



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A semi-supervised learning approach for COVID-19 detection from chest CT scans

Yong Zhang^{a,b}, Li Su^{a,b}, Zhenxing Liu^{a,b}, Wei Tan^{d,*}, Yinuo Jiang^c, Cheng Cheng^{c,*}

^aSchool of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

^bEngineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, Wuhan 430081, China

^cSchool of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

^dHospital of WUST, Wuhan University of Science and Technology, Wuhan 430081, China

ARTICLE INFO

Article history:

Received 23 March 2022

Revised 11 May 2022

Accepted 18 June 2022

Available online 23 June 2022

Keywords:

COVID-19

Computed tomography

Semi-supervised learning

Deep learning

Attention mechanisms

ABSTRACT

COVID-19 has spread rapidly all over the world and has infected more than 200 countries and regions. Early screening of suspected infected patients is essential for preventing and combating COVID-19. Computed Tomography (CT) is a fast and efficient tool which can quickly provide chest scan results. To reduce the burden on doctors of reading CTs, in this article, a high precision diagnosis algorithm of COVID-19 from chest CTs is designed for intelligent diagnosis. A semi-supervised learning approach is developed to solve the problem when only small amount of labelled data is available. While following the MixMatch rules to conduct sophisticated data augmentation, we introduce a model training technique to reduce the risk of model over-fitting. At the same time, a new data enhancement method is proposed to modify the regularization term in MixMatch. To further enhance the generalization of the model, a convolutional neural network based on an attention mechanism is then developed that enables to extract multi-scale features on CT scans. The proposed algorithm is evaluated on an independent CT dataset of the chest from COVID-19 and achieves the area under the receiver operating characteristic curve (AUC) value of 0.932, accuracy of 90.1%, sensitivity of 91.4%, specificity of 88.9%, and F1-score of 89.9%. The results show that the proposed algorithm can accurately diagnose whether a chest CT belongs to a positive or negative indication of COVID-19, and can help doctors to diagnose rapidly in the early stages of a COVID-19 outbreak.

© 2022 Published by Elsevier B.V.

1. Introduction

In December 2019, a case of unexplained pneumonia was diagnosed and it spread rapidly throughout the country and around the world. Severe cases appeared as acute respiratory distress, multiple organ failure and other symptoms [1,2]. It has been shown that the pneumonia was caused by a new Coronavirus infection and was identified as an international public health emergency by the WHO in January 2020. As of April 1, 2022, more than 618 million people have been diagnosed with Covid and 4.9 million people have died. Early screening of suspected infected patients plays a vital role in preventing and fighting new cases of Corona pneumonia. COVID-19 is usually diagnosed by a reverse transcription polymerase chain reaction (RT-PCR) in the early stages of the epidemic. However, due to the rapid outbreak of

the epidemic, many countries still lack sufficient kits to detect suspected patients. Moreover, it takes several days for RT-PCRs to get results, which leads to the delay of epidemic controls and treatment. In addition, RT-PCR detection sensitivity is low: one test may not be able to make accurate judgments, so it needs multiple tests to make the final diagnosis [3]. In clinical practice, researchers found that the CT images of COVID-19 patients showed ground-glass opacities (GGO), multifocal patchy consolidation, and interstitial changes with a peripheral distribution and other image features [4,5]. And compared with RT-PCR detection, doctors can get chest CT scans and corresponding diagnosis results faster. CT is an important component in modern medical care systems, and plays a key role in combating the disease, so CT has become another effective way to screen and diagnose COVID-19 [6].

With the increasing severity of the epidemic, the number of CT images which contain a large amount of disease information has increased dramatically. The large number and high complexity of image data also easily fatigues doctors engaged in high-intensity diagnosis work, which makes it difficult for doctors to keep

* Corresponding author.

E-mail addresses: zhangyong77@wust.edu.cn (Y. Zhang), zhenxingliu@wust.edu.cn (Z. Liu), tanwei63317@163.com (W. Tan), c_cheng@hust.edu.cn (C. Cheng).

focused. It even runs the risk of doctors making wrong diagnoses. Once misdiagnosis occurs, patients will miss the best opportunity for treatment, and could even spread the epidemic. Deep learning, which has emerged as a powerful tool for improving the efficiency of CT diagnosis can automatically classify medical images, effectively helping doctors to make correct judgments and recommend corresponding treatment for patients. It can also reduce the risk of misdiagnosis in the process of early diagnosis, and improve the cure rate. In recent years, with the breakthrough of deep learning in computer vision, it has been widely used in image classification [7], image location and detection [8], medical image segmentation [9] and other fields, greatly reducing the burden of massive medical image data on doctors. There are a large number of published studies that the role of deep learning in disease diagnosis. Arevalo et al. [10] propose a feature learning framework for breast cancer diagnosis, which uses CNN to automatically learn discriminative features and classify breast X-ray lesions. Gerard et al. [11] propose a supervised discriminant learning framework for simultaneous feature extraction and classification (See Fig. 1).

Although these studies have good performance in medical image classification, they also have some limitations. First of all, most of them are based on supervised learning and need a lot of labelled data. But in a lot of practical work, there may be only a few labelled samples available, because the cost of labelling data is very high. For example, CT acquisition and labelling of Covid-19 requires a lot of time and energy from professional doctors, which is more difficult during the epidemic. Training deep learning models needs a lot of labelled data to achieve a clinical standard of performance. Insufficient data will lead to over fitting and poor

performance of the model. Secondly, because medical image data involves the privacy of patients, many CT image datasets are not public. The models trained with these non-public datasets cannot be used in other hospitals.

Traditional machine learning technology is divided into supervised learning and unsupervised learning. In some scenarios, labelled data is difficult to obtain, while unmarked data is relatively easy to obtain. Semi-supervised learning aims to introduce unlabelled samples when the information of sample labels is limited and it is difficult for the classifier to determine accurate classification decision boundaries accurately. The hidden distribution information learned from the model is used to help the classifier to move towards the correct decision, thus achieving higher generalization and accuracy. In the field of natural image recognition, semi-supervised learning can use a small number of labelled samples and a large number of unlabelled samples to alleviate the problem of insufficient data [12–14]. A classic example of the application of semi-supervised learning in medical imaging is Liu et al. [15] who proposed a new relationship-driven semi-supervised medical image classification framework to classify chest X-ray diseases. Additionally, Su et al. [16] propose an interactive cell segmentation algorithm based on active annotation and verification propagation.

Based on these findings, we propose a semi-supervised learning method based on deep learning to automatically diagnose CT scans of COVID-19 in this paper. While following the MixMatch rules to conduct sophisticated data augmentation, we introduce a model training technique to reduce the risk of over-fitting of the model by marking data. At the same time, a new data enhancement

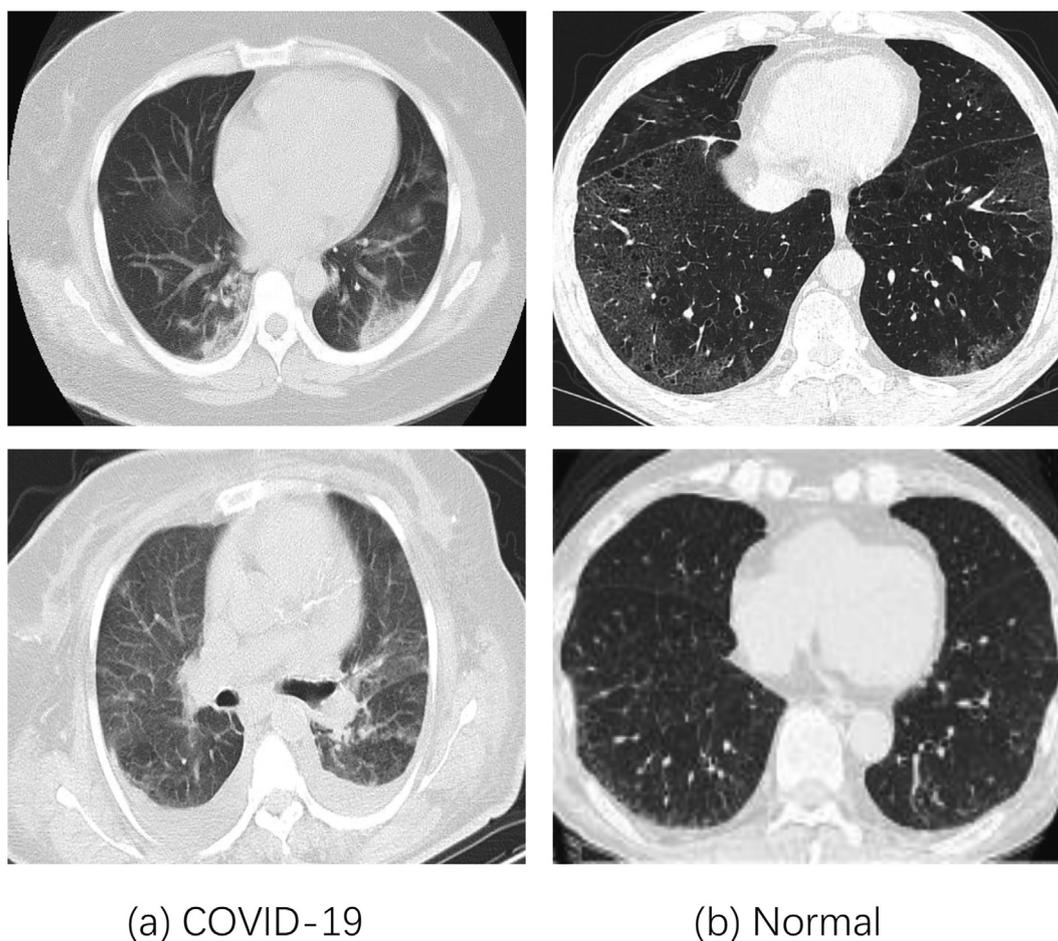


Fig. 1. Examples of (a) COVID-19 infections and (b) non-infected CT images as shown in the left and the right column respectively.

method is proposed to make the model focus on the areas that are difficult to distinguish. In order to further improve the performance of the model, a convolutional neural network based on attention mechanisms will be developed to achieve the accurate classification of CT scans. The decision-making process of the model is not transparent, so doctors pay special attention to the interpretability of the model, which is also very important for the diagnosis of COVID-19. We make a visual analysis of model to increase the interpretability of the model. Our method is evaluated on the CT public dataset of COVID-19 and achieves better performance in the case of a small number of samples compared with other methods.

As a summary, the contributions of our work are threefold:

- We improve MixMatch technique to release the training signal of labelled data, which effectively prevents the model overfitting with the labelled data.
- A new data enhancement method is proposed to replace the regularization method in MixMatch.
- We modify the attention module that is able to extract multi-scale features, which can be added to the existing network to ensure that the network focuses on the exact infected area and increase the performance of the model.

We organize the remainder of this paper as follows. Section 2 briefly reviews the basic principles of MixMatch and the related works in terms of a deep learning model. In Section 3, we describe in detail the proposed semi-supervised learning strategy and improved models. We then give detailed descriptions of collected datasets, experiment settings, and exhaustive results in Section 4. Finally, we conclude this work in Section 5.

2. Related works

2.1. MixMatch

MixMatch [17] is a semi-supervised learning method, which follows two principles: consistency regularization and entropy minimization. The principle of consistency regularization is that the decision boundary of learning must be located in the low-density region, that is, if an unlabelled data is disturbed, the output of the model should remain unchanged or approximate as far as possible. Mixmatch adds its rules to the loss function in the following form:

$$\|p_{\text{model}}(y|\text{Augment}(x); \theta) - p_{\text{model}}(y|x; \theta)\|_2^2 \quad (1)$$

where x is unlabeled data, $\text{Augment}(x)$ is the new data generated by randomly data enhancement. θ is the model parameter, and y is the model prediction result. That means new samples are generated by data expansion, and the prediction results of the model should be consistent.

The entropy minimization principle forces the classifier to make low entropy prediction for unlabelled data. MixMatch uses a sharpening function to minimize the entropy of the i -th unlabelled data:

$$\text{Sharpen}(p, T)_i = \frac{p_i^{\frac{1}{T}}}{\sum_{j=1}^L p_j^{\frac{1}{T}}} \quad (2)$$

where p is prediction label of the input and L is the number of categories. T is the temperature parameter, which can adjust the classification entropy.

Mixup [18] is a kind of image enhancement algorithm used in computer vision. It can mix different kinds of images to expand the training dataset. MixMatch mixes labelled data and unlabelled data by mixup and the mixed sample is:

$$\lambda \sim \text{Beta}(\alpha, \alpha) \quad (3)$$

$$\lambda' = \max(\lambda, 1 - \lambda) \quad (4)$$

$$x' = \lambda'x_1 + (1 - \lambda')x_2 \quad (5)$$

$$y' = \lambda'y_1 + (1 - \lambda')y_2 \quad (6)$$

where x_1 and x_2 are input data with corresponding labels y_1 and y_2 , while x' and y' are the output data and corresponding labels. α is a hyperparameter used for generating the beta distribution.

2.2. Training signal annealing

The basic principle of Training Signal Annealing (TSA) [19] is that in the training process, with the increase of unlabelled data, the label data is gradually removed. The training signal of supervised data is gradually released, so as to avoid an over fitting problem of the model to the label data.

At t time of training, set a threshold η_t ($\frac{1}{L} \leq \eta_t \leq 1$, L is the number of categories). When the probability of the correct category p_θ of a label example is higher than the threshold η_t , the model removes this example from the loss function and trains only other labelled examples under this minibatch:

$$\min_{\theta} \frac{1}{Z} \sum_{x,y \in B} [-I(p_\theta(y|x) < \eta_t) \log p_\theta(y|x)] \quad (7)$$

$$Z = \sum_{x,y \in B} I(p_\theta(y|x) < \eta_t) \quad (8)$$

where B is minibatch sample, Z is the filtered sample set and I is the indicator function.

Threshold η_t is used to prevent the model from being over fitting to label data. With η_t is close to 1, the model can only be monitored slowly from the annotation examples, which greatly alleviates the problem of over fitting. η_t values is varying while training:

$$\eta_t = \exp\left(\left(\frac{t}{S} - 1\right) \times N\right) \times \left(1 - \frac{1}{L}\right) + \frac{1}{L} \quad (9)$$

where S is the total training steps, t is the current training steps and N is a constant equals to 5.

2.3. Baseline models

We experimented with different backbone networks and chose two best performing networks as baseline models. The first baseline model is DenseNet121 [20] whose structure is that all the front layers are directly connected with those of the back layer. Therefore, the back layer of the network not only accepts the output of the previous layer as the input, but also accepts the output of all the previous layers as the additional input. This connection method can use the previously extracted features many times, so it saves a lot of calculation process. The second model is ResNet50 [21] which uses residual blocks to directly connect the shallow feature layer with the deeper feature layer. Through direct connection, the back-propagation process can be transmitted to the feature layer closer to the input layer more conveniently. Before the emergence of residual blocks, the effect of networks with more hidden layers will decline. The addition of residual blocks effectively solves this problem.

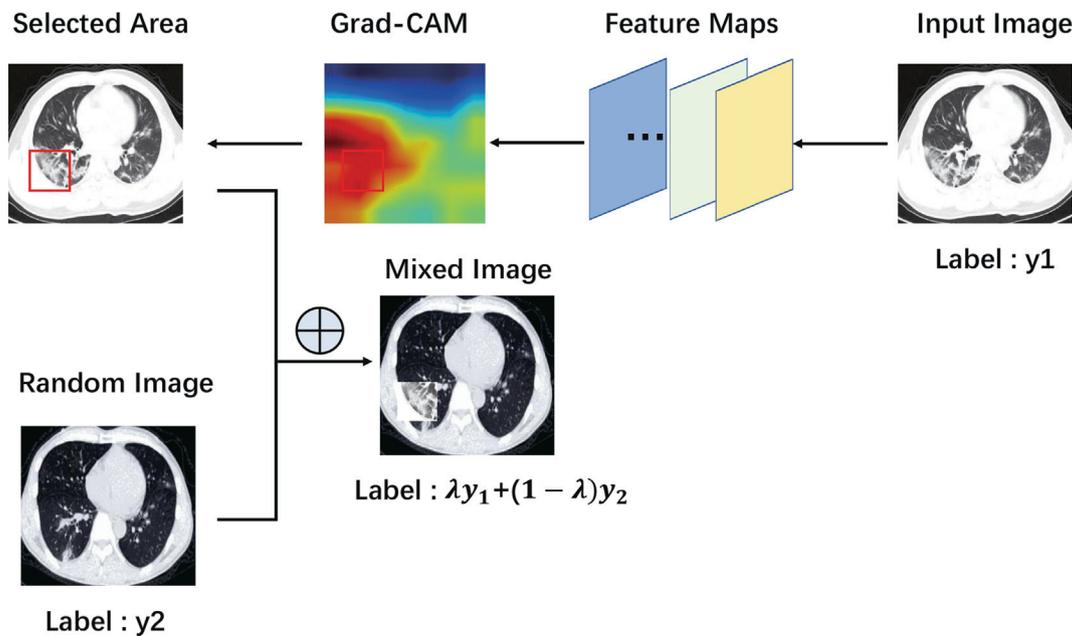


Fig. 2. Framework overview of proposed CAMMix.

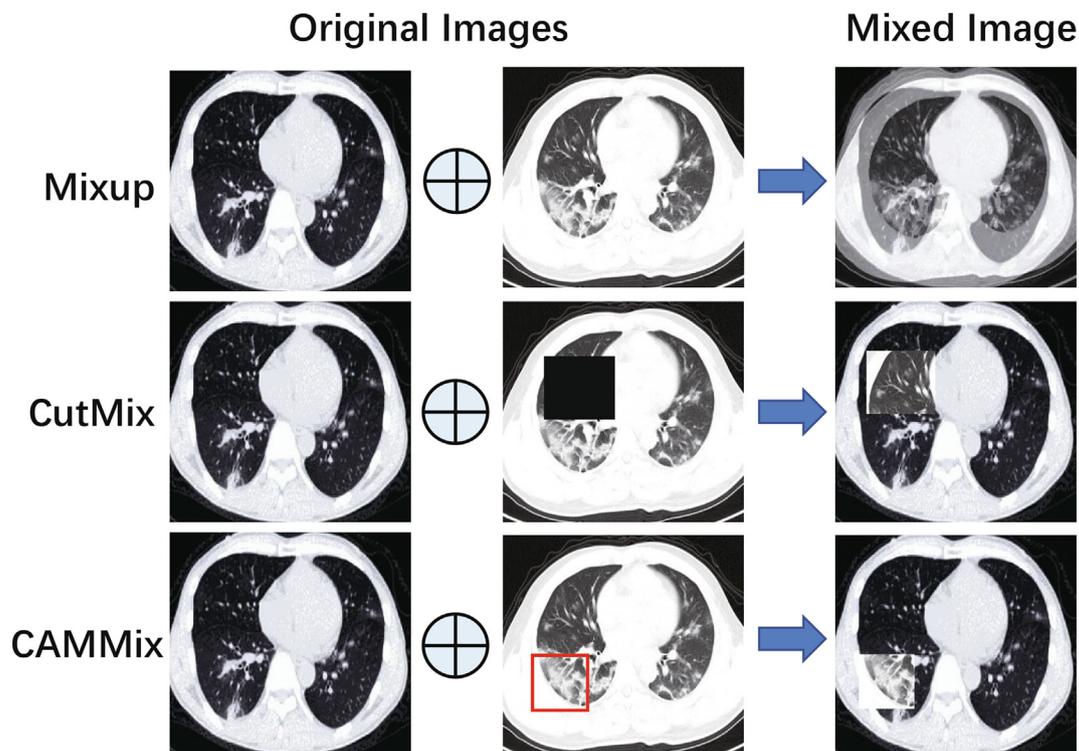


Fig. 3. Comparison results of CAMMix, Mixup, and CutMix.

2.4. Visual explanations from deep networks

CAM (Class Activation Mapping) [22] is a tool to help us visualize CNN. We can clearly observe which area of the image the network focuses on by using CAM, but also need to change the network structure and retrain. Guided backpropagation [23] visualizes the gradient of network back propagation to understand the network. This visualization method has high resolution and

can show the fine-grained details in the image, but the visualization effect is not good for category discrimination. Grad-CAM [24] calculates the weight of each feature graph through the global average of the gradient, and then makes a weighted summation according to the weight of the corresponding categories of all feature graphs to get the final thermal graph. Grad-CAM does not need to modify the model structure and retrain the model, thus can be applied to a variety of different tasks. In order to better

understand the decision of the model, we used Gard-CAM to visualize the attention map of the lesion area.

3. Method

The overall framework is shown in Fig. 4 and our method is divided into two stages: Stage 1 is that a small number of labelled samples and a large number of unlabelled samples are used to generate new samples by semi-supervised learning which are sent them to the network for training. Stage 2 involves training the two models with mixed images and labelled images respectively,

and uses ensemble learning to integrate predictions from the two trained networks to get a final diagnosis.

3.1. SSL strategy

MixMatch is an effective semi-supervised strategy, it has some defects: in the training process, the model is easy to over fit with a small amount of labelled data. In our study, we added TSA to MixMatch to gradually release labelled samples in the training process, so as to avoid model overfitting. Algorithm 1 shows an overview of the semi-supervised learning process. After K times of data

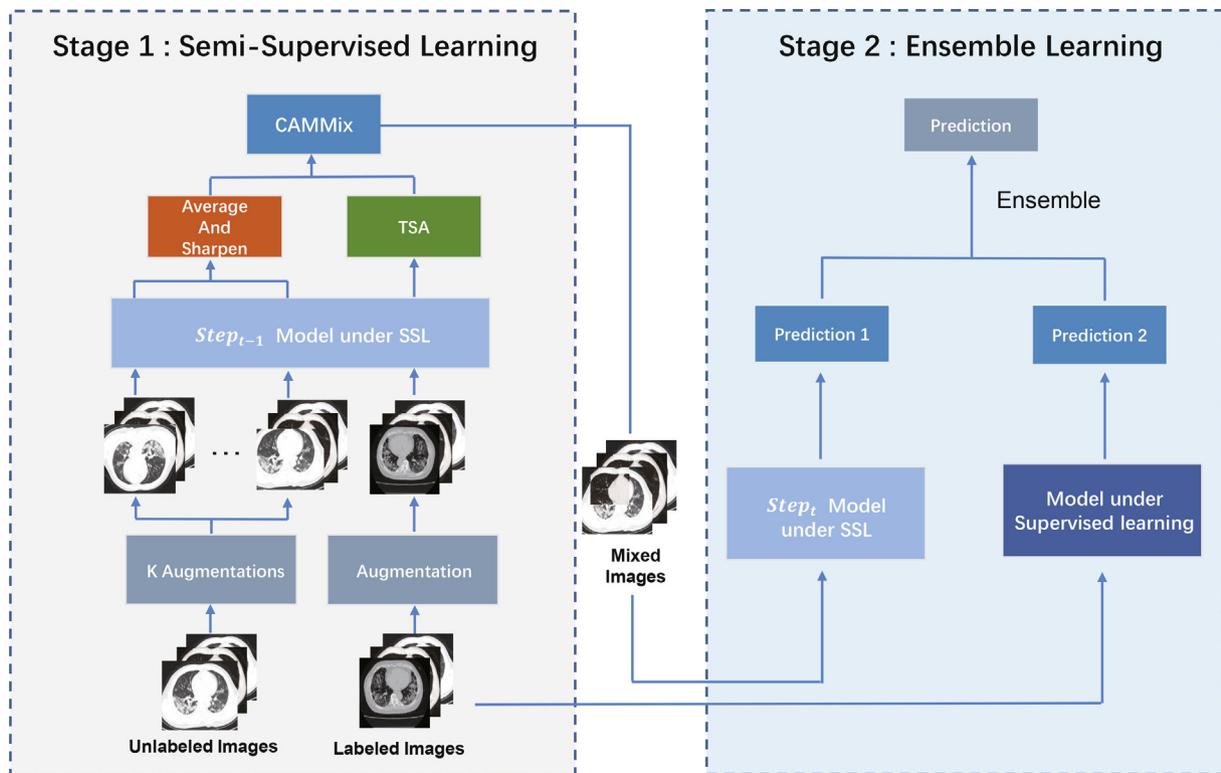


Fig. 4. The proposed method includes two stages: 1) a small number of labelled samples and a large number of unlabelled samples are used to generate new samples by semi-supervised learning which are sent to the network for training. 2) We train the two models with mixed images and labelled images respectively, and use the ensemble learning to integrate predictions from the two trained networks.

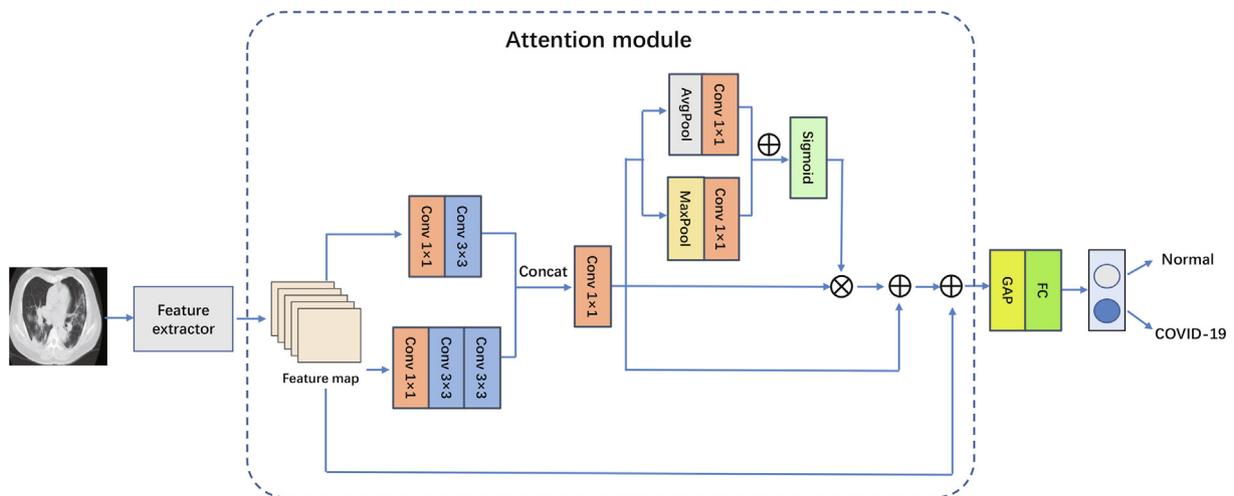


Fig. 5. Overview of Attention module.

Table 1
Details of the labelled dataset and unlabelled dataset.

Dataset	Labelled		Unlabelled	
	Training	Validation	Test	Training
COVID-19	191	60	93	500
Normal	234	58	99	500
Total	425	118	192	1000

enhancement, the unlabelled samples are predicted by using network models, and then the soft-labels are obtained by averaging and sharpening. The labelled samples are enhanced once through the same procedure for prediction. If the probability of the correct category of the predicted labelled sample is higher than the threshold, the model removes this example from the loss function. Then, new images are generated from the augmented labelled data and unlabelled data through CAMMix, which are sent to the network model for training.

3.2. CAMMix

Mixup is used as the regularization method in MixMatch, however, it produces noise that impacts the model to learn the accurate characteristic diagram response distribution. Cutout [25], Cutmix [26] and other methods promote better generalization of the network by partially occluding the distinguished parts of the object. However, these methods are hard to capture most important regions in the image. To solve this problem, we propose a new data enhancement method called CAMMix, which is based on a Grad-CAM to replace Mixup in MixMatch. In each training process, we selected the most descriptive area in the image according to Grad-CAM and cut it to another image to get a new mixed image. The main process is shown in Fig. 2:

Grad-CAM is a response based visual interpretation method. The weights of the FC layer and the feature map are weighted and summed to generate the attention map, highlighting the important areas closely related to the prediction results. Therefore, we first calculate the weight of the Gard-CAM of the input image:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial Y^c}{\partial A_{ij}^k} \quad (10)$$

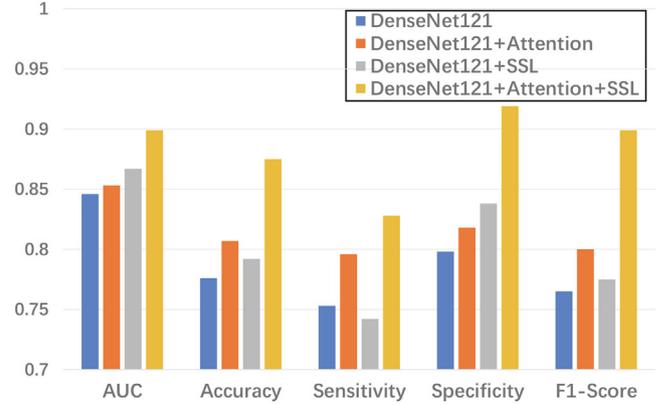
where c denotes categories, Y is the logits corresponding to the categories. A is the feature map. k denotes the channel of the feature map and i, j represents the abscissa and ordinate of the feature map. Z denotes the size of the feature map.

After obtaining weights, the channels of the feature map are linearly weighted and through the ReLU layer to obtain Gard-CAM:

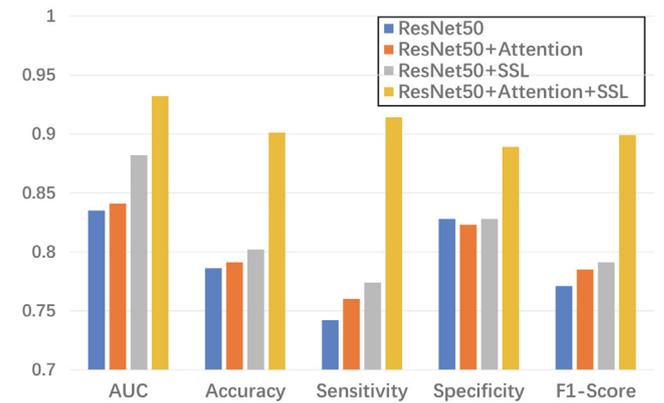
$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right) \quad (11)$$

Table 2
Performance of SSL and attention modules in the ablation studies.

Method	AUC	Accuracy	Sensitivity	Specificity	F1-score
DenseNet121	0.846	0.776	0.753	0.798	0.765
DenseNet121 + Attention	0.853	0.807	0.796	0.818	0.800
DenseNet121 + SSL	0.867	0.792	0.742	0.838	0.775
DenseNet121 + Attention + SSL	0.899	0.875	0.828	0.919	0.865
ResNet50	0.835	0.786	0.742	0.828	0.771
ResNet50 + Attention	0.841	0.791	0.760	0.823	0.785
ResNet50 + SSL	0.882	0.802	0.774	0.828	0.791
ResNet50 + Attention + SSL	0.932	0.901	0.914	0.889	0.899



(a) Densenet121



(b) ResNet50

Fig. 6. Performance of SSL and attention modules in the ablation studies.

Since the output size of the last layer of the convolution layer is usually not equal to the size of the input image, the result needs to be up sampled to the size of the input image. Then the highest contribution is selected whose size is 56×56 from Grad-CAM as the attentive region, and the region is cut from the given image and pasted it to the corresponding position of another image:

$$x' = B \odot x_1 + (1 - B) \odot x_2 \quad (12)$$

$$y' = \lambda y_1 + (1 - \lambda) y_2 \quad (13)$$

where $B \in \{0, 1\}$ is the binary mask, indicating which pixels belong to which of the two images. λ is the ratio of the area cut from the first image to the second image of the total size of the second image. Considering that we select an area of 56×56 from image of 224×224 , λ is set to $\frac{1}{16}$.

3.3. Attention module

Attention mechanisms [27,28] are widely used in natural language processing and computer vision. They are similar to a human visual mechanism, and tend to pay attention to the parts in the image that are more helpful for decision-making, meanwhile, ignoring the unimportant information. Attention mechanisms can help the model assign different weights to each part of the input, extract more critical and important information, and help the model make more accurate judgments. CBAM (Convolutional Block Attention Module) [29] proposes a convolutional block attention module to extract information features by fusing cross channel and spatial information (See Fig. 3).

First, the feature map is extracted, a parallel multi convolution operation is then carried out on the input features to extract multi-scale image features with different depths. Features of the two branches are then concatenated. A 1×1 convolution layer is added before 3×3 convolution layer to reduce the number of parameters. The fused feature map is passed through 1×1 convolution layer to ensure the same dimension of the input and output layers.

Finally, the feature map goes through global Max pooling and global average pooling based on width and height respectively to get two feature maps which go through the convolution layer. Then, the output features of the convolution layer are added. The final attention map is generated by the sigmoid layer. Because the input scale and output scale are unchanged, it can be easily embedded into the current mainstream network architecture.

3.4. Loss function

Two losses are used to train model: the labelled data loss L_{label} and the unlabelled data loss $L_{unlabel}$. We adopt the binary cross entropy as the labeled data loss and adopt MSE as the unlabelled data loss:

$$L_{label} = -\frac{1}{|X|} \sum_{x,p \in X'} (p \log(p_{model}(y|x)) + (1-p) \log(1-p_{model}(y|x))) \quad (14)$$

$$L_{unlabel} = \frac{1}{L|U|} \sum_{u,q \in U'} (q - p_{model}(y|u))^2 \quad (15)$$

where L is the number of classification categories, x and p are the augmented labelled data input and corresponding labels, u and q are unlabelled data inputs and corresponding labels.

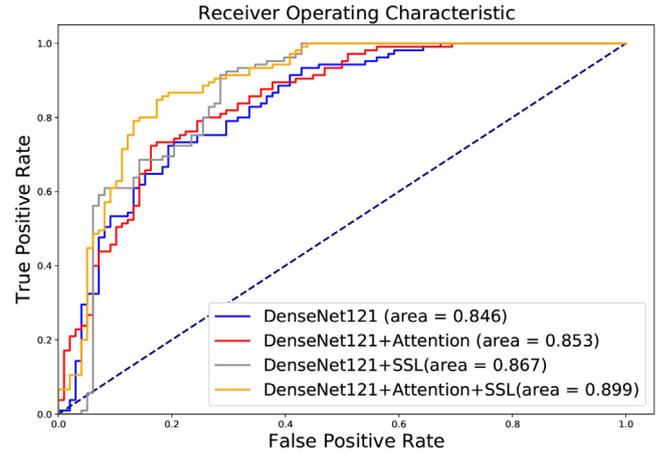
Then, the overall loss function for training model is expressed as:

$$L_{total} = L_{label} + \zeta L_{unlabel} \quad (16)$$

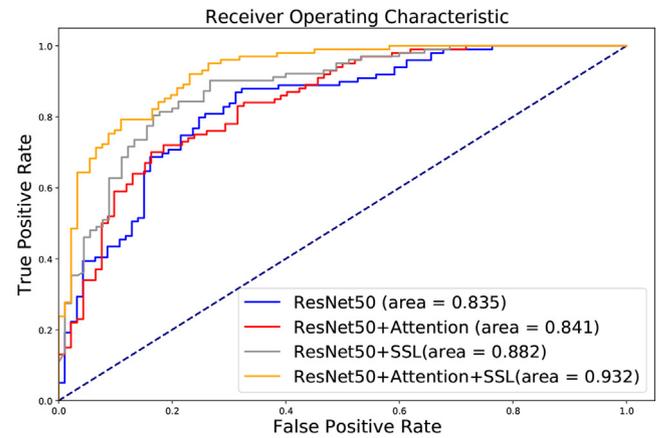
where ζ is the weighting factor of the unlabelled data loss function.

3.5. Ensemble learning

Semi-supervised learning can introduce unlabelled images for training to solve the problem of insufficient data, but it may also cause model over fitting on unlabelled data. In contrast, supervised learning could learn feature representation from the original data distribution in a relatively robust way. Taking the advantages of both sampling methods, we use ensemble learning [30] to gauge the weight for the prediction results produced by the two models. Ensemble learning completes the learning task by building and combining multiple models. Two models are trained with different learning strategies at the same time, and then the prediction results of the two models are integrated into the final diagnosis results by using an ensemble learning layer:



(a) DenseNet121



(b) ResNet50

Fig. 7. The receiver operating characteristic curve of binary classification between COVID-19 and Normal.

$$p_{total} = \mu p_1 + (1 - \mu) p_2 \quad (17)$$

where p_1 is the prediction score of models trained by semi-supervised learning, and p_2 is the prediction score of models trained by supervised learning (See Fig. 5).

Algorithm 1: SSL Algorithm

Input:

- 1: Initialization parameters number of augmentations K , sharpening temperature T , network parameters θ , loss function coefficient λ ;
- 2: Batch of labelled samples $x = ((x_b, y_b); b \in (1, \dots, B))$;
- 3: Batch of unlabelled samples $u = ((u_b); b \in (1, \dots, B))$;
- 4: **foreach** minibatch B **do**
- 5: **for** $b = 1$ to B **do**
- 6: $\hat{x}_b = \text{Augment}(x_b)$
- 7: $\hat{u}_{b,k} = \text{Augment}(u_b)$ for K times
- 8: $\bar{q}_b = \text{Average}(p_{model}(\hat{u}_{b,k}; \theta))$
- 9: $q_b = \text{Sharpen}(\bar{q}_b, T)$: by using Eq. (2)
- 10: **end for**
- 11: $w = \text{Shuffle}(\text{Concatenate}(\hat{x}_b, y_b), (\hat{u}_{b,k}, q_b))$

```

12:  $(x', p') = \text{CAMMix}(\hat{x}_{bi}, y_{bi}, w_i), i \in (1, \dots, |\hat{x}_b|)$ 
13:  $(u', q') = \text{CAMMix}(\hat{u}_{bi,k}, q_{bi}, w_{i+|\hat{x}_b|}), i \in (1, \dots, |\hat{u}_b|)$ 
14: if  $s_1 \leq \eta_t$  then
15:    $L_{\text{label}} = \text{CrossEntropy}(p'_i, p_{\text{model}}(x'_i; \theta)), i \in (1, \dots, |\hat{x}_b|)$ :
   calculate by using Eq. (14)
16: else
17:   pass
18: end if
19:  $L_{\text{unlabel}} = \text{MSE}(q'_j, p_{\text{model}}(u'_j; \theta)), j \in (1, \dots, |\hat{u}_b|)$ :
   calculate by using Eq. (15)
20:  $L_{\text{total}} = L_{\text{label}} + \zeta L_{\text{unlabel}}$ 
21: update  $\theta$  using optimizer ADAM
22: end for
23: return  $\theta$ 
24: Using ensemble learning to get the sample's final
   prediction: by using Eq. (17)

```

4. Experimental results

4.1. Description of experimental dataset

We used a labelled CT dataset and an unlabelled CT dataset for evaluating the proposed methods in the diagnosis of COVID-19 in this study. The labelled CT dataset is a public dataset collected by

He et al. [31], which contains 349 positive and 397 negative CT scans. The positive samples were COVID-19 preprints from medRxiv and bioRxiv, and negative samples include a CT scan of health people or other types of diseases. The unlabelled sample dataset is derived from several open source COVID-19 CT image datasets [32], which are used by researchers to accurately diagnose COVID-19 using CT images. We randomly selected 500 positive and 500 negative samples as unlabelled samples for training. Table 1 presents the dataset information.

4.2. Training details and evaluation methods

Our method is implemented by Pytorch. The input shape of the CT image is resized to 256×256 , and different data augmentations have been applied including horizontal flipping, random cropping and scaling. We use the Adam optimizer with the momentum set to 0.9, weight attenuation of 0.0001 and learning rate of 0.001 and minibatch size set to 32. μ is the ensemble learning weight factor that is set to 0.6 in our experiment. ζ is the weighting factor of the unlabelled data loss function and is set to 100. We use a cosine learning rate scheduler to adjust the learning rate during the training. All models first train the networks from scratch on ImageNet and then fine-tune them on the dataset. The training is conducted on GTX 2080Ti GPUs with data parallelism.

In order to evaluate the performance of the model, five different metrics are used to measure the classification results, namely area under the receiver operating characteristic curve (AUC), accuracy,

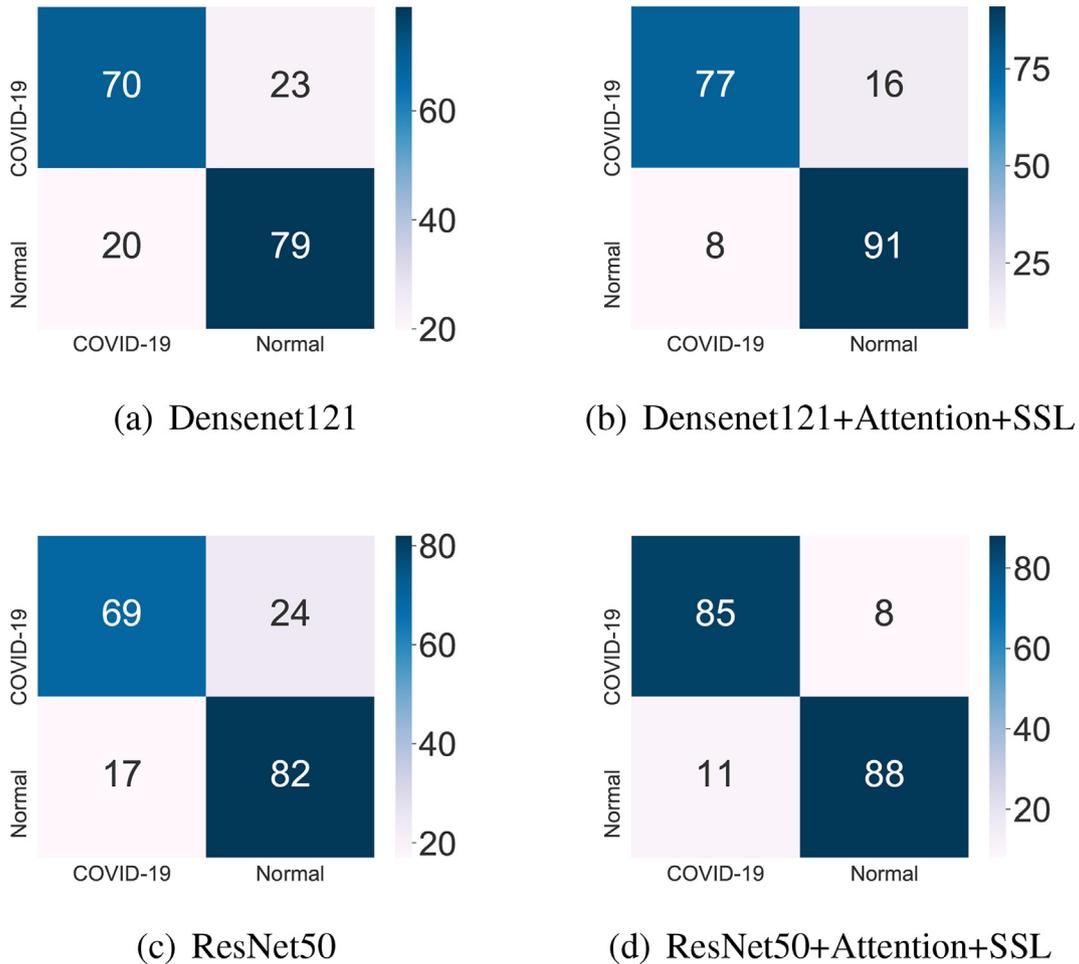


Fig. 8. The confusion matrix of the binary classification task.

sensitivity, specificity, and F1-score. For all five metrics, the highest score means the best performance of the model.

4.3. Simulation results analysis

4.3.1. Experimental result

ResNet50 and DenseNet121 are chosen as feature extractors of the proposed model. Table 2 and Fig. 6 compare different ablation study settings. Firstly, as can be seen from Table 2, the model with the attention module performs better than without the attention module. This shows that the proposed attention module suppress the contribution of irrelevant parts in the image, as a result, the model make diagnose decisions based on the actual infected area. In addition, Table 2 shows that the performance of the model is improved when the SSL is applied, which indicates the SSL technique can increase the generalization of the model by expanding the dataset and reduce the risk of model over fitting. It can be seen that when attention modules and SSLs are both included in the network, models achieve the best performance.

The receiver operating characteristic (ROC) curve of different models are shown in Fig. 7 to further evaluate the performance of different settings. The results show that both the SSL strategy and attention modules can improve the diagnostic ability of the model for patients with lung CT. When two modules are both included, which achieves the highest AUC scores of 0.899 and 0.932 for DenseNet121 and ResNet50 respectively, much higher than that of the basic network 0.846 and 0.835. These results further confirm the robustness and stability of the proposed algorithm.

The confusion matrix in Fig. 8 shows that the proposed algorithm significantly reduces the error of model judgments compared with the basic model. That means the proposed algorithm improves the accuracy of the model for the diagnosis of COVID-19.

4.3.2. Comparison with other algorithm

To demonstrate the efficacy of the proposed approach, we compare the design of it with other SSL methods and state-of-the-art methods, include: 1) mean teacher method proposed in [14], 2) Virtual Adversarial Training (VAT) [33] method that realizes the adversarial training of the model under semi supervision, 3) Interpolation Consistency Training (ICT) [34] method, and 4) Self-Trans [31] method that introduces comparative self supervised learning into the process of transfer learning, adjusting the network weights of source data pre training, so as to reduce the deviation of source data and reduce the risk of over fitting.

Table 3 and Fig. 9 show the evaluation metrics of the above five methods on the dataset. The proposed algorithm outperforms other semi-supervised algorithms for all the evaluation metrics, except 1% AUC score lower than the Self-Trans method. However, it significantly outperforming the Self-Trans on accuracy and F1-score metrics. Note that while all the methods achieve promising performance, our algorithm provides a more reliable result with respect to the ability of identifying the infection areas.

4.3.3. Visualization analysis

Fig. 10 shows the Grad-CAM visualizations of the baseline and our model. From left to right: column (1) shows original images with COVID-19; column (2–3) shows Grad-CAM visualizations of the baseline model; Specifically, column (2) is the Grad-CAM from the baseline. In column (3), the Grad-CAM is superimposed on the original image to show the active area. The color from dark red to dark blue corresponds to the value of the pixel's category significance from large to small. Columns (4–5) are Grad-CAM visualizations for our method. Grad-CAM visualizations is a visual interpretation of the network predicting COVID-19 CT scan lesions. By comparing columns (3) and (5), we find that the baseline

Table 3
Comparison of classification result of different algorithms on testset.

Method	AUC	Accuracy	F1-score
Mean teacher	0.869	0.802	0.808
ICT	0.884	0.860	0.863
VAT	0.873	0.813	0.824
Self-Trans	0.940	0.860	0.850
Our method	0.932	0.901	0.897

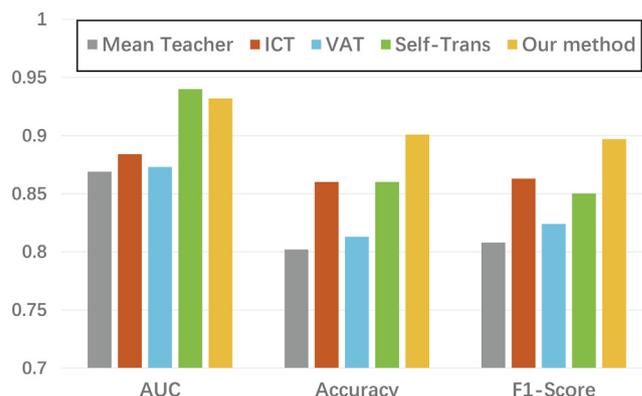


Fig. 9. Performance of the proposed method and other algorithms.

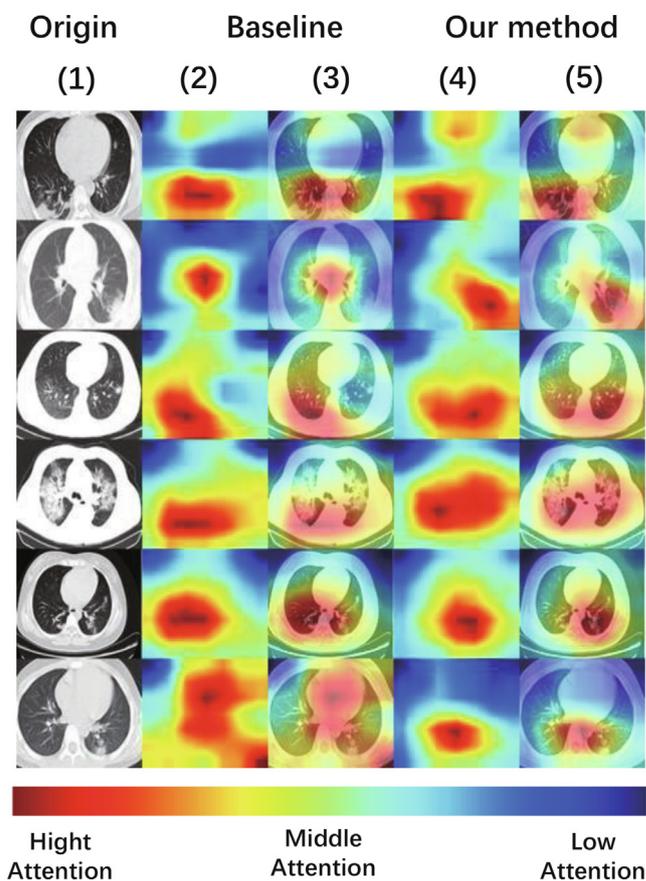


Fig. 10. Grad-CAM visualizations for baseline and the proposed method.

incorrectly focuses on some image edges and corners that are unrelated to the features of COVID-19 on CTs. In contrast, the proposed method has more accurate disease-related visual localization and can capture almost all significant regions affected by COVID-19.

5. Conclusion

This study set out to propose a new method of semi-supervised learning based on deep learning to automatically diagnose CT scans of COVID-19. While following the MixMatch rules to conduct sophisticated data augmentation, we introduce a model training technique to reduce the risk of over fitting the model by marking data. At the same time, a new data enhancement method is proposed to help the model focuses on the areas that are difficult to distinguish. In order to further improve the performance of the model, a convolutional neural network based on attention mechanisms is then designed to achieve accurate classification of CT scans. We experiment on an independent CT dataset on the chest of a patient with COVID-19 to evaluate the feasibility of our method which achieves an AUC of 0.932, accuracy of 90.1%, sensitivity of 91.4%, specificity of 88.9%, and an F1-score of 89.9%. Additionally, to better understand the decision of our model, we also visualized the Grad-CAM of the model, which is able to reveal important regions for diagnosis. The results of this research prove the proposed method can accurately diagnose whether chest CTs belong to the positive or negative diagnoses of COVID-19, and can help doctors to diagnose rapidly in the early stages of a COVID-19 outbreak.

More experiments are required to verify the feasibility of the proposed method in the future. Further research should improve the attention model to better focus on the lesion areas and reduce the influence of irrelevant regions. Some efficient filtering techniques [35–38] will also improve the accuracy of diagnosis and they are worth exploring.

CRedit authorship contribution statement

Yong Zhang: Conceptualization, Writing – original draft, Funding acquisition. **Li Su:** Data curation, Software, Investigation. **Zhenxing Liu:** Validation, Visualization. **Wei Tan:** Supervision, Investigation. **Yinuo Jiang:** Validation, Investigation. **Cheng Cheng:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China [Grant Nos. 61873197, 51905197].

References

- [1] S.P. Adhikari, S. Meng, Y.J. Wu, Y.P. Mao, R.X. Ye, Q.Z. Wang, C. Sun, S. Sylvania, S. Rozelle, H. Raat, Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review, *Infect. Diseases Poverty* 9 (2020) 1–12.
- [2] B. Fwca, S.Y. A. A. Khk, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, *Lancet* 395 (2020) 514–523.
- [3] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases, *Radiology* 296 (2020) 32–40.
- [4] M. Chung, A. Bernheim, X. Me, N. Zhang, H. Shan, CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV), *Radiology* 295 (2020) 202–207.
- [5] C. Huang, Y. Wang, X. Li, L. Ren, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (2020) 497–506.
- [6] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR, *Radiology* 296 (2020) 115–117.
- [7] J. Tan, P.J. Pickhardt, Y. Gao, Z. Liang, A.F. Abbasi, 3D-GLCM CNN: A 3-dimensional gray-level co-occurrence matrix based CNN model for polyp classification via CT colonography, *IEEE Trans. Med. Imaging* 39 (2020) 2013–2024.
- [8] Hoo-Chang Shin, Matthew R. Orton, David J. Collins, Simon J. Doran, Martin O. Leach, Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2013) 1930–1943.
- [9] G. Wang, W. Li, M.A. Zuluaga, R. Pratt, P.A. Patel, M. Aertsen, T. Doel, A.L. David, J. Deprest, S. Ourselin, Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning, *IEEE Trans. Med. Imaging* 37 (2018) 1562–1573.
- [10] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A.G. Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, *Comput. Methods Programs Biomed.* 127 (2016) 248–257.
- [11] S.E. Gerard, T.J. Patton, J.E. Bayouth, J.M. Reinhardt, G.E. Christensen, FissureNet: A Deep Learning Approach For Pulmonary Fissure Detection in CT Images, *IEEE Trans. Med. Imaging* 38 (2018) 156–166.
- [12] D.H. Lee, Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks, *ICML*, vol. 3, 2013, pp. 896.
- [13] A. Rasmus, H. Valpola, M. Honkala, M. Berglund, T. Raiko, Semi-Supervised Learning with Ladder Networks, *Comput. Sci.* 9 (2015) 1–9.
- [14] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, *Advances in neural information processing systems*, 30, 2017.
- [15] Q. Liu, L. Yu, L. Luo, Q. Dou, P.A. Heng, Semi-supervised Medical Image Classification with Relation-driven Self-ensembling Model, *IEEE Trans. Med. Imaging* 39 (2020) 3429–3440.
- [16] H. Su, Z. Yin, S. Huh, T. Kanade, J. Zhu, Interactive Cell Segmentation Based on Active and Semi-Supervised Learning, *IEEE Trans. Med. Imaging* 35 (2016) 762–777.
- [17] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, C. Raffel, MixMatch: A Holistic Approach to Semi-Supervised Learning, *Advances in Neural Information Processing Systems*, 32, 2019.
- [18] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: Beyond Empirical Risk Minimization, *arXiv preprint arXiv:1710.09412*, 2017.
- [19] Q. Xie, Z. Dai, E. Hovy, M.T. Luong, Q.V. Le, Unsupervised Data Augmentation for Consistency Training, *Advances in Neural Information Processing Systems* 33 (2020) 6256–6268.
- [20] G. Huang, Z. Liu, L.V.D. Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, *CVPR* (2017) 2261–2269.
- [21] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *CVPR* (2016) 770–778.
- [22] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning Deep Features for Discriminative Localization, *CVPR*, pp. 2921–2929, 2016.
- [23] J.T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, Striving for Simplicity, *The All Convolutional Net* (2014).
- [24] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, *ICCV* 1 (2020) 618–626.
- [25] T. Devries, G.W. Taylor, Improved regularization of convolutional neural networks with cutout, *arXiv preprint arXiv:1708.04552*, 2017.
- [26] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo and J. Choe, CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features, *ICCV*, pp. 6022–6031, 2019.
- [27] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, Z. Zhang, The application of two-level attention models in deep convolutional neural network for fine-grained image classification, *CVPR*, pp. 842–850, 2015.
- [28] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual Attention Network for Image Classification, *CVPR*, pp. 3156–3164, 2017.
- [29] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, CBAM: Convolutional Block Attention Module, *ECCV*, pp. 3–19, 2018.
- [30] A. Krogh, P. Sollich, Statistical mechanics of ensemble learning, *Phys. Rev. E* 55 (1997) 811.
- [31] X. He, X. Yang, S. Zhang, J. Zhao, P. Xie, Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans, *medRxiv*, 2020.
- [32] H. Gunraj, L. Wang, A. Wong, COVIDNet-CT: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest CT Images, *Frontiers in medicine*, 1025, 2020.
- [33] M. Takeru, M. Shin-Ichi, I. Shin, K. Masanori, Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (8) (2018) 1979–1993.
- [34] Vikas Sharma, Kenji Kawaguchi, Alex Lamb, Juho Kannala, Arno Solin, Yoshua Bengio, David Lopez-Paz, Interpolation consistency training for semi-supervised learning, *Neural Networks*, vol. 145, pp. 90–106, 2022.
- [35] L. Liu, L. Ma, J. Zhang, Y. Bo, Distributed non-fragile set-membership filtering for nonlinear systems under fading channels and bias injection attacks, *Int. J. Syst. Sci.* 52 (6) (2021) 1192–1205.
- [36] L. Zou, Z. Wang, H. Geng, X. Liu, Set-membership filtering subject to impulsive measurement outliers: a recursive algorithm, *IEEE/CAA J. Autom. Sin.* 8 (2) (2021) 377–388.
- [37] J. Mao, Y. Sun, X. Yi, H. Liu, D. Ding, Recursive filtering of networked nonlinear systems: A survey, *Int. J. Syst. Sci.* 52 (6) (2021) 1110–1128.
- [38] H. Geng, H. Liu, L. Ma, X. Yi, Multi-sensor filtering fusion meets censored measurements under a constrained network environment: advances, challenges and prospects, *Int. J. Syst. Sci.* 52 (16) (2021) 3410–3436.



Yong Zhang received the Ph.D. degree in control theory and control engineering from Huazhong University of Science and Technology, Wuhan, China, in 2010. From 2014 to 2015, he was a Visiting Scholar with the Department of Information Systems and Computing, Brunel University London, Uxbridge, U.K. He is currently an Professor with the School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, China. He has authored over 20 papers in refereed international journals. His current research interests include remaining useful life prediction of key equipment, fault diagnosis and fault tolerant control of networked systems. Dr. Zhang is a very active Reviewer for many international journals.

1st prize of Hubei Science and Technology Progress Award, 1st prize of Hubei Province, and participated in four monographs.



Yinuo Jiang received her B. Eng. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2020, where she is currently pursuing the Master's Degree in School of Artificial Intelligence and Automation. Her research interests include deep learning applications in ECG diagnose.



Li Su received the B.E. in the School of Electrical and Information Engineering, Hubei University Of Automotive Technology, Hubei, China. Since 2019, he has been studying for master's degree of control engineering in the School of Information Science and Engineering, Wuhan University of Science and Technology. His current research interests include medical image classification and recognition.



Cheng Cheng received the B. Eng. degree in Measurement, Control Technology and Instrument in 2012 from Tianjin University, China. In 2013 and 2018, she respectively received the MSc and the Ph.D. in Control Systems from Imperial College London, UK. She is currently a lecturer in the School of Artificial Intelligence and Automation at Huazhong University of Science and Technology, Wuhan, China. Her research interests include robust control, mechatronic systems modelling and simulation, and deep learning applications.



Zhenxing Liu received the M.Sc. and Ph.D. degrees in engineering in 1990 and 2004, respectively, from Huazhong University of Science and Technology, Hubei, China. Currently, he is a Professor with the School of Information Science and Engineering from Wuhan University of Science and Technology, Wuhan, China. His research interests include monitoring and diagnosis of electrical machines and automatic devices.



Tan Wei Chief physician/professor, received the Ph.D. degree in Medical Imaging Department from Union Hospital Tongji Medical College Huazhong University of Science and Technology, Wuhan, Hubei, China. Since 2007, he has been working in the Radiology Department of Tianyou Hospital affiliated to Wuhan University of Science and Technology, a key clinical specialty in Hubei Province, as the academic leader of medical imaging. Sametime, in 2011, he also act as the academic leader of medical imaging in Wuhan University of Science and Technology Hospital. Doctoral supervisor of Medical Imaging Department, master tutor of General Medicine.

In 2022, he was elected as Deputy director of the Medical Division, Wuhan University of Science and Technology, Wuhan, China. In the past five years, he won