni G, Glocker B, King A, Matthews PM, Rueckert D (2017) Semi-supervised learning for network-based cardiac mr image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 253–260 . Springer

Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848

Chen S, Bortsova G, García-Uceda Juárez A, Tulder Gv, Bruijne Md (2019) Multi-task attention-based semi-supervised learning for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 457–465 . Springer

Chen X, Yao L, Zhang Y (2020) Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images. arXiv preprint arXiv:2004.05645

Cui W, Liu Y, Li Y, Guo M, Li Y, Li X, Wang T, Zeng X, Ye C (2019) Semi-supervised brain lesion segmentation with an adapted mean teacher model. In:International Conference on Information Processing in Medical Imaging, pp. 554–565 . Springer

Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K (2021) Attentional feature fusion. In:Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3560–3569

Duta IC, Liu L, Zhu F, Shao L (2020) Pyramidal convolution: rethinking convolutional neural networks for visual recognition. arXiv preprint arXiv:2006.11538

Fan D-P, Zhou T, Ji G-P, Zhou Y, Chen G, Fu H, Shen J, Shao L (2020) Inf-net: automatic covid-19 lung infection segmentation from ct images. IEEE Trans Med Imaging 39(8):2626–2637

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In:Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778

Kamble B, Sahu SP, Doriya R (2020) A review on lung and nodule segmentation techniques. Adv Data Inform Sci. https://doi.org/10.1007/978-981-15-0694-9_52

Kanne JP (2020) Chest CT findings in 2019 novel coronavirus (2019-ncov) infections from Wuhan, china: key points for the radiologist. Radiology. https://doi.org/10.1148/radiol.2020200241

Kumar Singh V, Abdel-Nasser M, Pandey N, Puig D (2021) Lunginfseg: segmenting covid-19 infected regions in lung ct images based on a receptive-field-aware deep learning framework. Diagnostics 11(2):158

Li H, Xiong P, Fan H, Sun J (2019) Dfanet: Deep feature aggregation for real-time semantic segmentation. In:Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9522–9531

Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q et al (2020) Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest CT. Radiology. https://doi.org/10.1148/radiol.2020200905

Li X, Yu L, Chen H, Fu C-W, Xing L, Heng P-A (2020) Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Trans Neural Netw Learn Syst 32(2):523–534

Liu S, Xu D, Zhou SK, Mertelmeier T, Wicklein J, Jerebko A, Grbic S, Pauly O, Cai W, Comaniciu D (2017) 3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes. arXiv preprint arXiv:1711.08580

Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In:Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440

Milletari F, Navab N, Ahmadi S-A (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. In:2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE

Munusamy H, Muthukumar KJ, Gnanaprakasam S, Shanmugakani TR, Sekar A (2021) Fractalcovnet architecture for covid-19 chest x-ray image classification and CT-scan image segmentation. Biocybern Biomed Eng 41(3):1025–1038

Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, Agha M, Agha R (2020) The socio-economic implications of the coronavirus pandemic (covid-19): a review. Int J Surg Open 78:185–193

Nie D, Gao Y, Wang L, Shen D (2018) Asdnet: attention based semi-supervised deep networks for medical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 370–378 Springer

Rajamani KT, Siebert H, Heinrich MP (2021) Dynamic deformable attention network (ddanet) for covid-19 lesions semantic segmentation. J Biomed Inform 119:103816

Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, Xue Z, Shen D, Shi Y (2021) Abnormal lung quantification in chest ct images of covid-19 patients with deep learning and its application to severity prediction. Med Phys 48(4):1633–1645

Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C (2020) Radiological findings from 81 patients with covid-19

pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 20(4):425–434

Sluimer I, Schilham A, Prokop M, Van Ginneken B (2006) Computer analysis of computed tomography scans of the lung: a survey. IEEE Trans Med Imaging 25(4):385–405

Wang G, Shapey J, Li W, Dorent R, Dimitriadis A, Bisdas S, Paddick I, Bradford R, Zhang S, Ourselin S (2019) Automatic segmentation of vestibular schwannoma from t2-weighted mri by deep spatial attention with hardness-weighted loss. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 264–272 . Springer

Wang X, Deng X, Fu Q, Zhou Q, Feng J, Ma H, Liu W, Zheng C (2020) A weakly-supervised framework for covid-19 classification and lesion localization from chest CT. IEEE Trans Med Imaging 39(8):2615–2625

Wang G, Liu X, Li C, Xu Z, Ruan J, Zhu H, Meng T, Li K, Huang N, Zhang S (2020) A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images. IEEE Trans Med Imaging 39(8):2653–2663

Wang L, Lin ZQ, Wong A (2020) Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Sci Rep 10(1):1–12

Wang X, Jiang L, Li L, Xu M, Deng X, Dai L, Xu X, Li T, Guo Y, Wang Z (2021) Joint learning of 3d lesion segmentation and classification for explainable covid-19 diagnosis. IEEE Trans Med Imaging 40(9):2463–2476

Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, Sun C, Tang X, Jin L, Zhang M et al (2020) A machine learning-based model for survival prediction in patients with severe covid-19 infection. MedRxiv. https://doi.org/10.1101/2020.02.27.20028027

Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2019) Unet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 39(6):1856–1867

Zhou T, Canu S, Ruan S (2020) An automatic covid-19 ct segmentation network using spatial and channel attention mechanism . arXiv preprint arXiv:2004.06673

Zhou Z, Sodha V, Rahman Siddiquee MM, Feng R, Tajbakhsh N, Gotway MB, Liang J (2019) Models genesis: Generic autodidactic models for 3d medical image analysis. In:International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 384–393. Springer