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.

Neurocomputing 458 (2021) 232–245

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Deep supervised learning using self-adaptive auxiliary loss for COVID-19 diagnosis from imbalanced CT images

Kai Hu [a,d,1], Yingjie Huang [a,1], Wei Huang [b,1], Hui Tan [a], Zhineng Chen [c], Zheng Zhong [b], Xuanya Li [e], Yuan Zhang [a,*], Xieping Gao [a,f,*]

[a] Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University, Xiangtan 411105, China
[b] Department of Radiology, the First Hospital of Changsha, Changsha 410005, China
[c] School of Computer Science, Fudan University, Shanghai 200438, China
[d] Key Laboratory of Medical Imaging and Artifical Intelligence of Hunan Province, Xiangnan University, Chenzhou 423000, China
[e] Baidu Inc, Beijing 100085, China
[f] Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha 410081, China

ARTICLE INFO

Article history:
Received 16 October 2020
Revised 2 May 2021
Accepted 4 June 2021
Available online 7 June 2021
Communicated by Zidong Wang

Keywords:
COVID-19
Classification
Data imbalance
Deep supervised learning
Self-adaptive auxiliary loss

ABSTRACT

The outbreak and rapid spread of coronavirus disease 2019 (COVID-19) has had a huge impact on the lives and safety of people around the world. Chest CT is considered an effective tool for the diagnosis and follow-up of COVID-19. For faster examination, automatic COVID-19 diagnostic techniques using deep learning on CT images have received increasing attention. However, the number and category of existing datasets for COVID-19 diagnosis that can be used for training are limited, and the number of initial COVID-19 samples is much smaller than the normal's, which leads to the problem of class imbalance. It makes the classification algorithms difficult to learn the discriminative boundaries since the data of some classes are rich while others are scarce. Therefore, training robust deep neural networks with imbalanced data is a fundamental challenging but important task in the diagnosis of COVID-19. In this paper, we create a challenging clinical dataset (named COVID19-Diag) with category diversity and propose a novel imbalanced data classification method using deep supervised learning with a self-adaptive auxiliary loss (DSN-SAAL) for COVID-19 diagnosis. The loss function considers both the effects of data overlap between CT slices and possible noisy labels in clinical datasets on a multi-scale, deep supervised network framework by integrating the effective number of samples and a weighting regularization item. The learning process jointly and automatically optimizes all parameters over the deep supervised network, making our model generally applicable to a wide range of datasets. Extensive experiments are conducted on COVID19-Diag and three public COVID-19 diagnosis datasets. The results show that our DSN-SAAL outperforms the state-of-the-art methods and is effective for the diagnosis of COVID-19 in varying degrees of data imbalance.

1. Introduction

The coronavirus disease 2019 (COVID-19) is spreading all over the world and is a serious threat to human life and health. By 2 May 2021, the cumulative number of confirmed cases around the world was close to 153 million, with nearly 3.2 million deaths. Reverse Transcription-Polymerase Chain Reaction (RT-PCR) test is considered as the gold standard of confirming COVID-19 patients, which needs 4–6 h to obtain the results and tends to be inadequate in many areas where the disease is severe [1]. In clinical diagnosis, as easily available imaging equipment, chest CT provides huge assistance to clinicians when characteristic manifestations such as ground glass opacity (GGO) or bilateral patchy shadows in CT scans were observed [2]. However, the rapidly increasing demand for medical imaging reading has brought a heavy burden to clinicians. Meanwhile, due to the complexity of medical imaging, the long and tedious reading of medical imaging may cause misinterpretation and misjudgment to clinicians.

In recent years, the explosion of all kinds of data has made convolutional neural networks (CNNs) achieve great success in many fields such as computer vision [3,4]. Similarly, CNNs have been

* Corresponding authors at: Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University, Xiangtan 411105, China (X. Gao).
E-mail addresses: yuanz@xtu.edu.cn (Y. Zhang), xpgao@xtu.edu.cn (X. Gao).
[1] These authors contributed equally to the paper.

https://doi.org/10.1016/j.neucom.2021.06.012
0925-2312/

shown to be effective in assisting in the diagnosis of COVID-19. However, due to the fact that CNNs is a data-driven approach, this means high demand for data. On the one hand, the number and size of datasets available for COVID-19 diagnosis are limited. Moreover, most datasets are used to distinguish COVID-19 from non-COVID-19, that is, the diversity of classes is limited, so the trained models do not have good migration ability. On the other hand, in clinical practice, like most medical image datasets, the original COVID-19 samples are much smaller than the normal ones, so there are different degrees of imbalance between target and non-target samples. The problem of data imbalance means that some classes in the training set have much more samples than others [5], which makes it challenging to build a well-performed classifier. The classification algorithms are often forced to be biased towards the majority classes and neglect the minority classes, resulting in low classification accuracy. Therefore, it is essential to study the diagnosis of COVID-19 based on imbalanced data.

In general, two categories of approaches have been proposed to tackle the data imbalance problem, i.e., data-level methods and algorithm-level methods [5]. To make the distribution balanced in the data-level, the prior distribution is either modified by under-sampling the majority classes, or over-sampling the minority classes, or a combination of both [6]. It is well known that under-sampling might discard useful information, but medical image datasets are usually smaller and under-sampling side effects are more obvious, so over-sampling is preferred in most methods. For example, Xu et al. expanded the number of minority classes by 3 times to balance the majority class so as to weaken the influence of the imbalance of different image types [7]. Gozes et al. used image rotations, horizontal flips, and cropping to overcome the limited numbers of COVID-19 cases [8]. However, over-sampling often makes the training procedure computationally burdensome by increasing the size of training data. Besides, simple forms of over-sampling such as random replication only increase the number of images without increasing the diversity of features, and the models are susceptible to overfitting when using over-sampling. Thus, this paper focuses on the algorithm-level method.

While most existing algorithm-level approaches attempt to affect the loss functions by considering more prior information for the minority class, including the class-to-class separability [9] and the sample distribution of the raw datasets. However, the learning of the prior distribution of data for different classes is not automatic and the model parameters need to be adjusted manually. Furthermore, data overlap due to similarity between CT slices and possible mislabeling by clinicians may further degrade model performance.

To address the above problems, in this paper we construct a challenging COVID-19 diagnosis clinical dataset (named COVID19-Diag), and propose a novel Deep Supervised Network with a Self-Adaptive Auxiliary Loss (DSN-SAAL) for COVID-19 diagnosis with imbalanced CT images. Our COVID19-Diag is collected from different models of a hospital, including three categories of normal, bacterial pneumonia, and COVID-19. Although bacterial pneumonia and COVID-19 have their own characteristics in image performance, they also have a large part of overlapping characteristics. Therefore, our dataset has a certain diversity, which is more challenging than the binary dataset. Moreover, the data scale is relatively large, which is reflected in the number of cases and the number of CT slices. Our DSN-SAAL is used as a deep learning method to solve the problem of COVID-19 diagnosis with imbalanced CT images. With several new proposed techniques, DSN-SAAL can effectively learn the features for the samples of both majority classes and minority classes. Specifically, we first present a novel deep supervised learning framework for multi-scale feature learning for image classification. Then, we introduce an effective self-adaptive auxiliary loss for automatically learning the data distributions of both majority and minority classes. The loss function considers the data overlap by measuring the effective number of samples for reweighting the cross entropy (CE) loss. Meanwhile, a reverse cross entropy (RCE) as a regularization term is further proposed to handle incorrect labels. DSN-SAAL is self-adaptive since all model parameters are learned through the network automatically, and thus can be applied to various datasets. To verify the effectiveness of our proposed method, we conduct extensive experiments on our COVID19-Diag and three public COVID-19 diagnosis datasets. The experimental results demonstrate that our method outperforms the state-of-the-art approaches on both balanced and imbalanced datasets and is effective for the diagnosis of COVID-19.

Overall, the key contributions of this paper can be summarized as follows:

- We propose a novel deep supervised learning model with self-adaptive auxiliary loss, called DSN-SAAL, for the diagnosis of COVID-19 based on CT scans with the problem of data imbalance.
- We design a self-adaptive auxiliary loss, which considers both data overlap between CT slices and possible noisy labels during data collection for imbalanced data classification.
- We create a new COVID-19 diagnosis dataset (named COVID19-Diag) consisting of 6982 CT slices from 225 clinical cases in three categories: COVID-19, normal, and bacterial pneumonia. Extensive experiments are conducted on this dataset to verify the effectiveness of our DSN-SAAL in varying degrees of imbalance.
- We evaluate DSN-SAAL on three additional publicly available COVID-19 datasets. The results show that DSN-SAAL outperforms the state-of-the-art methods and can achieve significant performance on generalization ability.

2. Related work

2.1. COVID-19 diagnosis studies on CT scans

There has been plenty of studies on the diagnosis of COVID-19 on medical images (such as CT scans) using traditional machine learning methods or deep learning methods since the outbreak of COVID-19. Shi et al. first used a VB-Net to segment COVID-19 infection regions and then trained a random forest model with some hand-crafted features for the diagnosis [10]. He et al. combined self-supervised deep learning with transfer learning and proposed a Self-Trans approach, which can achieve high diagnosis accuracy of COVID-19 with limited training data [11]. Ying et al. proposed a deep learning-based CT diagnosis system called DeepPneumonia to identify patients with COVID-19 [12]. Li et al. proposed the Transfer-CheXNet which used the pre-trained network CheXNet for the COVID-19 classification task to better help the parameter learning of small and medium-sized datasets in the target task [13]. Gunraj et al. introduced a deep convolutional neural network architecture named COVIDNet-CT, which explored a machine-driven method for the diagnosis of COVID-19 from CT images [14]. Besides, there are some methods to diagnose COVID-19 by deep learning based on the extraction of regions of interest (ROIs). For example, Chen et al. trained a UNet++ to segment COVID-19 related lesions to help the diagnosis of COVID-19 [15]. Jin et al. proposed a CNN to segment the lung and then identify slices of COVID-19 cases [16]. In [7], a deep learning model based on V-Net is firstly designed to segment the infection regions, and a ResNet-18 network is then used to diagnose COVID-19.

Three dimensional (3D) models have also been applied to address the diagnosis of COVID-19. Considering the high cost of manual labeling of COVID-19, Wang et al. proposed a 3D Deep CNN (DeCoVNET) for the detection of weakly labeled COVID-19 [17]. However, the number of positive and negative samples used in these studies mentioned above are approximately equal, that is, the data is relatively balanced. Moreover, Wang et al. collected 1065 CT images, including 325 images of COVID-19 cases and 740 images of typical viral pneumonia, to train their deep learning model [18]. The raw data was somewhat imbalanced, but they used 160 images from both classes for training, which means the training data was absolutely balanced. The data compositions of part of the researches mentioned above are shown in Table 1.

Although these studies achieved good results in their own datasets, they did not take into account the impact of imbalanced data in the clinical diagnosis of COVID-19. In most cases, the number of open COVID-19 samples from CT scans are much less than that of normal samples. However, the COVID-19 characteristics obtained from the models trained with these imbalanced samples are often inadequate, which easily leads to a poor diagnosis of COVID-19. Moreover, the robustness of the model may not be good enough to be applied to other scenarios of COVID-19 diagnosis. Therefore, it is necessary to study the diagnosis of COVID-19 based on imbalanced data, which is more suitable for clinical practice.

2.2. Data-imbalanced loss

Since data-level methods consume more additional resources and are not stable in execution, this paper focuses on algorithm-level methods. Among them, the design of the loss function is the main embodiment of algorithm-level methods.

Generally, there are two main categories of loss functions for solving the data imbalance problem. The first one is a single form such as hinge loss, soft-max loss, Euclidean loss, and contrastive loss [20]. The other attempts to improve those methods, including class-balanced loss [21], focal loss [22], and cost-sensitive CE loss [9]. These methods adopt the weighted term based on CE [23], which makes the decision boundary of the classifier be biased to minority classes.

Due to the limited performance of these loss functions in realizing the identifiability of feature space, recent studies have begun to explore better combinations of multiple loss functions to solve the data imbalance problem. Inspired by the symmetric KL-divergence, Wang et al. [24] proposed a symmetric cross entropy to address the under learning problem of minority classes and the overfitting problem of noisy labels. Zhang et al. [25] and Deng et al. [26] proposed the combination of CE and center loss or range loss to concurrently enforce intra-class compactness and inter-class separability.

Nevertheless, the existing combined loss functions are mostly limited to the adjustment of hyper-parameters and do not have good generalization because they cannot judge the degree of action between multiple items. Recently, Shu et al. proposed a new meta-learning method, called Meta-Weight-Net [27], which adaptively extracts sample weights to ensure robust deep learning in the case of training data biases. In particular, considering that different depths of the network can learn diverse expressions of characteristics, the supervised role of loss functions should be various in different stages of the network, which is reflected in the extent of partiality for minority classes in solving the problem of data imbalance. Our approach considers these issues simultaneously and constructs an adaptive auxiliary loss in the deep supervised network to effectively combine the learning degree of different types of characteristics in each stage of the network, so as to promote the feature learning of minority classes in the network.

Table 1
The number of samples of each class in the datasets with CT scans.

Literature	Samples
Wang et al. [18]	325 COVID-19 740 Viral pneumonia
He et al. [11]	349 COVID-19 397 non-COVID-19
Soares et al. [19]	1252 COVID-19 1229 non-COVID-19
Xu et al. [7]	219 COVID-19 224 Influenza-A 175 Normal
Ying et al. [12]	777 COVID-19 505 Bacterial pneumonia 708 Normal
Gunraj et al. [14]	21395 COVID-19 36856 Common pneumonia 45758 Normal

3. Method

In this section, we present our method in detail, including the image preprocessing, the DSN-SAAL architecture, self-adaptive auxiliary loss, and the parameter optimization algorithm. The overall flowchart of the proposed method is shown in Fig. 1.

3.1. Image preprocessing

It is well known that the images obtained by different types of CT machines are different. For example, the CT scans in some cases do not have cylindrical scan boundaries. Besides, in the whole CT image, other parts except the lung parenchyma are not effective for the diagnosis of COVID-19. Therefore, we adopt some preprocessing operations before the training of our model. First of all, we adjust the window level (WL) and window width (WW) of all clinical CT images to be consistent (WL: −400, WW: 1500). Then, we normalize them linearly to [0, 255] to fit into the digital image format. After that, some morphological operations are used to extract the area of lung parenchyma as following. (1) We use binarization to obtain the binarized images with pixel value equals 170. (2) The unrelated regions are removed by erosion, dilation, floodfill, and other operations to obtain the largest connected regions. (3) We adopt the convex hull operation to create the mask of the lung parenchyma and then multiply them by the images before binarization to obtain the final regions of lung parenchyma, which are used as the input images for model training [28].

3.2. DSN-SAAL

In general, the shallow layers of CNNs contain more local information, while the deep layers contribute more abstract information. As the number of network layers increases, the influence of the shallow layers on the deep layer decreases gradually. Thus, in the process of image analysis using deep learning methods, the effective combination of shallow and deep information is conducive to the full learning of the features including minority and majority classes. To comprehensively extract the multi-scale feature information of an image, we propose a deep supervised network by considering several stages in the network architecture. Fig. 2 illustrates the proposed network, which is extended from the VGG-16 network [29] with five stages. Each stage contains several convolutional layers with a 3×3 filter. Different stages are connected by a 2×2 max pooling layer. For the richer

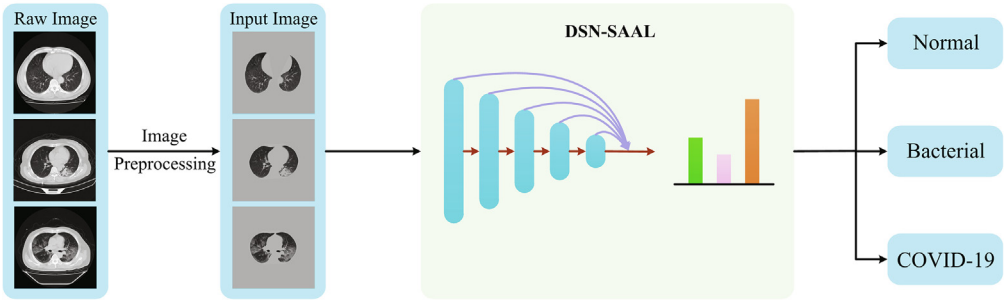

Fig. 1. Overview of the proposed COVID-19 diagnosis method.

convolutional features [30], we add a 21 × 1 × 1 convolution operation after each intermediate convolutional layer, and fuse these features with a 1 × 1 convolution operation in each stage [31]. In our architecture, each convolutional layer is closely followed by batch normalization (BN), which can effectively accelerate the convergence of the networks and improve the stability of training. Finally, an adaptive average pooling layer is used to modify the size of the feature map in each stage to 7 × 7, and then the classification is performed by a fully connected layer. The sum of all six losses, including the final classification loss of the VGG-16 network, is set as the final loss under deep supervised learning.

In particular, given a sample vector $\mathbf{x}$ with class label y, where $y \in \{1, \ldots, n, \ldots, C\}$ and C is the number of the classes, the loss of the architecture can be represented by

$$\ell(\mathbf{x}, y) = \sum_{i=1}^{N} \ell_i(\mathbf{x}, y) + \ell_{final}(\mathbf{x}, y) \tag{1}$$

where ℓ_i is the loss of the i^{th} auxiliary classifier, ℓ_{final} is the final loss of the network, and N is the number of auxiliary losses. Since ℓ_i and ℓ_{final} are set as the same form, Eq. (1) can be rewritten as

$$\ell(\mathbf{x}, y) = \sum_{i=1}^{N+1} \ell_i(\mathbf{x}, y) \tag{2}$$

which is used to minimize the difference between the network prediction and the ground-truth. Its optimization objective is

$$\arg\min_{(\mathbf{x},y)} \ell(\mathbf{x}, y) \tag{3}$$

As shown in Eq. (3), ℓ_i can be set as a suitable surrogate loss, such as cross entropy. In this study, we design a self-adaptive auxiliary loss as $\ell(\mathbf{x}, y)$.

3.3. Self-adaptive auxiliary loss

A novel self-adaptive auxiliary loss is proposed to help the training with imbalanced data by introducing a self-adaptive factor, which reflects the feature distribution and emphasizes minority classes. It is measured as the ratio of the effective number of samples to the total number of samples with a weighting regularization item, which is used for addressing the problem of noisy labels. Both of them are based on the cross entropy, as introduced below.

3.3.1. Cross entropy

The cross entropy for measuring the divergence between the output and ground-truth in the label space is computed by

$$\ell_{CE} = -\sum_{n=1}^{C} q(n|\mathbf{x}) \log p(n|\mathbf{x}) \tag{4}$$

Assuming $\{\mathbf{x}, y\}$ is an input-label pair, the ground-truth distribution over labels for sample $\mathbf{x}$ is in the term of $q(n|\mathbf{x})$, and it satisfies $\sum_{n=1}^{C} q(n|\mathbf{x}) = 1$. The inputs are transformed into a feature space representation with models and produces a set of probabilities as the output $p(n|\mathbf{x})$ via the Softmax function.

The cross entropy cannot solve the data imbalance problem in classification, since it treats minority classes in the same way as majority ones, resulting in the trained model biased towards majority classes. We further use the following techniques to address the problem.

3.3.2. Effective number of samples

We first leverage a novel theory proposed by [21] which argues that the sum of information from all samples in the dataset cannot be measured by the total number of samples. Particularly in medical images, such as CT scans, multiple images can be obtained after one CT scan of the same patient. On the one hand, because the slices are very similar to each other, especially in thin-slice scanning, where the slice thickness is only 1 mm, there is a great deal of overlap between the features provided by each image. On the other hand, each slice cannot be easily discarded to avoid the loss of information. Thus, the data overlap can be measured by the effective number of samples as follows:

$$\mathbf{E} = \left\{ \left(\frac{1 - \alpha^{k_n}}{1 - \alpha} \right)^{(n)} \right\}_{n=1}^{C} \tag{5}$$

where α is the effective sample factor to measure the ratio of the effective number of samples, and k_n is the number of samples in k^{th} class. In fact, α is used to control the rate of increase of the effective number of samples when k_n increases. In our model, α is updated in the network as a learnable parameter. After random initialization, α is activated and modified by Sigmoid function as an effective sample factor. The detailed process of parameter optimization is described in Section 3.4. As shown in Eq. (5), there is asymptotic property that $E_{(n)} = 1$ if $\alpha = 0$, which means there is only one valid sample of the class and all other samples of the class provide the same features. And $E_{(n)} \to k_n$ if $\alpha \to 1$, it can be obtained from L 'Hopital's rule that:

$$\lim_{\alpha \to 1} E = \lim_{\alpha \to 1} \frac{1 - \alpha^{k_n}}{1 - \alpha} = \lim_{\alpha \to 1} \frac{-k_n \alpha^{k_n - 1}}{-1} = k_n \tag{6}$$

The number of valid samples of the class is approximately equal to the total number of samples of the class, that is, the features provided by each sample are different and valid. We skip the detailed proof here, as it has been given in [21].

Different from previous methods weighting the loss by the inverse of the number of samples of the class, we set the weight as the inverse of E_n, which can obtain better performance. Considering the effective number of samples, the formulation of loss can be defined by

$$\ell_{DCE} = -\sum_{n=1}^{C} \frac{1-\alpha}{1-\alpha^{k_n}} q(n|\mathbf{x}) \log p(n|\mathbf{x}) \tag{7}$$

3.3.3. Regularization

In clinical datasets, mislabeling is a problem that can easily occur, such as labeling other pneumonia as COVID-19, which is mainly reflected in the fact that COVID-19 shares some common characteristics with other pneumonia caused by similar viruses. Meanwhile, in a group of CT slices, not every slice has the discriminative characteristics of COVID-19, which leads to a poor generalization of the models learned by mislabeling. In the context of noisy labels, some samples' labels are incorrect, namely $q(n|\mathbf{x})$ does not represent the true class distribution. Instead, $p(n|\mathbf{x})$ can reflect the true distribution to some extent. Inspired by the study in [24], we utilize $p(n|\mathbf{x})$ as the ground-truth and $q(n|\mathbf{x})$ as the class probability of the outputs, referring as the reverse cross entropy. The reverse cross entropy for the sample $\mathbf{x}$ is computed as follows:

$$\ell_{RCE} = -\sum_{n=1}^{C} p(n|\mathbf{x}) \log q(n|\mathbf{x}) \tag{8}$$

Same as the ℓ_{DCE}, we use the inverse of E_n as the weight by

$$\ell_{DRCE} = -\sum_{n=1}^{C} \frac{1-\alpha}{1-\alpha^{k_n}} p(n|\mathbf{x}) \log q(n|\mathbf{x}) \tag{9}$$

When labels are one-hot, computational problems might exist as the distribution $q(n|\mathbf{x}) = 0$. To solve this problem, we set $\log 0 = A$, where A is a certain constant and it satisfies $A < 0$ [24]. This approach uses less bias into the model at finite number of points like $q(n|\mathbf{x}) = 0$, but no bias at $q(n|\mathbf{x}) = 1$.

Since ℓ_{DRCE} and ℓ_{DCE} play different roles in calculating the difference between the prediction and the ground-truth, a hyper-parameter β is used to balance them. Due to the difference of the ground-truth between ℓ_{DRCE} and ℓ_{DCE}, the sample size of each class corresponding to the ground-truth is also discrepant, which is the same as the effective number of samples. While the process of reconstructing the RCE label is not feasible, β can help to adjust it with the network accordingly.

3.3.4. The proposed loss function

Based on the above presentation, the proposed self-adaptive auxiliary loss can be formulated by combining Eqs. (7) and (9) as follows:

$$\begin{aligned} \ell_{SAAL} = & -\sum_{n=1}^{C} \tfrac{1-\alpha}{1-\alpha^{k_n}} q(n|\mathbf{x}) \log p(n|\mathbf{x}) \\ & -\beta \sum_{n=1}^{C} \tfrac{1-\alpha}{1-\alpha^{k_n}} p(n|\mathbf{x}) \log q(n|\mathbf{x}) \end{aligned} \tag{10}$$

The first term on the right of Eq. (10) is the cross entropy weighted by the effective number of samples, and the second term is the reverse cross entropy weighted by the effective sample number as the regularization term. α and β are the hyper-parameters, which can be learned automatically through the network. The proposed self-adaptive auxiliary loss can learn the data distributions of both majority and minority classes by considering the effective number of samples to reweight the cross entropy loss and introducing a reverse cross entropy as a regularization term to handle incorrect labels.

3.4. Parameter optimization

Our goal is to jointly learn the network weight θ and the hyper-parameters α and β. As shown in Eq. (10), α and β represent the effective sample factor and weight of ℓ_{DRCE} respectively. ℓ_{DRCE} is based on cross entropy changing the position of $p(n|\mathbf{x})$ and $q(n|x)$ in form. Given the case of noise labels, the prediction $p(n|\mathbf{x})$ of the network can reflect the distribution of original data more than the ground-truth $q(n|x)$. Since the output in different stages have different levels of reflection to data distribution, we set up a set of weighted item β in the different stages. Different with β, the value of α is not related to the training process. It is the effective number of samples distribution reflecting the original data, so we only set it as a single parameter. For both types of parameters, we use the stochastic gradient descent with the back-propagation of error to update them. The whole iterative optimization process for the parameters is shown in Algorithm 1.

Algorithm 1 Iterative Optimization for Parameters (θ, α, β)

Input: Training set $(\mathbf{x}, y)$, Maximum epochs (M_e), Number of batches B, Learning rate γ, Self-adaptive auxiliary loss L
Output: $(\theta^*, \alpha^*, \beta^*)$

```
Net ←deep supervised network
θ ←a pretrained model of ImageNet is used on the backbone network and random initialization is used on the auxiliary chain
{β_1, ..., β_6, α} ← random initialization
out ← 0, loss ← 0
for e ∈ [1, M_e] do
    for b ∈ [1, B] do
        for i ∈ [1, 6] do
            out_i ←forward (x, y, Net, θ)
            lost_i ← L(out_i, y, Net, θ)
            loss = loss + lost_i
        end for
        grad_b ←backward (loss, Net, θ, α, β)
        θ*, α*, β* ←update (Net, θ, α, β, grad_b, γ)
        θ, α, β ← θ*, α*, β*
    end for
end for
return θ*, α*, β*
```

In the experiments, we transfer the parameters α and β to $1/(1+e^{-\alpha})$ and $1/(1+e^{-\beta})$ respectively, since they can increase the corresponding loss to a large value potentially. During the training of the network, the loss can make the training procedure unstable and lead to the non-convergence of the loss function. Therefore, we introduce the form of exponential function to compress the weight to $[0, 1)$.

4. Experimental protocol

In this section, we describe the details of our COVID-19-Diag dataset and three publicly available datasets including the COVIDx-CT [14], COVID19-CT [11], and SARS-CoV-2 CT-scan [19] datasets, as well as evaluation metrics and implementation details.

4.1. Data

4.1.1. Our COVID19-Diag dataset

In this study, we create a new COVID-19 dataset, named COVID19-Diag, which consists of 69 CT volumes of COVID-19, 95 CT volumes of normal cases, and 62 CT volumes of bacterial pneumonia from the First Hospital of Changsha[2]. The CT volumes of all the cases are performed on a CT scanner as SIEMENS or GE MEDICAL

[2] https://github.com/MLMIP/COVID19-Diag

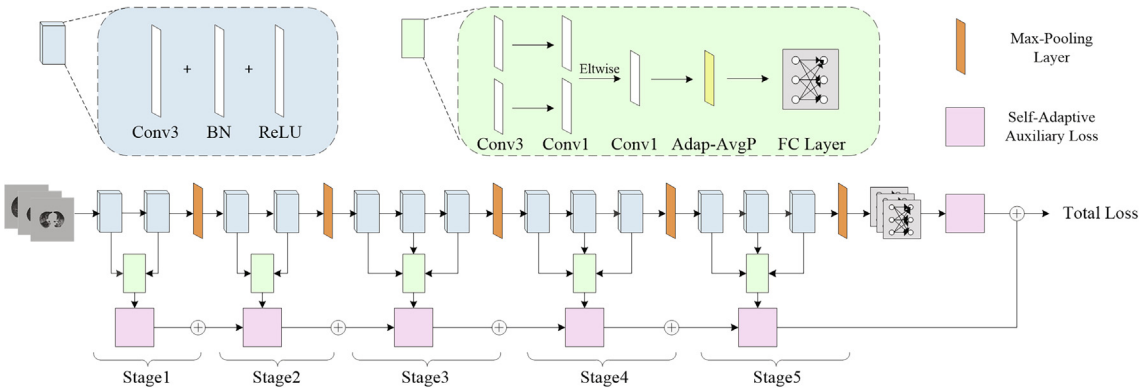

Fig. 2. The architecture of the proposed deep supervised network with self-adaptive auxiliary loss.

SYSTEMS with 5 mm of slice thickness and 512 × 512 of the matrix. We extract 1769, 3824, 1389 two dimensional (2D) CT axial slices from COVID-19, normal cases, and bacterial pneumonia respectively, in which the number of slices selected for each CT scan ranges from 2 to 54. A training set and a test set are randomly divided by cases with a ratio of 7 to 3. The size of the images are set to $1 \times 224 \times 224$ to accommodate the input of our model. The statistics of the dataset are shown in Table 2. It is worth mentioning that in the process of collecting and constructing our COVID19-Diag dataset, we do not deliberately enlarge the imbalance ratio of the number of samples in each category, but the problem of data imbalance has a great impact on the process of using data-driven deep learning methods to the diagnosis of COVID-19. In order to further verify the classification performance of our method for imbalanced data, we adopt different proportions of COVID-19 images for experiments, as shown in Section 5.3.

Fig. 3 shows the samples of the three classes from the dataset. The images we collected come from various positions in CT scans of the lungs, and each section contains different sizes of lung regions. We can find that some images such as the last column in bacterial pneumonia and COVID-19 samples have focal areas but are not obvious. There are also similar features between bacterial pneumonia and COVID-19 such as the first and second column, both of which have GGO. Therefore, our dataset is challenging and representative for the diagnosis of COVID-19 and can be well used to verify the classification performance of the algorithm.

Table 2
Dataset split of our COVID19-Diag.

Item	Class	Training	Testing
Cases	COVID-19	43	18
	Normal	67	28
	Bacterial Pneumonia	48	21
Images	COVID-19	1256	513
	Normal	2674	1150
	Bacterial Pneumonia	980	409

4.1.2. Public COVID-19 diagnosis datasets

To verify the generalization ability of the proposed method in the diagnosis of COVID-19, three additional public challenging datasets are used. Among them, the first is the COVIDx-CT dataset [14], which is derived from CT imaging data collected by the China National Center for Bioinformation comprising 104,009 images across 1,489 patient cases. The COVIDx-CT dataset contains 21395 COVID-19 (NCP) slices (12520 for training, 4529 for validation, and 4346 for testing), 36856 Common pneumonia (CP) slices (22061 for training, 7400 for validation, and 7395 for testing), and

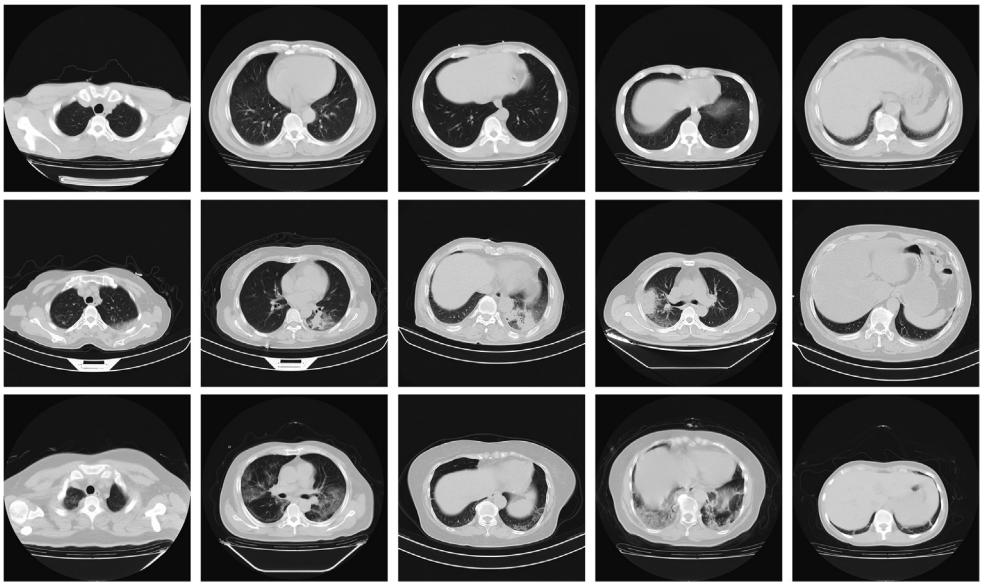

Fig. 3. Samples of normal (top row), bacterial pneumonia (middle row), and COVID-19 (bottom row).

45758 Normal slices (27201 for training, 9107 for validation, and 9450 for testing). The COVIDx-CT dataset is currently the largest dataset used for the diagnosis of COVID-19 with two-dimensional slices, which contains three categories and has a certain diversity.

The second is the COVID19-CT dataset [11], which contains 349 COVID-19 images (191 for training, 60 for validation, and 98 for testing) and 397 non-COVID-19 images (234 for training, 58 for validation, and 105 for testing). The COVID19-CT dataset was collected from some articles about COVID-19 diagnosis on medRxiv and bioRxiv services. The data volume is relatively small, and there is a large difference between these images since they come from different sources. Therefore, this dataset has certain challenges and research value.

The last is the SARS-CoV-2 CT-scan dataset [19], which is the clinical dataset from a hospital including 1252 for positive novel coronavirus infection and 1229 for patients non-infected, where 80% of images is used for training and the remaining is for validation. For a fair comparison, we apply fivefold cross-validation to report the results on the SARS-CoV-2 CT-scan dataset.

4.2. Evaluation metrics

We use the overall classification accuracy (ACC), F1-score, and G-mean as the main evaluation metrics, where the F1-score and G-mean are important indexes to evaluate the problem of data imbalance. In addition, the area under the ROC curve (AUC), Sensitivity (SEN), Specificity (SPE), and Precision(PRE) are required to evaluate the corresponding datasets to be consistent with other methods. These metrics are defined by Eqs. (11)–(16), respectively.

$$ACC = \frac{Right}{All} \tag{11}$$

$$F1 - score = \frac{2 \cdot PRE \cdot SEN}{PRE + SEN} \tag{12}$$

$$G - mean = \sqrt{SEN \cdot SPE} \tag{13}$$

$$SEN = \frac{TP}{TP + FN} \tag{14}$$

$$SPE = \frac{TN}{TN + FP} \tag{15}$$

$$PRE = \frac{TP}{TP + FP} \tag{16}$$

where the *Right* and *All* represent the number of correctly classified samples and total samples, respectively. TP, TN, FP, and FN are corresponding to true positives, true negatives, false positives, and false negatives, respectively.

4.3. Implementation details

We use our proposed deep supervised network to learn the discriminative feature representations for the image classification task (see Fig. 2 for details). The backbone of the network is consists of a VGG-16 network with batch normalization layers. We initialize them using pre-trained models on ImageNet [32], which is of great significance for the improvement of the convergence and performance of the model with the help of parameters trained under a large dataset [33]. Like the other layers of the network, random initialization of the weights and biases are adopted.

In the experiments, we construct a baseline CNN for comparison using the VGG-16 with batch normalization but without the adaptive auxiliary loss for network training. Similarly, we use the pre-trained model on ImageNet to initialize the weights and biases of the baseline, and cross entropy is adopted as the loss to help the network convergence. We calculate the confidence intervals, including the mean and standard deviation, when conducting experiments on our COVID19-Diag dataset. Each value is a comprehensive evaluation of the results of 10 experiments.

For the COVIDx-CT, COVID19-CT, and SARS-CoV-2 CT-scan datasets, because the size of the images is various and not suitable for the input of the network, we resize them to $3 \times 224 \times 224$ using the bilinear interpolation, respectively. The output of the network depends on the number of categories in each dataset. Stochastic gradient descent (SGD) with a momentum 0.9 and weight decay 5×10^{-4} are adopted as the optimizer. The learning rate is initialized as 0.001 and divided by 10 every 40 epochs (120 epochs in total). We implement our model based on PyTorch 1.3.0, and all experiments are performed on an NVIDIA GeForce RTX 2080Ti 11G.

5. Results

5.1. Overall performance on Our COVID19-Diag dataset

In order to verify the performance of DSN-SAAL on our COVID19-Diag dataset, we compare our method with some common used classification models with CE including VGG-16 [29], ResNet-50 [34], DenseNet-169 [35], MobileNet-V2 [36], and ResNeXt-50 [37]. As seen from Table 3, DSN-SAAL outperforms all competing methods in all evaluation metrics, and it improves the ACC, F1-score, and G-mean by 5.3–8.4%, 7.6–12.5%, and 5.2–8.7%, respectively when compared with other models. The ROC curves of all models are shown in Fig. 4. We can observe that the red curve of DSN-SAAL is clearly above all the other curves.

To evaluate the effectiveness of our DSN-SAAL, we have also reproduced three works (i.e., Self-Trans [11], Transfer-CheXNet [13], and Meta-Weight-Net [27]) for automatic diagnosis of COVID-19, as shown in Table 3. The codes for all of them are publicly available. For a fair comparison, the settings of the parameters remain the same in accordance with the relevant official codes. We only adjust the input of the model to single-channel images, so as to adapt to our COVID19-Diag dataset. As to the Meta-Weight-Net, we have made some modifications according to its main architecture and add some convolutional layers and sub-sampling layers to adapt to the 224×224 input. The size of the final output feature map is 7×7. From Table 3, we can see that our DSN-SAAL achieves high performance in the diagnosis of COVID-19 and outperforms the competing methods in all metrics.

5.2. Ablation study on DSN-SAAL

To better validate the role of each component of our model in the diagnosis of COVID-19, we design a set of ablation experiments, including a deep supervised network (DSN) and a self-adaptive auxiliary loss (SAAL) study.

We consider four commonly used CNNs including VGG-16, ResNet-50, DenseNet-169, and ResNeXt-50, which are staged or modular and suitable for joining the auxiliary supervised chain. The comparison results between baseline and DSN are shown in Table 4. The results show that ACC, F1-score, and G-mean are significantly improved compared with the baseline after the addition of auxiliary supervision chain in the network, while the number of parameters is only slightly increased, not more than 0.1 M (Flops-counter.pytorch[3]). We believe that a slight increase in the number of parameters is acceptable compared to a large increase in the

[3] https://github.com/sovrasov/flops-counter.pytorch

Table 3
Performance comparison of DSN-SAAL with other classification models on our COVID19-Diag dataset. The best results are highlighted in bold.

Model	ACC	F1-score	G-mean
VGG-16 [29]	0.836 ± 0.013	0.748 ± 0.018	0.849 ± 0.008
ResNet-50 [34]	0.837 ± 0.009	0.763 ± 0.017	0.838 ± 0.017
DenseNet-169 [35]	0.844 ± 0.002	0.767 ± 0.004	0.862 ± 0.001
MobileNet-V2 [36]	0.867 ± 0.008	0.797 ± 0.010	0.873 ± 0.008
ResNeXt-50 [37]	0.846 ± 0.012	0.765 ± 0.017	0.850 ± 0.007
Self-Trans [11]	0.909 ± 0.016	0.866 ± 0.017	0.896 ± 0.020
Transfer-CheXNet [13]	0.899 ± 0.015	0.848 ± 0.017	0.885 ± 0.018
Meta-Weight-Net [27]	0.888 ± 0.005	0.825 ± 0.009	0.868 ± 0.016
DSN-SAAL (Proposed)	**0.920 ± 0.004**	**0.873 ± 0.007**	**0.925 ± 0.006**

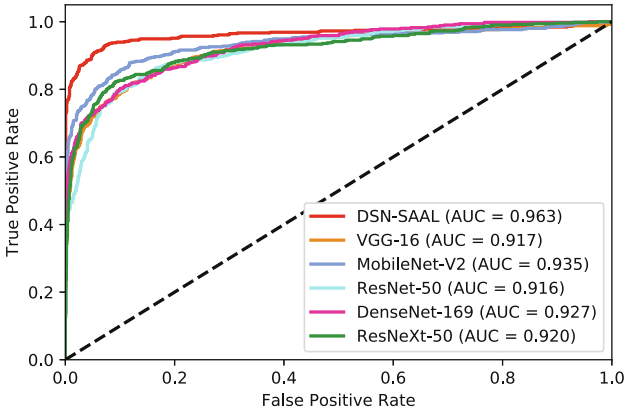

Fig. 4. The ROC curves of our DSN-SAAL and some popular models.

diagnosis result. We can conclude that the performance of deep neural networks can be effectively improved by adding auxiliary supervision.

Moreover, it can be seen that it is more beneficial to increase the depth supervision mechanism on the shallow and simple layer network than the deep layer network, because the complex network structure has the learning ability of different layer characteristics to a certain extent.

The comparison results of our SAAL and some other loss functions are shown in Table 5. From the results, we can observe that SAAL outperforms other loss functions by providing more accurate diagnosis performance on the COVID19-Diag dataset. Especially for the F1-score, our results increase even higher than other loss functions, showing the learning effect of SAAL on the features of each category. Furthermore, when DSN and SAAL are used in VGG-16, the ACC result is 0.920 ± 0.004, F1-score is 0.873 ± 0.007, and G-mean is 0.925±0.006, all of which are higher than either of them alone, reflecting the advantages of DSN-SAAL. In summary, our DSN-SAAL effectively integrates the shallow and deep information through the auxiliary supervision, which promotes the feature learning of the model. Besides, SAAL makes the model pay more attention to the learning of the minority classes, thus effectively solving the problem of data imbalance.

Table 5
Performance comparison of VGG-16 with different loss functions to solve the data imbalance on our COVID19-Diag dataset.

Method	ACC	F1-score	G-mean
CE	0.836 ± 0.013	0.748 ± 0.018	0.849 ± 0.008
Focal Loss [22]	0.851 ± 0.007	0.766 ± 0.011	0.855 ± 0.008
CB Loss [21]	0.861 ± 0.005	0.782 ± 0.007	0.868 ± 0.006
SCE [24]	0.856 ± 0.014	0.804 ± 0.014	0.885 ± 0.010
SAAL (Proposed)	**0.875 ± 0.005**	**0.832 ± 0.014**	**0.890 ± 0.010**

5.3. Evaluation on data imbalance problem

To further demonstrate the effectiveness of our model on the data imbalance problem, we increase the imbalance ratio of our COVID19-Diag dataset for comparison. Our goal is to get better results with a small number of COVID-19 samples, and thus we reduce the samples of the target class as COVID-19 to 25%, 10%, 5%, and 1% respectively, to evaluate our model against the baseline. The results are shown in Table 6. In order to better evaluate the classification results of the target class and non-target classes, sensitivity and specificity are used. In particular, we record changes in sensitivity at different levels of deletion with the samples of COVID-19 compared to the standard distribution in brackets. We observe that all indicators decline when the sample size of the target class decreases. Especially for F1-score and G-mean, which are the comprehensive indicators focusing on the learning of each class, show a large decline. However, our DSN-SAAL still has some advantages compared to baseline. With the increase of the imbalance ratio, the decline in sensitivity of DSN-SAAL is slower than that of baseline. Besides, the specificity is slightly improved. Fig. 5 shows the confusion matrix obtained by different degrees of imbalance ratios for the COVID-19 samples separately. Overall, the results show that our DSN-SAAL can effectively maintain the classification accuracy of minority classes without affecting the feature learning of majority classes.

5.4. Samples analysis and visualization of CAM

The studies reported that the pulmonary abnormalities on COVID-19 CT scans include bilateral and subpleural GGO, bronchovascular thickening, air space consolidation, traction bronchiectasis, pleural effusion, and crazy paving appearance. However, there are some overlaps between the biological characteristics of COVID-19 and other pneumonia in CT slices, such as GGO, space consolidation, and frantic pavement, which are common findings of COVID-19 and bacterial pneumonia on CT images.

Table 4
Performance comparison of whether to add auxiliary supervision in classification models on our COVID19-Diag dataset.

Model	Method	ACC	F1-score	G-mean	Params
VGG-16	Baseline	0.836 ± 0.013	0.748 ± 0.018	0.849 ± 0.008	134.28 M
	DSN	**0.894 ± 0.004**	**0.830 ± 0.010**	**0.901 ± 0.004**	134.32 M
ResNet-50	Baseline	0.837 ± 0.009	0.763 ± 0.017	0.838 ± 0.017	23.51 M
	DSN	**0.862 ± 0.007**	**0.790 ± 0.017**	**0.865 ± 0.009**	23.59 M
DenseNet-169	Baseline	0.844 ± 0.002	0.767 ± 0.004	0.862 ± 0.001	12.48 M
	DSN	**0.868 ± 0.009**	**0.792 ± 0.017**	**0.872 ± 0.011**	12.49 M
ResNeXt-50	Baseline	0.846 ± 0.012	0.765 ± 0.017	0.850 ± 0.007	22.98 M
	DSN	**0.887 ± 0.007**	**0.819 ± 0.011**	**0.888 ± 0.006**	23.06 M

Table 6
Performance comparison of DSN-SAAL with the baseline under both balanced and imbalanced distributions of our COVID19-Diag dataset.

Imbalanced	ACC		F1-score		G-mean		SEN		SPE	
	Base.	D.S.	Base.	D.S.	Base.	D.S.	Base.	D.S.	Base.	D.S.
Stand. split	0.836 ± 0.013	**0.920 ± 0.004**	0.748 ± 0.018	**0.873 ± 0.007**	0.849 ± 0.008	**0.925 ± 0.006**	0.822 ± 0.009	**0.907 ± 0.012**	0.877 ± 0.018	**0.944 ± 0.005**
25% of COVID-19	0.837 ± 0.004	**0.900 ± 0.006**	0.719 ± 0.005	**0.828 ± 0.012**	0.803 ± 0.006	**0.878 ± 0.006**	0.702 ± 0.013(-0.120)	**0.810 ± 0.016(-0.097)**	0.918 ± 0.006	**0.952 ± 0.013**
10% of COVID-19	0.832 ± 0.003	**0.889 ± 0.010**	0.692 ± 0.014	**0.793 ± 0.019**	0.758 ± 0.008	**0.829 ± 0.017**	0.602 ± 0.012(-0.220)	**0.702 ± 0.030(-0.205)**	0.955 ± 0.009	**0.979 ± 0.006**
5% of COVID-19	0.804 ± 0.005	**0.870 ± 0.007**	0.581 ± 0.015	**0.740 ± 0.029**	0.666 ± 0.011	**0.778 ± 0.029**	0.464 ± 0.015(-0.358)	**0.616 ± 0.055(-0.291)**	0.956 ± 0.005	**0.985 ± 0.014**
1% of COVID-19	0.731 ± 0.010	**0.773 ± 0.004**	0.141 ± 0.030	**0.373 ± 0.036**	0.278 ± 0.033	**0.480 ± 0.029**	0.080 ± 0.018(-0.742)	**0.232 ± 0.029(-0.675)**	0.983 ± 0.005	**0.997 ± 0.002**

Base.: Baseline; D.S.: DSN-SAAL

Besides, for some patients with mild disease in the early stage, there is no obvious lesion area in the CT slices, which is easy to cause false negative in the diagnosis process and cannot effectively prevent the progression of the disease. These conditions will affect the diagnosis of COVID-19.

Fig. 6 shows some samples that are diagnosed using our model. Figs. 6a and b are the slices where bacterial pneumonia samples are diagnosed as normal. It can be seen that there is no obvious lesion area, especially for samples like Fig. 6b, which is located at the top or bottom of the CT scans. The pulmonary parenchyma area itself is very small, but the lesion features cannot be ignored, which becomes a difficulty in model learning. For Figs. 6c and d, the bacterial pneumonia samples are diagnosed as COVID-19. The lesions of bacterial pneumonia appeared on the periphery of the slices, which is similar to the characteristics of the COVID-19 samples and is more prone to misclassification. Slices as shown in Figs. 6e and f indicate that the samples of COVID-19 are judged to normal. It can be seen that the features of the lesions in both of them are not obvious, and even appear to be very similar to pulmonary blood vessels. Figs. 6g and h represent the samples of the COVID-19 that are determined to be bacterial pneumonia. Similarly, the samples of the COVID-19 show similar characteristics with bacterial pneumonia. Especially for Fig. 6h, the lung parenchyma shows a large area of low-density GGO and is difficult to distinguish.

The results also show that our model has higher sensitivity in the diagnosis of COVID-19, which also indicates that the model has fewer false negatives and is less likely to be misdiagnosed. Although the deep learning model can distinguish COVID-19 from bacterial pneumonia and normal cases to a certain extent, the model is limited by the diversity and class imbalance of the training data, and our work is trying to solve these problems.

As shown in Fig. 7, the class activation mapping (CAM) [38] is used to visualize the attention regions on our COVID19-Diag dataset for VGG-16 with CE and DSN-SAAL. This can be obtained by the convolutional layer at the end of the models. As seen from the first and fifth columns, the raw images of the lesion area are not obvious, VGG-16 can not accurately distinguish between pulmonary vessels and lesion areas, resulting in a large area of the red area covering the pulmonary area. While our method can notice them more accurately. As shown in the second and fourth columns, we find that our method can better separate the lung lobes from the background and find the GGO more precisely. For a challenging sample as the third and sixth columns, DSN-SAAL can still distinguish the lung parenchyma and lesion areas to obtain more accurate results than VGG-16.

5.5. Evaluation on public COVID-19 diagnosis datasets

To verify the generalization of DSN-SAAL, we conduct comparative experiments on other three publicly available COVID-19 datasets. We use the data division which is mentioned in Section 4.1, and the results are shown in Tables 7–9. For each dataset, we compare our method with the state-of-the-art approaches, and the results of the comparison methods in the tables are all from the original papers.

For the COVIDx-CT dataset, we first conduct experiments under the original data distribution and compare them with relevant methods. It can be seen that our DSN-SAAL performs better than the other two approaches in terms of the overall accuracy and the SEN and PRE of the three categories. COVIDNet-CT [14] is the method proposed in conjunction with the original COVIDx-CT dataset, and the VisionPro [39] is deep learning software that has been widely used in various fields from factory automation to life science. As the COVIDx-CT dataset itself has a large number of slices, the samples in the three categories are relatively rich for

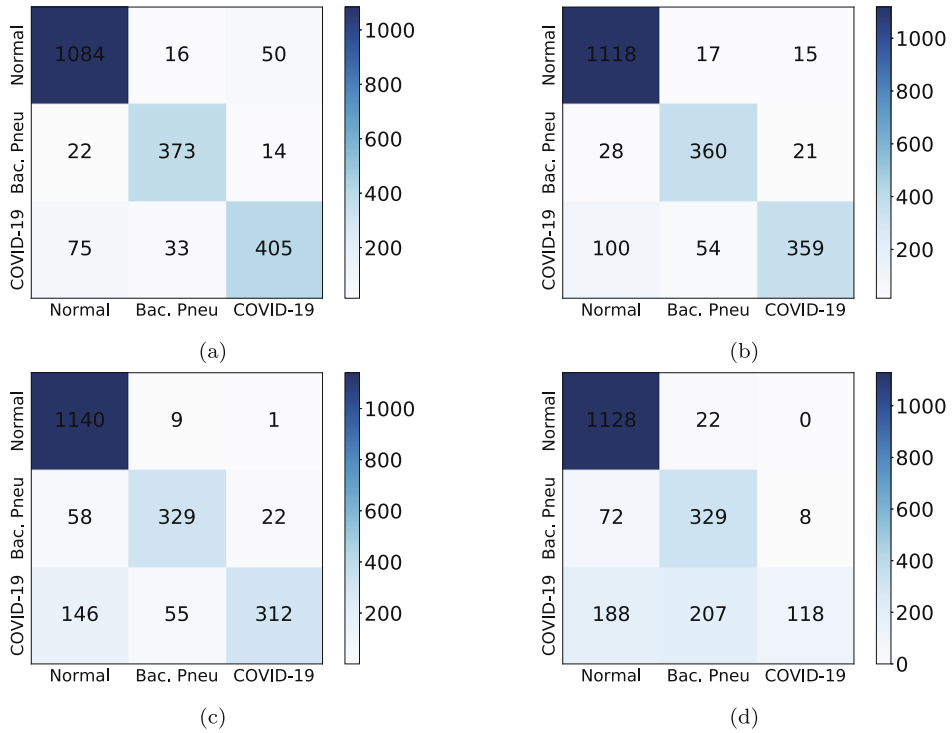

Fig. 5. Confusion matrix for DSN-SAAL on our COVID19-Diag dataset. Fig. 5a–d are the results with 25%, 10%, 5% and 1% of COVID-19 samples respectively.

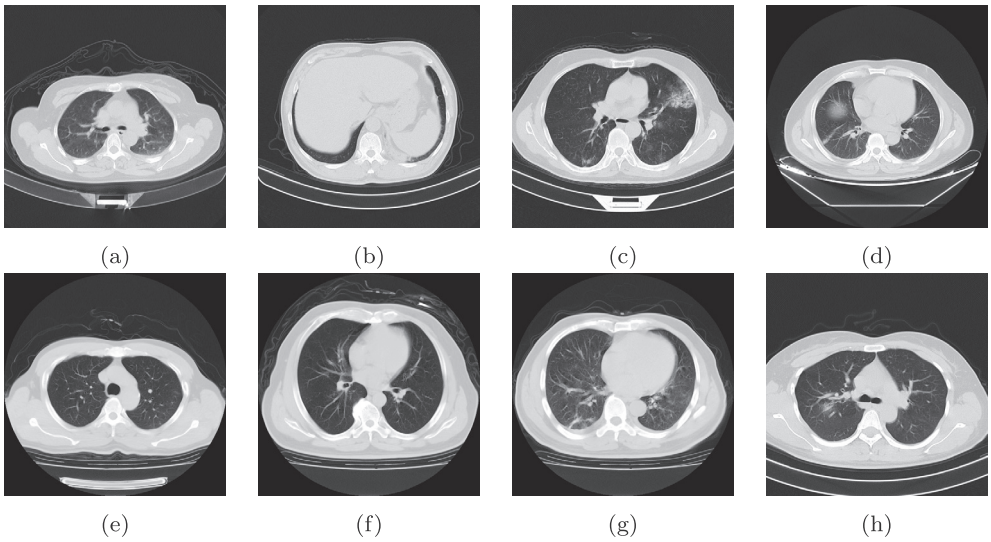

Fig. 6. Examples of the diagnostic results obtained using our DSN-SAAL.

the classification tasks, so the three methods can all get good results on this dataset, as shown in the first part of Table 7.

In order to better verify the effectiveness of our DSN-SAAL, we extract 1000 CT slices from each of the three categories of training sets for training, that is, only about 5% of the original training set is used for training, and the test set is kept unchanged for a fair evaluation. The samples taken are consistent with that in reference [40]. As shown in the second part of Table 7, COVID-CT-MaskNet [40], Two Stage Model [41], Lightweight Model [42], and One Shot Model [43] all adopted a two-stage learning strategy. First, the relevant regions of interest (ROI) containing GGO and consolidation shadows were obtained from the images by detection and segmentation methods, and then each category was distinguished by a classification network. From Table 7, we can observe that our DSN-SAAL outperforms the state-of-the-art approaches and can still maintain a high degree of differentiation for each category even when the dataset is greatly reduced. Besides, our DSN-SAAL is a single-stage network, which makes the training process more convenient than a multi-stage model.

For the COVID19-CT dataset, as shown in Table 8, it can be seen that although the AUC value of our DSN-SAAL is slightly lower than that of Self-Trans, it still ranks 2nd, and our results in ACC and F1-score are higher than that of Self-Trans. It is worth mentioning that, for Self-Trans, 1000 additional unlabeled CT slices from the Lung Nodule Analysis (LUNA) database were trained with the pre-trained model on the ImageNet dataset, and then the COVID19-CT dataset was trained on the obtained model to complete the final classification task. However, our method does not

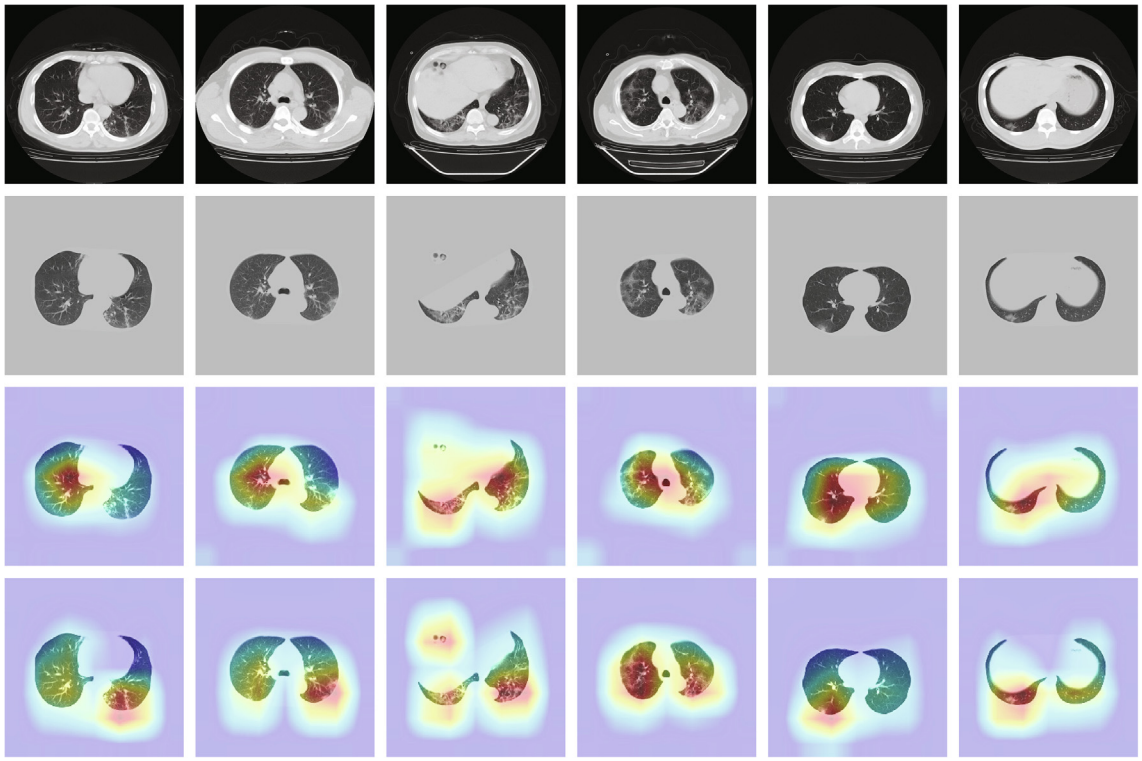

Fig. 7. Visualization of raw images with COVID-19 (the first row), the input images (the second row), and CAMs obtained by VGG-16 with CE (the third row) and our DSN-SAAL (the fourth row).

Table 7
Comparison with the state-of-the-art methods on the COVIDx-CT dataset.

Section	Method	ACC	SEN (Normal/CP/NCP)	PRE (Normal/CP/NCP)
Stand.split	COVIDNet-CT [14]	0.9911	1.0000/0.9904/0.9731	0.9940/0.9844/0.9969
	VisionPro [39]	0.9960	0.9992/0.9922/0.9959	0.9962/0.9966/0.9947
	DSN-SAAL (Proposed)	**0.9987**	**1.0000/0.9977/0.9977**	**0.9983/0.9988/0.9995**
5% of training set	COVID-CT-Mask-Net [40]	0.9166	0.9110/0.9162/0.9080	0.9433/0.8708/0.9475
	Two Stage Model [41]	0.9564	0.9691/0.9506/0.9388	0.9766/0.9300/0.9588
	Lightweight Model [42]	0.9395	0.9698/0.9163/0.9135	–
	One Shot Model [43]	–	0.9927/0.9813/0.9574	–
	COVIDNet-CT [14]	0.9757	–/–/0.9249	–
	DSN-SAAL (Proposed)	**0.9891**	**0.9998/0.9831/0.9758**	**0.9870/0.9867/0.9976**

Table 8
Comparison with the state-of-the-art methods on the COVID19-CT dataset.

Method	ACC	F1-score	AUC
VGG-16 [29]	0.76	0.76	0.82
ResNet-50 [34]	0.80	0.81	0.88
DenseNet-169 [35]	0.83	0.81	0.87
Self-Trans [11]	0.86	0.85	**0.94**
Contrastive-COVIDNet [44]	0.79	0.79	0.85
Transfer-CheXNet [13]	0.87	**0.86**	0.75
Cross-Datasets Analysis [45]	**0.88**	**0.86**	0.91
DSN-SAAL (Proposed)	0.87	**0.86**	0.91

use additional data for training. In addition, our ACC value is slightly lower than Cross-datasets Analysis [45], but we get the highest F1-score. Furthermore, we can find that our DSN-SAAL significantly outperforms the existing methods including Cross-datasets Analysis [45] in each index, when conducting experiments on the SARS-CoV-2 CT-scan dataset (see Table 9). It effectively verifies that our DSN-SAAL has better generalization performance than the state-of-the-art methods.

We further conduct experiments on another dataset, i.e., SARS-CoV-2 CT-scan. Table 9 shows the comparison of results between DSN-SAAL and a series of classical traditional machine learning methods and existing deep learning models. Among them, xDNN combined deep neural networks with prototype learning, aiming to propose an interpretable deep learning model for the automatic diagnosis of COVID-19. MAD-DBM [46] used a deep bidirectional long short-term memory network with a mixture density network model as a real-time COVID-19 diagnostic system. It can be found from Table 9 that DSN-SAAL is still superior to other popular traditional machine learning and deep learning methods in each evaluation index, which effectively verifies the performance of our model.

6. Discussion

Deep learning has proven to be an effective tool for assisting the diagnosis of COVID-19 due to its rapid and accurate characteristics. However, data volume and diversity have a profound impact on the performance of deep learning models [49] (see Table 6). On the one hand, there are relatively few public datasets available on the diag-

Table 9
Comparison with the state-of-the-art methods on the SARS-CoV-2 CT-scan dataset.

Method	ACC	F1-score	AUC	PRE	SEN
AdaBoost	0.9516	0.9514	0.9519	0.9363	0.9671
Decision Tree	0.7944	0.7984	0.7951	0.7681	0.8313
AlexNet [47]	0.9375	0.9361	0.9368	0.9498	0.9228
VGG-16 [29]	0.9496	0.9497	0.9496	0.9402	0.9543
GoogleNet [48]	0.9173	0.9182	0.9179	0.9020	0.9350
ResNet [34]	0.9496	0.9503	0.9498	0.9300	0.9715
xDNN [19]	0.9738	0.9731	0.9736	0.9916	0.9553
Contrastive-COVIDNet [44]	0.9083	0.9087	0.9624	0.9575	0.8589
Cross-Datasets Analysis [45]	0.9889	–	–	0.9920	0.9880
MADE-DBM [46]	0.9837	0.9814	0.9832	0.9874	0.9887
DSN-SAAL (Proposed)	**0.9943**	**0.9944**	**0.9995**	**0.9952**	**0.9936**

nosis of COVID-19, most of which are intended to distinguish between normal and COVID-19. Although CT scan images of COVID-19 are distinctly different from normal images, there are many common manifestations with other pneumonia. On the other hand, the sample size of the original target class is much smaller than that of the non-target class. In the diagnosis of COVID-19, there are often few COVID-19 targets, which leads to the problem of data imbalance, but the clinical need to distinguish them effectively. However, most of the existing deep learning methods for diagnosing COVID-19 did not consider the problem of data imbalance or simply amplify the data through affine transformation, and the diversity of samples did not increase. Taking the problems mentioned above into account, we first collected and created a category diversity dataset for COVID-19 diagnosis. Second, we proposed a novel method called DSN-SAAL, which can better distinguish COVID-19 from normal and bacterial pneumonia in the case of data imbalance.

The image features learned by the network at different stages are diverse and need to be effectively utilized. In this paper, we integrated shallow features and deep features through auxiliary supervision to promote the simultaneous learning of minority and majority class features. Furthermore, considering the similarity between different slices of CT scans and the possible mislabeling of clinical data, we designed an adaptive auxiliary loss for supervision, which is effectively combined with the deep supervision network to promote the learning of minority class features. Tables 4 and 5 show the advantages of deep supervision network and adaptive auxiliary loss over baseline, respectively. The results illustrated in Table 3 and Fig. 4 also show that our method has great advantages when compared to popular deep learning models.

To fully demonstrate the superiority of our method under the imbalance problem, we designed a series of comparison experiments under the imbalanced ratio (see Table 6). It can be seen that our method has better stability than the baseline in the case of an increased imbalance ratio. We also show the confusion matrix of different proportions of COVID-19 samples (see Fig. 5). It can be found that with the decrease of COVID-19 samples, the performance of our model decreases correspondingly. However, it is worth mentioning that when the COVID-19 samples are 1% of the original ones, that is, only 13 samples are used for training at this time, our model still has a certain recognition ability of the COVID-19 in the same test set. To verify the generalization ability of our model, some experiments were conducted on other three public COVID-19 diagnosis datasets. The results as shown in Tables 7–9 verified the superiority of the proposed method. Besides, CAMs showed that our model can focus on the pulmonary parenchymal areas to further find relevant lesion areas more accurately, even if the lesion area is not obvious (see Fig. 7).

Although we have demonstrated that our model performed well in the COVID-19 diagnosis, there are still some limitations. First, the COVID19-Diag dataset is limited. Compared with existing COVID-19 diagnostic datasets, our COVID19-Diag has some advantages in data volume, but it is still not enough, which will have a certain impact on the training of deep learning models. We plan to evaluate our method using additional CT scans from more centers in the future. Second, our method only focuses on the identification of COVID-19 and does not quantify the lesion area for analyzing the severity to help clinicians make further diagnoses. Thus, we are going to look at that to help with monitoring and treatment in future research. We also plan to replicate our model on the open source deep learning platform paddle.

7. Conclusion

In this paper, we create a challenging clinical dataset named COVID19-Diag and propose a novel deep supervised learning using self-adaptive auxiliary loss for COVID-19 diagnosis from imbalanced CT images. We first present a novel deep supervised network for multi-scale feature learning of imbalanced data (i.e., the equivalence learning of majority and minority classes). Then, we propose an efficient self-adaptive auxiliary loss by considering the effective number of samples and the regularization item with an RCE. Our method can be applied to different datasets since all model parameters are automatically learned through the network iteration.

Finally, the results on our COVID19-Diag and three publicly available COVID-19 diagnosis datasets show that using a convolutional neural network without any data amplification can effectively identify COVID-19 from imbalanced CT images.

CRediT authorship contribution statement

Kai Hu: Supervision, Conceptualization, Methodology, Visualization, Formal analysis, Validation, Software, Writing - review & editing. **Yingjie Huang:** Conceptualization, Methodology, Visualization, Formal analysis, Validation, Software, Writing - review & editing. **Wei Huang:** Data curation, Writing - review & editing. **Hui Tan:** Methodology, Visualization, Validation, Writing - review & editing. **Zhineng Chen:** Methodology, Visualization, Formal analysis, Writing - review & editing. **Zheng Zhong:** Data curation, Writing - review & editing. **Xuanya Li:** Methodology, Visualization, Writing - review & editing. **Yuan Zhang:** Methodology, Visualization, Formal analysis, Validation, Writing - review & editing. **Xieping Gao:** Supervision, Methodology, Visualization, Formal analysis, Validation, 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.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants 61802328, 61972333 and 61771415, the Natural Science Foundation of Hunan Province of China under Grant 2019JJ50606, the Research Foundation of Education Department of Hunan Province of China under Grant 19B561, and the Baidu Pinecone Program.

References

[1] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19, IEEE Reviews in Biomedical Engineering 14 (2020) 4–15.
[2] X. Xu, C. Yu, J. Qu, L. Zhang, S. Jiang, D. Huang, B. Chen, Z. Zhang, W. Guan, Z. Ling, et al., Imaging and clinical features of patients with novel coronavirus sars-cov-2, European Journal of Nuclear Medicine and Molecular Imaging 47 (5) (2020) 1275–1280.
[3] K. Hu, B. Shen, Y. Zhang, C. Cao, F. Xiao, X. Gao, Automatic segmentation of retinal layer boundaries in oct images using multiscale convolutional neural network and graph search, Neurocomputing 365 (2019) 302–313.
[4] K. Hu, K. Chen, X. He, Y. Zhang, Z. Chen, X. Li, X. Gao, Automatic segmentation of intracerebral hemorrhage in ct images using encoder–decoder convolutional neural network, Information Processing & Management 57 (6) (2020) 102352.
[5] M. Buda, A. Maki, M.A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural Networks 106 (2018) 249–259.
[6] C.L. Castro, A.P. Braga, Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data, IEEE Transactions on Neural Networks and Learning Systems 24 (6) (2013) 888–899.
[7] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, et al., A deep learning system to screen novel coronavirus disease pneumonia, Engineering 6 (10) (2020) 1122–1129.
[8] Gozes O., Frid-Adar M., Greenspan H., Browning P.D., Zhang H., Ji W., Bernheim A., Siegel E., Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis, arXiv preprint arXiv:2003.05037 (2020).
[9] S.H. Khan, M. Hayat, M. Bennamoun, et al., Cost-sensitive learning of deep feature representations from imbalanced data, IEEE Transactions on Neural Networks and Learning Systems 29 (8) (2018) 3573–3587.
[10] F. Shi, L. Xia, F. Shan, B. Song, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. He, Y. Gao, Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification, Physics in Medicine & Biology 66 (6) (2021) 065031.
[11] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, P. Xie, Sample-efficient deep learning for covid-19 diagnosis based on ct scans, MedRxiv (2020).
[12] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Zha, J. Shen, Y. Chong, Y. Yang, Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2021).
[13] C. Li, Y. Yang, H. Liang, B. Wu, Transfer learning for establishment of recognition of covid-19 on ct imaging using small-sized training datasets, Knowledge-Based Systems 218 (2021) 106849.
[14] 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 7 (2020) 1025.
[15] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography, Scientific Reports 10 (1) (2020) 1–11.
[16] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, et al., Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks, MedRxiv (2020).
[17] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, C. Zheng, A weakly-supervised framework for covid-19 classification and lesion localization from chest ct, IEEE Transactions on Medical Imaging 39 (8) (2020) 2615–2625.
[18] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al., A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19), European Radiology (2021) 1–9.
[19] P. Angelov, E. Almeida Soares, SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification, MedRxiv (2020).
[20] M. Hayat, S. Khan, S.W. Zamir, J. Shen, L. Shao, Gaussian affinity for max-margin class imbalanced learning, IEEE/CVF International Conference on Computer Vision (ICCV) 2019 (2019) 6468–6478.
[21] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, S. Belongie, Class-balanced loss based on effective number of samples, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 9268–9277.
[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 2980–2988.
[23] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, European Conference on Computer Vision (ECCV), Springer, 2016, pp. 499–515.
[24] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, J. Bailey, Symmetric cross entropy for robust learning with noisy labels, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE, 2019, pp. 322–330.
[25] X. Zhang, Z. Fang, Y. Wen, Z. Li, Y. Qiao, Range loss for deep face recognition with long-tailed training data, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 5409–5418.
[26] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface, Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 4690–4699.
[27] J. Shu, Q. Xie, L. Yi, Q. Zhao, S. Zhou, Z. Xu, D. Meng, Meta-weight-net: Learning an explicit mapping for sample weighting, NeurIPS (2019) 1–23.
[28] X. Li, Y. Zhou, P. Du, G. Lang, M. Xu, W. Wu, A deep learning system that generates quantitative ct reports for diagnosing pulmonary tuberculosis, Applied Intelligence (2020) 1–12.
[29] Simonyan K., Zisserman A., Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
[30] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, X. Bai, Richer convolutional features for edge detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 3000–3009.
[31] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, X. Gao, Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function, Neurocomputing 309 (2018) 179–191.
[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 248–255.
[33] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2014, pp. 580–587.
[34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016 (2016) 770–778.
[35] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017 (2017) 2261–2269.
[36] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, Mobilenetv 2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2018, pp. 4510–4520.
[37] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 1492–1500.
[38] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 618–626.
[39] A. Sarkar, J. Vandenhirtz, J. Nagy, D. Bacsa, M. Riley, Detection of covid-19 from chest computed tomography (ct) images using deep learning: Comparing cognex visionpro deep learning 1.0 software with open source convolutional neural networks, arXiv preprint arXiv:2010.00958 (2020).
[40] A. Ter-Sarkisov, Covid-ct-mask-net: Prediction of covid-19 from ct scans using regional features, MedRxiv (2020).
[41] A. Ter-Sarkisov, Detection and segmentation of lesion areas in chest ct scans for the prediction of covid-19, MedRxiv (2020).
[42] A. Ter-Sarkisov, Lightweight model for the prediction of covid-19 through the detection and segmentation of lesions in chest ct scans, International Journal of Automation, Artificial Intelligence and Machine Learning 2 (1) (2021) 01–15.
[43] A. Ter-Sarkisov, One shot model for covid-19 classification and lesions segmentation in chest ct scans using lstm with attention mechanism, MedRxiv (2021).
[44] Z. Wang, Q. Liu, Q. Dou, Contrastive cross-site learning with redesigned net for covid-19 ct classification, IEEE Journal of Biomedical and Health Informatics 24 (2020) 2806–2813.
[45] P. Silva, E. Luz, G. Silva, G. Moreira, R. Silva, D. Lucio, D. Menotti, Covid-19 detection in ct images with deep learning: A voting-based scheme and cross-datasets analysis, Informatics in Medicine Unlocked 20 (2020) 100427.
[46] Y. Pathak, P.K. Shukla, K. Arya, Deep bidirectional classification model for covid-19 disease infected patients, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2020).
[47] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems 25 (2012) 1097–1105.
[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015, pp. 1–9.
[49] Y. Zhao, A. Gafita, B. Vollnberg, G. Tetteh, F. Haupt, A. Afshar-Oromieh, B. Menze, M. Eiber, A. Rominger, K. Shi, Deep neural network for automatic characterization of lesions on 68 ga-psma-11 pet/ct, European Journal of Nuclear Medicine and Molecular Imaging 47 (2020) 603–613.

Kai Hu received the B.S. degree in Computer Science and the Ph.D. degree in Computational Mathematics from Xiangtan University, Hunan, China, in 2007 and 2013, respectively. He was a Visiting Scholar at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2016 to 2017. Currently, he is an Associate Professor in the Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education and the School of Computer Science, Xiangtan University, China. His current research interests include machine learning, pattern recognition, bioinformatics, and medical image processing.

Yingjie Huang received the B.S. degree in Communication Engineering from Xiangtan University, Hunan, China, in 2018. Now he is pursuing the M.S. degree in Information and Communication Engineering from Xiangtan University, Hunan, China. His current research interests are deep learning and medical image processing.

Wei Huang received the B.S. degree in Medical Imaging from University of South China, Hunan, China, in 2005, and the M.S. degree in Clinical Medicine from Central South University, Hunan, China, in 2014. Now, he is an Associate Professor in the Radiology Department, the First Hospital of Changsha. His research interests focus on medical imaging, medical image processing, and disease diagnosis.

Hui Tan received the B.S. degree in Information Engineering from Huaqiao University, Fujian, China, in 2019. Now he is pursuing the M.S. degree in Information and Communication Engineering from Xiangtan University, Hunan, China. His current research interests are deep learning and medical image processing.

Zhineng Chen received the M.Sc and B.Sc degrees in computer science from the College of Information Engineering, Xiangtan University, China, in 2004 and 2007, respectively, and the Ph.D. degree in Computer Science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2011. He is now a Pre-tenured Professor with the school of computer science, Fudan University, Shanghai, China. He was an Associate Professor with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, and was a senior research associate with the Department of Computer Science, City University of Hong Kong, Hong Kong, China. He has published over 60 academic papers in prestigious journals and conferences. His research interests include large-scale multimedia analytics, medical image analysis and pattern recognition.

Zheng Zhong received the B.S. and M.S. degrees in medical imaging from the Central South University in 1996 and 2017, respectively. Now, he is a professor in the Department of radiology of the first hospital of Changsha. His research interests are medical image processing and disease diagnosis.

Xuanya Li received the Ph.D. degree from Beijing Institute of Technology, Beijing, China, in 2012. He is currently the director of Baidu Campus and the executive member of China Computer Federation. His main research interests include Internet of Things and artificial intelligence.

Yuan Zhang received the B.S. degree in Biomedical Engineering from Zhengzhou University, Henan, China, in 2009, and the M.S. degree in Information and Communication Engineering from Xiangtan University, Hunan, China, in 2012. She is currently pursuing the Ph. D. degree in Computational Mathematics from Xiangtan University. Her research interests focus on wavelet analysis, machine learning, and biomedical signal processing.

Xieping Gao was born in 1965. He received the B.S. and M.S. degrees from Xiangtan University, China, in 1985 and 1988, respectively, and the Ph.D. degree from Hunan University, China, in 2003. He is a Professor in the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, China. He is also with the Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University, China. He was a visiting scholar at the National Key Laboratory of Intelligent Technology and Systems, Tsinghua University, China, from 1995 to 1996, and at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2002 to 2003. He is a regular reviewer for several journals and he has been a member of the technical committees of several scientific conferences. He has authored and co-authored over 110 journal papers, conference papers, and book chapters. His current research interests are in the areas of wavelet analysis, neural network, bioinformatics, image processing, and computer network.