

Received April 8, 2020, accepted April 24, 2020, date of publication April 29, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991237

# Class Balanced Loss for Image Classification

LIN WANG<sup>1</sup>, CHAOLI WANG<sup>1</sup>, (Member, IEEE), ZHANQUAN SUN<sup>1</sup>,  
SHUQUN CHENG<sup>2</sup>, AND LEI GUO<sup>2</sup>

<sup>1</sup>School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

<sup>2</sup>Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital, Shanghai 200438, China

Corresponding authors: Chaoli Wang (clwang@usst.edu.cn) and Shuqun Cheng (chengshuqun@aliyun.com)

This work was supported in part by the Shanghai Natural Science Foundation under Grant 19ZR1436000, and in part by the National Natural Science Foundation of China under Grant 61374040.

**ABSTRACT** In the study of image classification, neural network learning relies heavily on datasets. Due to variability in the difficulty of collecting images in reality, datasets tend to have class imbalance problems, which undoubtedly increases the difficulty of classification. During the training of a neural network, classes with a large number of images are naturally trained more often than classes with a small number of images. Because of imbalanced training, the classification ability of neural networks on test and validation sets differs greatly in different categories. The test results of more training classes are better, and the test results of classes with less training are poor. In this paper, we propose two kinds of balanced loss functions, namely, CEFL loss and CEFL2 loss, by rebalancing the cross-entropy loss function and focal loss function. The experimental results show that the proposed loss functions are significantly able to improve classification accuracy on class-imbalanced datasets.

**INDEX TERMS** Classification, class imbalance, loss, accuracy.

## I. INTRODUCTION

In the past decade, deep learning methods have been widely used in image classification, target detection, voice recognition and other related fields. Many well-known network structures, such as AlexNet [1], VGG [2], ResNet [3], Inception [4], and Res2Net [5], have been widely applied and studied. Furthermore, breakthroughs are still being made in those fields [6], [7]. Despite these advances, network performance tends to decline significantly in the face of some class-imbalanced data problems [8]. Therefore, research addressing the network performance degradation caused by class-imbalanced data has great importance [9].

Supervised learning requires a large quantity of accurately labelled data. In image classification, every data sample belongs to a known category. However, real-world datasets often face a class imbalance problem [10], [11]. As illustrated in Fig. 1, the Kaggle Mushroom Classification dataset demonstrates an imbalanced distribution of different categories of data. In this dataset, the data samples of some classes occupy a large proportion of the total samples, while those of other classes take up a small proportion. The training process of the neural network will shift towards those classes

with a large number of samples, which is conducive to the learning of those classes with a large number of samples. Correspondingly, classes with a small number of samples are studied less. As a result, classes with a few samples are more likely to be misclassified. In extreme cases, the designed neural network may ignore the minority class [10]. Models trained with class-imbalanced datasets have poor generalization ability, are prone to overfitting and tend to perform well in the training set but poorly in the validation set or test set [9], [12].

When neural networks are used to study the problem of image classification, the loss function describes the relationship between the predicted output of the network  $\hat{y}$  and the true label  $y$ , reflects the fitting degree of the model to the data, and is an important index to evaluate network performance. Common loss functions, including the cross-entropy loss function, perform well when training datasets with a balanced distribution in some datasets, such as CIFAR-10/100 [13] and MNIST [14]. However, when the training datasets exhibit class imbalance phenomena, the cross-entropy loss function often does not classify well [15].

Samples from minor classes tend to have higher loss than those from major classes, as the features learned in minor classes are usually poorer [11], [15]. Focal loss [16], proposed by Kaiming He, can solve the class imbalance problem in

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaochun Cheng.

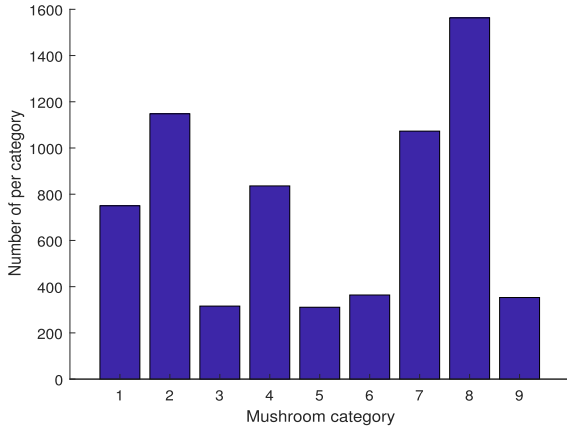


FIGURE 1. Kaggle Mushroom Classification dataset distribution.

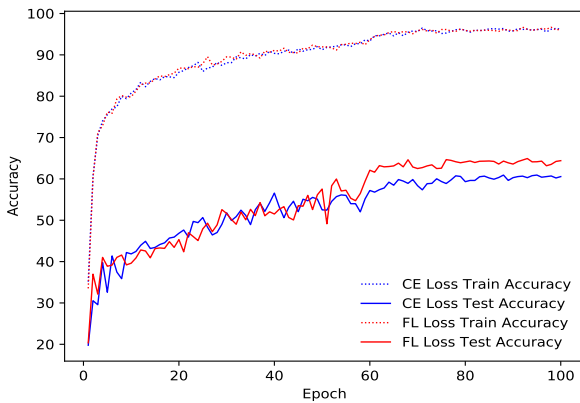


FIGURE 2. Experimental results on the Kaggle Mushroom dataset with ResNet-34 (CE Loss: cross-entropy loss, FL Loss: focal loss).

classification to a certain extent [17]. Fig. 2 shows the performances of both the focal loss and cross-entropy loss functions when dealing with the class imbalance problem. It can be seen from the figure that focal loss has a certain degree of improvement when compared with the cross-entropy loss function.

Focal loss has the following form:

$$FL(p) = -(1 - p)^\gamma \log(p) \quad (1)$$

where  $p$  is the neural network output of ground truth and  $\gamma$  is a constant hyperparameter. When  $\gamma = 0$ , the focal loss is formally equivalent to the cross-entropy loss. Focal loss distinguishes the weights of different samples through  $(1 - p)^\gamma$ . With an increase in network training frequency, classes with a large number of samples in the training set will have a higher ratio of training frequency, and neural networks will naturally learn more from these classes. Therefore, the classification of classes with a large number of samples will have greater accuracy, and samples in such classes will be easier to classify ( $p > 0.5$ , where  $p$  is the predicted probability of network output for the ground truth) [16]. Specifically, focal loss assigns a smaller loss to these classes. In contrast, the accuracy of classes with fewer samples is relatively lower, which makes it more likely that they will

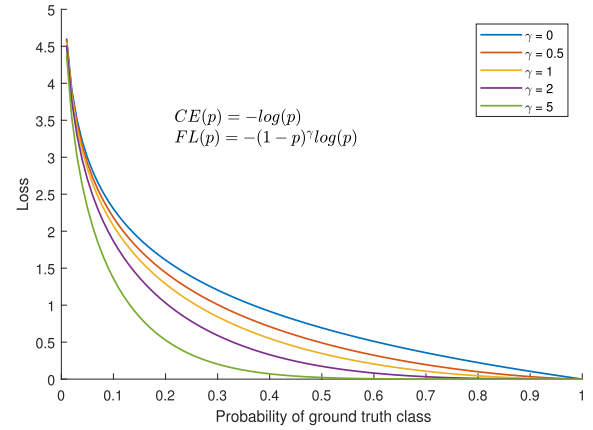
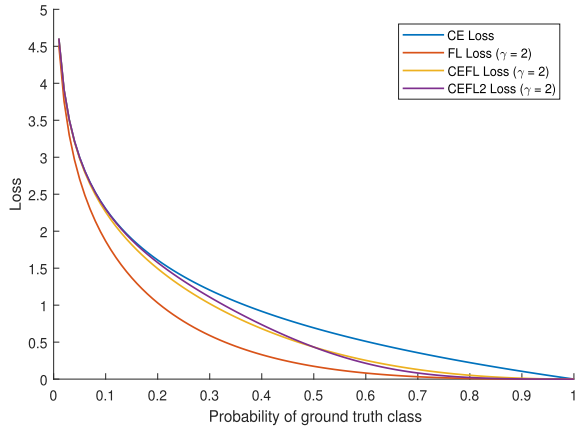


FIGURE 3. Focal loss graphs with different  $\gamma$ .

become poorly classified ( $p \leq 0.5$ ) [16] classes. Focal loss assigns a relatively larger loss to them. Therefore, focal loss focus more on poorly classified samples from minor classes. The focal loss graph is shown in Fig. 3.

Fig. 3 shows an interesting phenomenon: when using the focal loss function in network training (e.g.,  $\gamma = 2$ ), the loss of easy samples ( $p > 0.5$ ) is relatively small; instead, the loss of hard samples ( $p \leq 0.5$ ) becomes relatively large, but compared to that of the cross-entropy loss function, it becomes smaller with the same probability  $p$ . Naturally, a bold idea arises: based on the characteristics of cross-entropy loss with regard to higher loss on well-classified samples, we can add a cross-entropy loss item to focal loss so that the neural network can better distinguish between both easy samples and hard samples. Aiming at solving this problem, we introduce two kinds of loss functions, namely CEFL loss and CEFL2 loss, based on cross-entropy loss and focal loss to further increase the loss for easy samples and hard samples. Renderings of the CEFL loss and CEFL2 loss functions are shown in the following Fig 4, which will be displayed in section III in more detail. The figure shows the proposed CEFL loss function and CEFL2 loss function. Assume  $p$  is the predicted probability of the network for the ground truth. When  $p \leq 0.5$  (the samples are poorly classified), CEFL loss and CEFL2 loss are better approaches to cross-entropy loss, while when  $p > 0.5$  (well-classified samples), CEFL loss and CEFL2 loss approach the focal loss.

Our main contributions can be summarized as follows: (1) Based on the existing cross-entropy loss function and focal loss function, we propose two new balanced loss functions, CEFL loss and CEFL2 loss, which help to improve the image classification accuracy of class-imbalanced datasets. (2) We verify the superiority of CEFL loss and CEFL2 loss on three datasets with imbalanced data, and the experimental results shows that the two loss functions performed similarly on the Imbalanced CIFAR-10/CIFAR-100 dataset, while CEFL loss function was slightly better on Kaggle Mushroom Classification dataset and the performance of CEFL2 loss function was slightly better on AI Challenger 2018 Crop Disease dataset. We believe that, like focal loss, CEFL loss



**FIGURE 4.** Proposed loss curves compared to cross the cross-entropy loss and focal loss curves.

and CEFL2 loss can be extended to the field of target detection to address extreme imbalances between foreground and background classes during training [16], [18]–[20].

## II. RELATED WORK

Recent research methods have been proposed to address class-imbalanced datasets [21]–[25] and can be broadly divided into two categories. One approach resamples the training dataset, and the other reweights the training loss for imbalanced classes.

### A. RESAMPLING

Resampling is the process of over-sampling (adding duplicate samples to) classes with minority samples, under-sampling (deleting samples from) classes with majority samples, and sometimes both. Resampling can rebalance the distribution of training data to some extent, but over-resampling may increase the number of duplicate samples of classes with few samples, which may lead to overfitting of the model. In addition, for some datasets, the numbers of classes with minority samples are extremely limited, perhaps only a few images. In this case, it is very inappropriate to delete images in classes with large samples to balance the dataset. To solve this problem, adjacent samples can be interpolated [26] or synthesized [27], [28] for classes with minority samples. However, these new samples will inevitably include noise, which could reduce the generation of models.

### B. DATA AUGMENTATION

Data augmentation [29] is one of the most widely used methods before data are put into the network. There are many ways to augment data, most of which are available in current deep learning frameworks (such as PyTorch, Caffe and TensorFlow), which users can easily call as needed. To some extent, data augmentation can expand classes with small sample sizes and reduce overfitting slightly. However, the data augmentation method is much less effective for classes with very few data, so it is often used in conjunction with other methods.

### C. HARD EXAMPLE MINING (HEM)

Features of samples learned in minor classes are usually poorer, which may cause samples from minor classes to tend to be hard samples. HEM is a common method to deal with hard samples. The core idea of this method is to select hard samples according to the loss of input samples and put them back into network training. Online hard example mining (OHEM) [30] is a typical hard sample mining method. It aims to screen out the first few samples with greater loss and then rejoin the network for retraining. Like focal loss, OHEM places more emphasis on misclassified examples, but unlike focal loss, OHEM completely discards easy samples.

### D. FOCAL LOSS

In 2017, He et al proposed focal loss. The paper makes it clear that data imbalance raises two problems: (1) training is inefficient, as most locations are easy negatives that contribute no useful learning signal; and (2) en masse, easy negatives can overwhelm training and lead to degenerate models [16]. In general, when training class-imbalanced datasets with neural networks, classes with a large number of samples will receive more training and become easier to classify, while classes with a small number of samples will receive less training and become more difficult to classify. Focal loss is proposed for the express purpose of making a large contribution to the total training loss when there is a small number of difficult samples and a small contribution to the total loss when there is a large number of simple samples. In other words, focal loss causes the network to focus more on hard samples whose training set only includes a few data points in the whole dataset. However, when the predicted probability  $p$  of the ground truth is small, for example,  $p \leq 0.5$ , focal loss decreases the loss value compared to that of cross-entropy loss, which may not contribute to increasing the loss difference between easy examples and hard examples.

### III. CEFL/CEFL2 LOSS

Suppose that the output of a  $j$ -classification model for all classes is a set of tensors  $s = \{s_1, s_2, \dots, s_j\}$  where  $j$  is the total number of classes. After softmax calculation,  $s$  can be mapped to the probability distribution of different classes. Given the model output  $s$  of a set of samples, the probability output after softmax is:

$$p_i = \frac{e^{s_i}}{\sum_j e^{s_j}} \quad (2)$$

where  $i$  is the label of the ground truth and  $p_i$  is the model's estimated probability of the ground truth.

### A. CROSS-ENTROPY LOSS

The cross-entropy loss function is a loss function often used in multi-classification problems, and its form is as follows:

$$CE = -[y \log(p) + (1 - y) \log(1 - p)] \quad (3)$$

where  $p$  is the model's estimated probability for all classes and  $y$  is the set of labels of all classes. In a multi-classification

problem, the label is given in a one-hot coding form. Therefore,  $y$  should be 1 when only considering the ground truth [16]. Thus, the cross-entropy loss can be reduced to:

$$CE = -\log(p) \quad (4)$$

For a batch with an input of  $N$  images, the cross entropy loss is:

$$CE = -\frac{1}{N} \sum_{i=1}^N \log(p) \quad (5)$$

### B. FOCAL LOSS

Focal loss adds a term of  $(1-p)^\gamma$  based on the cross-entropy loss function to reduce the relative loss of well-classified samples and increase the relative loss of poorly classified samples. The formula is shown as follows:

$$FL = -(1-p)^\gamma \log(p) \quad (6)$$

where  $p$  is the model's estimated probability for the ground truth and  $\gamma$  ( $\gamma > 0$ ) is a tuneable focusing parameter. The focal loss function is drawn according to different gamma values shown in Fig. 3. It is worth mentioning that when  $\gamma = 0$ , focal loss is formally the same as the cross-entropy loss function. It can be seen from Fig. 3 that with increasing  $\gamma$ , the value of focal loss generally declines compared to that of cross-entropy loss. In [16], it is mentioned that the author introduces the  $\alpha$ -balanced variable to improve accuracy. In this case, the  $\alpha$ -balanced focal loss is denoted as:

$$FL = -\alpha(1-p)^\gamma \log(p) \quad (7)$$

For a batch with input of  $N$  images, the focal loss should be:

$$FL = -\frac{1}{N} \sum_{i=1}^N \alpha(1-p)^\gamma \log(p) \quad (8)$$

### C. CEFL/ZZZZZ/WWWWW/CEFL2 LOSS

Samples from minor classes tend to have higher losses than those from major classes as the features learned in minor classes are usually poorer. Consequently, samples from major classes will be well-classified and the samples from minor classes will be poorly classified [15], [16]. Aiming at addressing the problem, cost-sensitive learning method [31] assigns loss function weight for each class according to a given data distribution. One of interesting problems is how to assign weight distribution for major and minor classes automatically.

Motivated by this idea of weight distribution, we proposed two loss functions called CEFL loss and CEFL2 loss to address the problem of training from imbalanced data samples by rebalancing cross-entropy loss and focal loss to expand the loss value of well-classified and poorly classified examples. Mathematically, we distributed the weights of  $(1-p)$  and  $p$  to the cross-entropy loss function and focal loss

function respectively. A new loss function, called CEFL loss, is chosen as following form:

$$CEFL = -(1-p)\log(p) - p(1-p)^\gamma \log(p) \quad (9)$$

where  $\gamma$  ( $\gamma > 0$ ) is a tuneable focusing parameter, as it is in focal loss. Note that when  $\gamma = 0$ , CEFL loss is formally cross-entropy loss.

When a sample is poorly classified ( $p \leq 0.5$ ),  $p$  is relatively smaller and  $(1-p)$  is relatively larger, which makes the weight of the cross-entropy loss function relatively larger while the weight of the focal loss function is relatively smaller; thus, CEFL loss is closer to cross-entropy loss. Conversely, when a sample is well classified,  $p$  is relatively larger and  $(1-p)$  is relatively smaller, which gives a smaller weight to the cross entropy loss function and a larger weight to the focal loss function; thus, CEFL loss is closer to focal loss. The polarized trend of well-classified and poorly classified samples is conducive to make minor classes in smaller losses and major classes in larger losses as is shown in Fig. 4. Therefore, the network will focus on poorly classified classes.

For a batch with an input of  $N$  images, the CEFL loss should be:

$$CEFL = -\frac{1}{N} \sum_{i=1}^N [(1-p)\log(p) + p(1-p)^\gamma \log(p)] \quad (10)$$

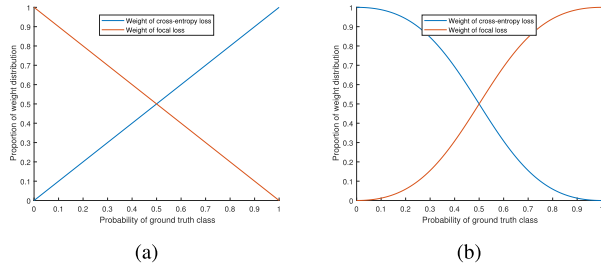
CEFL loss has two properties. (1) For well-classified samples from major classes, CEFL loss assigns them relatively smaller loss values by approaching focal loss. (2) For poorly classified samples from minor classes, CEFL loss assigns them relatively larger loss values by approaching cross-entropy loss. In order further to enhance the two properties, CEFL2 loss is introduced. CEFL2 loss is similar to CEFL loss in form, except that the weight distribution of the cross-entropy loss term and focal loss term in CEFL loss are different, which expands the effective loss of easy samples and hard samples to some extent, as shown in Fig. 4. The CEFL2 loss function is as follows:

$$CEFL2 = -\frac{(1-p)^2}{(1-p)^2 + p^2} \log(p) - \frac{p^2}{(1-p)^2 + p^2} (1-p)^\gamma \log(p) \quad (11)$$

For a batch with an input of  $N$  images, the CEFL2 loss should be:

$$CEFL2 = -\frac{1}{N} \sum_{i=1}^N \left[ \frac{(1-p)^2}{(1-p)^2 + p^2} \log(p) + \frac{p^2}{(1-p)^2 + p^2} (1-p)^\gamma \log(p) \right] \quad (12)$$

The proposed CEFL loss and CEFL2 loss are able to distinguish between well-classified and poorly classified samples according to different predicted class probability  $p$ . Fig. 5 demonstrates the polarized degree of cross-entropy loss and focal loss with the change of probability for ground truth class.



**FIGURE 5. Weight distributions of cross-entropy loss and focal loss. Left: CEFL loss. Right: CEFL2 loss.**

#### IV. EXPERIMENTS

To verify the effectiveness of the two loss functions, CEFL loss and CEFL2 loss, proposed in this paper, three different datasets were used, including the manually processed imbalanced CIFAR-10/CIFAR-100 (which we call the imbalanced CIFAR-10/CIFAR-100 dataset), the AI Challenger 2018 Crop Disease Detection dataset and the Kaggle Mushroom Classification dataset. In addition, all three datasets have different degrees of class imbalance problems; thus, they can be used to measure the loss function that we propose. Furthermore, all three datasets are available on the Internet.

ResNet is one of the most widely applied network architectures. The ResNet network introduced a shortcut structure and solved the problem of gradient disappearance to some extent with a deepening network layer. Its generalization ability in ImageNet, CIFAR, PASCAL and other datasets exhibited good performance [3]. Therefore, we use deep residual networks (ResNet), both pretrained and non-pretrained, with different depths as our network framework.

When analyzing the experimental results, we used several different metrics, including global *accuracy*, *precision*, *recall*, and *F1\_score*. Accuracy measures the classifier performance across all classes, while others are specific to each single class. They are written as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

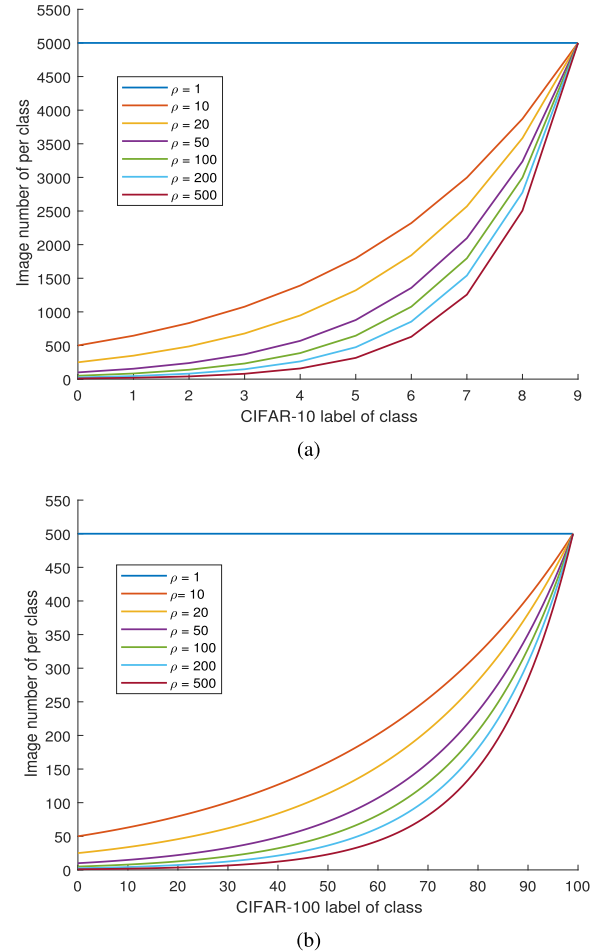
$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1\_score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (16)$$

#### A. DATASET INTRODUCTION

##### 1) IMBALANCED CIFAR-10/CIFAR-100 DATASET

The original CIFAR-10/CIFAR-100 contains a total of 60,000 images (50,000 training images and 10,000 test images). According to different subdivisions, the CIFAR datasets are divided into two categories, CIFAR-10 (the same number of sets in each category, with 5000 images) and CIFAR-100 (the same number of sets in each category, with 500 images). We conducted a certain amount of random sampling on the original CIFAR-10/CIFAR-100



**FIGURE 6. The imbalanced CIFAR-10/CIFAR-100 dataset distribution with different  $\rho$  values. Top: Imbalanced CIFAR-10. Bottom: Imbalanced CIFAR-100.**

training set for each class to obtain a class-imbalanced training set. To measure the imbalance of the new imbalanced CIFAR-10/CIFAR-100 training set, we adopt the imbalance ratio [9] of the numbers of images between the class with the most images and the class with the fewest images. The formula is expressed as

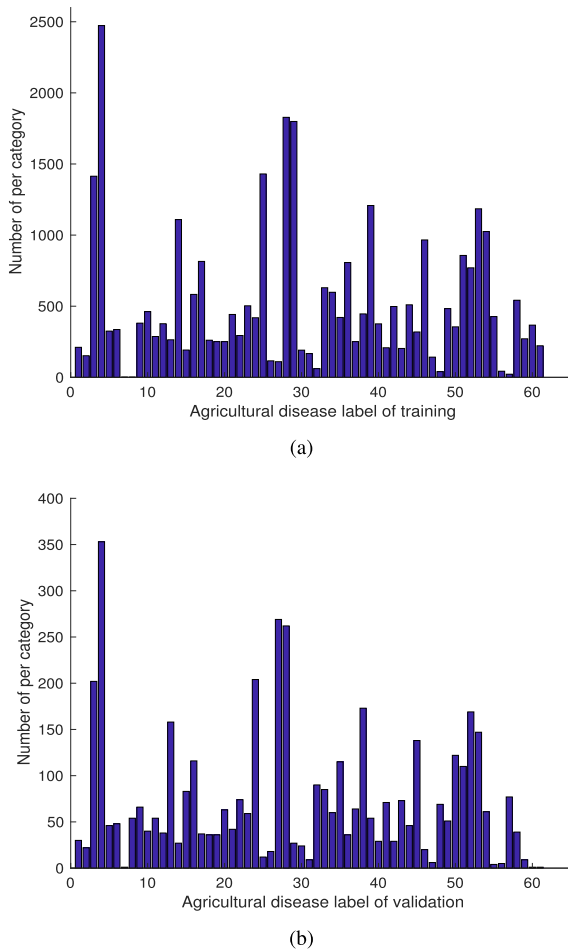
$$\rho = \frac{\max_i \{ |C_i| \}}{\min_i \{ |C_i| \}} \quad (17)$$

where  $C_i$  is a set of examples in class  $i$ ,  $\max_i \{ |C_i| \}$  is the number of images in the class with the most images in the newly defined dataset, and  $\min_i \{ |C_i| \}$  is the number of images in the class with the fewest images in the newly defined dataset. The new dataset samples different numbers of images from each class in the original CIFAR-10/CIFAR-100. To evaluate the imbalance degree of the dataset, a set of  $\rho$  values ( $\rho \in \{1, 10, 20, 50, 100, 200, 500\}$ ) is used in our experiments. The samples of imbalanced CIFAR-10 and CIFAR-100 obey the following power functions:

$$N_i = 5000m^{\frac{i}{9}-1} \quad (18)$$

$$N_j = 500m^{\frac{j}{99}-1} \quad (19)$$



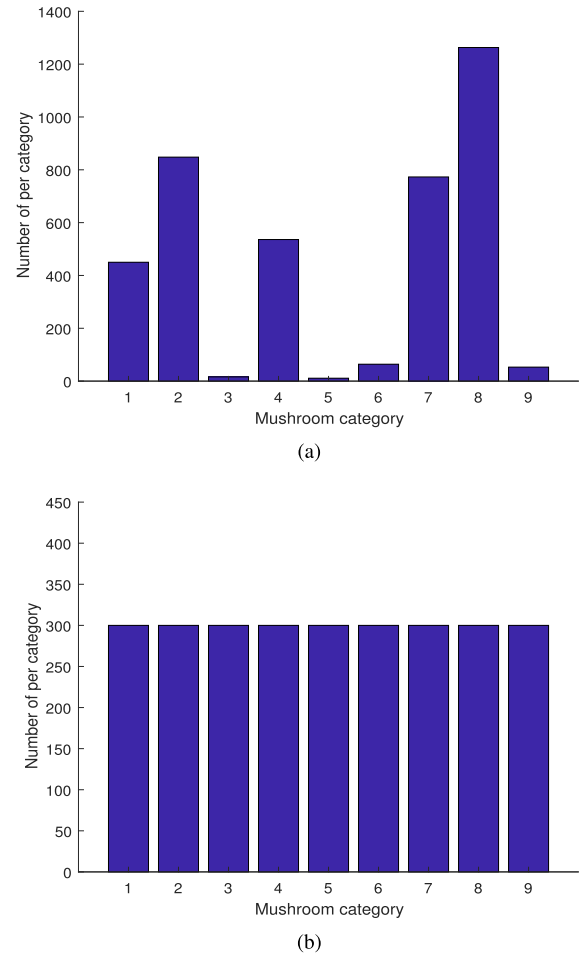


**FIGURE 7.** The imbalanced data statistics of training and validation for the AI Challenger 2018 Crop Disease Detection dataset. Top: training set. Bottom: validation set.

where  $i(i \in \{0, 1, \dots, 9\})$  and  $j(j \in \{0, 1, \dots, 99\})$  are the indexes of the class labels in CIFAR-10 and CIFAR-100, respectively. Therefore,  $i, j \in N^*$  and  $i \in [0, 9]$ ,  $j \in [0, 99]$ .  $N_i$  and  $N_j$  are the numbers of images for each class of the newly generated imbalanced CIFAR-10/CIFAR-100. For example, sampling from CIFAR-10 with the imbalance factor  $\rho = 10$ , the image numbers of each class in imbalanced CIFAR-10 should be  $N = [500, 645, 834, 1077, 1391, 1796, 2320, 2997, 3871, 5000]$ . Moreover, the test set is the same as the original one. We conducted experiments on imbalanced CIFAR-10/CIFAR-100. By sampling the original CIFAR-10/CIFAR-100 according to different imbalance factors, we created several different datasets, which are shown in Fig. 6.

## 2) AI CHALLENGER 2018 CROP DISEASE DATASET

The AI Challenger 2018 Crop Disease Detection dataset is the official open data set of the AI Challenger 2018 competition. The dataset is classified into 61 categories according to 'species - disease - degree'. There are 31,718 pictures in the training set and 4540 pictures in the validation set. We conducted experiments on training sets and validation



**FIGURE 8.** The imbalanced data statistics of training and testing for the Kaggle Mushroom dataset. Top: training set. Bottom: test set.

sets. The following figure can be obtained by conducting sample statistics for each class of the training set and validation set. As seen from Fig. 7, the total number of samples in categories with a large quantity of data can reach approximately 2500, while categories with a small quantity of data have only a few images.

## 3) KAGGLE MUSHROOM CLASSIFICATION DATASET

The Kaggle Mushroom Classification dataset is a publicly available dataset on Kaggle's official website. The dataset contains the most common images of the northern European mushroom species. There are 5614 images divided into 9 folders in the dataset. Each folder contains 300 to 1500 or more images of the mushroom genus, which makes it a typical class-imbalanced dataset.

We randomly selected 300 images from each category as the test set and the rest as the training set. Therefore, the data distribution of the training set and test set is shown in Fig. 8.

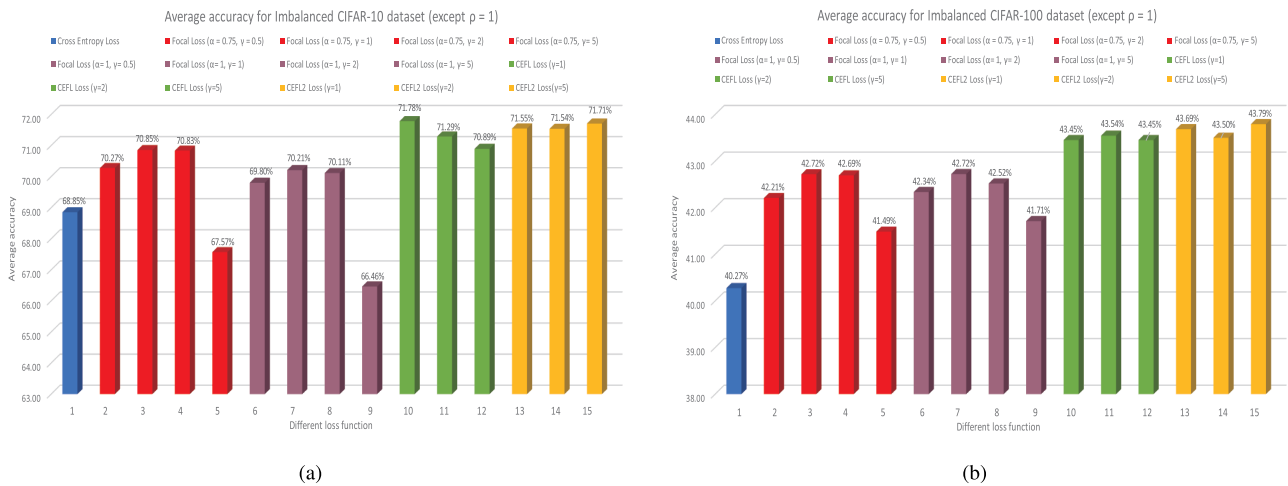
## B. IMPLEMENTATION

All experimental models are based on PyTorch. We used a stochastic gradient descent (SGD) optimizer with a

**TABLE 1.** Data augmentation implemented for different datasets.

Data Augmentation	Imbalanced CIFAR-10/ CIFAR-100	AI Challenger 2018 Crop Disease Detection Dataset	Kaggle Mushroom Classification Dataset
Resize	y	y	y
Random Resized Crop		y	y
Random Crop	y		
Random Horizontal Flip	y	y	y
Colour Jitter		y	y
Random Rotation		y	y
To Tensor	y	y	y
Normalize	y	y	y

y represents the specific augmentation strategy used.

**FIGURE 9.** Average accuracy on imbalanced CIFAR-10/CIFAR-100 with different imbalance ratios when using different loss functions. Left: Imbalanced CIFAR-10. Right: Imbalanced CIFAR-100.

momentum of 0.9 and weight decay of  $1 \times 10^{-4}$ . The data augmentation parts of each training set are shown in Table 1. None of the experiments uses a warm-up strategy [10] in the training.

For the imbalanced CIFAR-10/CIFAR-100 dataset, we used a ResNet-18 network without pretraining, and the size of the input image was  $32 \times 32$  pixels. A total batch size of 128 on a single GPU for 100 epochs was used, and the initial learning rate was 0.05. Then, when training to 60 epochs and 80 epochs, the learning rate was multiplied by 0.1. A number of ablation experiments were conducted on the imbalanced CIFAR-10/CIFAR-100.

For the AI Challenger 2018 Crop Disease Detection dataset, we used ResNet-18, ResNet-34, ResNet-50 and ResNet-101 for the pretraining model, and the image input network size was  $224 \times 224$  pixels. For each pretraining network, we used a batch size of 64 on a single GPU for 60 epochs, initialized the learning rate to 0.01, and then multiplied the learning rate by 0.1 when training to 30 epochs and 50 epochs.

For the Kaggle Mushroom Classification dataset, we used a pretraining model of ResNet-18, ResNet-34 ResNet-50 and ResNet-101 with an input image size of  $224 \times 224$  pixels. For each pretraining network, a batch size of 64 on a single GPU for 100 epochs was used, and the initial learning rate

was 0.01. Then, when training to 60 epochs and 80 epochs, the learning rate was multiplied by 0.1.

### C. EXPERIMENTAL RESULTS

To demonstrate that our proposed loss functions can be used in real-world imbalanced datasets, we conducted experiments on the above mentioned three datasets.

#### 1) EXPERIMENTS ON IMBALANCED CIFAR-10/CIFAR-100 DATASET

We conducted a series of ablation experiments on the imbalanced CIFAR-10/CIFAR-100 datasets based on ResNet-18, and the results show that the proposed CEFL loss and CEFL2 loss can improve the accuracy. We show the results in Tables 2 and 3. And the average accuracy averages the results of  $\rho = 10, 20, 50, 100, 200, 500$ . Obviously, the best accuracy results in each dataset with different  $\rho$  values and average accuracy are for the proposed CEFL and CEFL2 loss.

Fig. 9 shows the average accuracy of the imbalanced CIFAR-10/CIFAR-100 dataset containing various imbalanced ratios  $\rho$  (except  $\rho = 1$ ) listed in the last column of Tables 2 and 3. In Fig. 9, the proposed CEFL loss (green cylinder) and CEFL2 loss (yellow cylinder) significantly help the network improve classification accuracy compared to that

**TABLE 2.** Classification accuracy rate on the imbalanced CIFAR-10 dataset with different loss functions.

Imbalance Ratio ( $\rho$ )	Imbalanced CIFAR-10 Test Accuracy							Average Accuracy (except $\rho = 1$ )
	1	10	20	50	100	200	500	
Cross-Entropy Loss	93.43%	84.84%	80.74%	72.96%	65.59%	59.27%	49.71%	68.85%
Focal Loss ( $\alpha = 0.75, \gamma = 0.5$ )	93.27%	85.15%	81.01%	73.98%	67.25%	61.43%	52.82%	70.27%
Focal Loss ( $\alpha = 0.75, \gamma = 1$ )	93.35%	85.47%	81.06%	75.18%	67.70%	62.61%	53.08%	70.85%
Focal Loss ( $\alpha = 0.75, \gamma = 2$ )	93.53%	85.35%	81.50%	75.25%	67.46%	62.11%	53.33%	70.83%
Focal Loss ( $\alpha = 0.75, \gamma = 5$ )	91.85%	83.50%	79.49%	71.95%	63.42%	56.67%	50.37%	67.57%
Focal Loss ( $\alpha = 1, \gamma = 0.5$ )	93.36%	85.22%	80.96%	73.01%	66.44%	60.15%	53.02%	69.80%
Focal Loss ( $\alpha = 1, \gamma = 1$ )	93.74%	85.60%	81.29%	72.81%	67.73%	60.24%	53.58%	70.21%
Focal Loss ( $\alpha = 1, \gamma = 2$ )	93.47%	85.14%	81.26%	74.48%	65.61%	62.10%	52.09%	70.11%
Focal Loss ( $\alpha = 1, \gamma = 5$ )	91.75%	83.32%	78.73%	72.79%	61.06%	54.60%	48.27%	66.46%
CEFL Loss ( $\gamma = 1$ )	93.71%	85.77%	82.04%	75.29%	70.51%	62.95%	54.14%	71.78%
CEFL Loss ( $\gamma = 2$ )	93.54%	85.73%	82.38%	75.33%	67.42%	62.46%	54.41%	71.29%
CEFL Loss ( $\gamma = 5$ )	93.42%	85.02%	81.50%	76.19%	68.17%	60.66%	53.81%	70.89%
CEFL2 Loss ( $\gamma = 1$ )	93.81%	85.52%	82.46%	75.89%	69.22%	62.10%	54.09%	71.55%
CEFL2 Loss ( $\gamma = 2$ )	93.43%	85.51%	81.58%	75.40%	69.15%	63.34%	54.24%	71.54%
CEFL2 Loss ( $\gamma = 5$ )	93.79%	86.15%	81.91%	75.42%	69.19%	62.43%	55.14%	71.71%

**TABLE 3.** Classification accuracy rate on the imbalanced CIFAR-100 dataset with different loss functions.

Imbalance Ratio ( $\rho$ )	Imbalanced CIFAR-100 Test Accuracy							Average Accuracy (except $\rho = 1$ )
	1	10	20	50	100	200	500	
Cross-Entropy Loss	72.72%	55.68%	49.88%	40.41%	36.53%	32.49%	26.65%	40.27%
Focal Loss ( $\alpha = 0.75, \gamma = 0.5$ )	73.19%	57.19%	51.73%	42.45%	38.66%	33.98%	29.23%	42.21%
Focal Loss ( $\alpha = 0.75, \gamma = 1$ )	73.58%	57.56%	52.10%	43.35%	38.96%	34.72%	29.61%	42.72%
Focal Loss ( $\alpha = 0.75, \gamma = 2$ )	73.44%	57.92%	51.25%	43.73%	38.81%	34.14%	30.30%	42.69%
Focal Loss ( $\alpha = 0.75, \gamma = 5$ )	72.03%	55.84%	51.59%	41.63%	37.92%	33.41%	28.52%	41.49%
Focal Loss ( $\alpha = 1, \gamma = 0.5$ )	73.25%	57.28%	51.86%	42.53%	38.76%	34.03%	29.55%	42.34%
Focal Loss ( $\alpha = 1, \gamma = 1$ )	73.49%	57.77%	51.49%	43.94%	39.06%	34.34%	29.72%	42.72%
Focal Loss ( $\alpha = 1, \gamma = 2$ )	73.63%	57.30%	51.49%	43.11%	38.56%	34.28%	30.37%	42.52%
Focal Loss ( $\alpha = 1, \gamma = 5$ )	72.49%	57.05%	50.43%	43.08%	37.41%	33.46%	28.84%	41.71%
CEFL Loss ( $\gamma = 1$ )	74.25%	57.52%	53.01%	44.60%	39.51%	35.08%	30.98%	43.45%
CEFL Loss ( $\gamma = 2$ )	73.91%	58.35%	52.63%	44.76%	39.47%	35.86%	30.19%	43.54%
CEFL Loss ( $\gamma = 5$ )	74.23%	58.06%	52.55%	44.18%	39.97%	34.81%	31.13%	43.45%
CEFL2 Loss ( $\gamma = 1$ )	73.73%	58.10%	52.87%	44.49%	39.54%	35.95%	31.16%	43.69%
CEFL2 Loss ( $\gamma = 2$ )	74.16%	58.85%	52.08%	44.60%	39.63%	35.21%	30.63%	43.50%
CEFL2 Loss ( $\gamma = 5$ )	73.89%	58.74%	52.30%	44.57%	39.62%	35.69%	31.83%	43.79%

with cross-entropy loss (blue cylinder) and focal loss (red cylinder).

## 2) EXPERIMENTS ON THE IMBALANCED AI CHALLENGER 2018 CROP DISEASE DETECTION DATASET

For the AI Challenger 2018 Crop Disease Detection dataset, we conducted a set of ablation experiments, and the results are shown in Table 4 and Fig. 10. Table 4 shows that CEFL2 loss function is slightly better than CEFL loss function on AI Challenger 2018 Crop Disease Detection dataset. And Fig. 10

shows the experimental results with different loss functions and different depths of network. It can be seen that the proposed CEFL loss and CEFL2 loss obtain better performance than cross-entropy loss and focal loss on this imbalanced dataset.

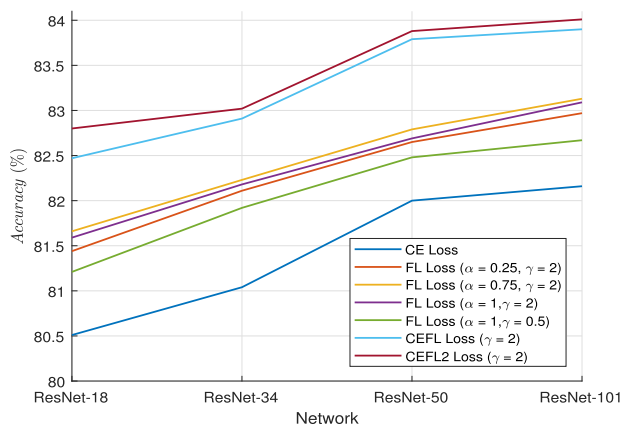
## 3) EXPERIMENTS ON THE KAGGLE MUSHROOM CLASSIFICATION DATASET

For the Kaggle Mushroom Classification dataset, we trained images on ResNet-18, ResNet-34, ResNet-50 and

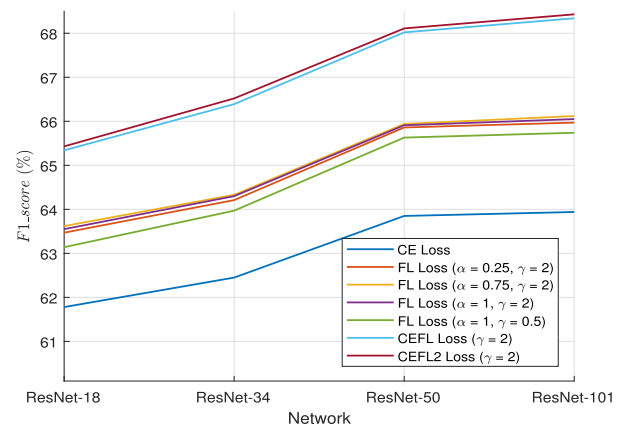


**TABLE 4.** Prediction accuracy of the AI Challenger 2018 Crop Disease Detection dataset and Kaggle Mushroom Classification dataset.

Network	Loss	$\alpha$	$\gamma$	AI Challenger 2018 Crop Disease Detection		Kaggle Mushroom Classification	
				Validation Set		Test Set	
				$F1\_score$	$Accuracy$	$F1\_score$	$Accuracy$
ResNet-18	Cross-Entropy Loss	*	*	61.78%	80.51%	55.67%	58.29%
	Focal Loss	0.25	2	63.47%	81.44%	59.98%	62.45%
	Focal Loss	0.75	2	63.62%	81.66%	60.24%	62.67%
	Focal Loss	1	2	63.55%	81.59%	60.04%	62.53%
	Focal Loss	1	0.5	63.14%	81.21%	59.35%	62.19%
	CEFL Loss	*	2	<b>65.34%</b>	<b>82.47%</b>	<b>63.30%</b>	<b>64.88%</b>
	CEFL2 Loss	*	2	<b>65.43%</b>	<b>82.80%</b>	<b>63.15%</b>	<b>64.67%</b>
ResNet-34	Cross-Entropy Loss	*	*	62.45%	81.04%	57.16%	60.56%
	Focal Loss	0.25	2	64.21%	82.11%	61.16%	64.07%
	Focal Loss	0.75	2	64.33%	82.23%	61.33%	64.29%
	Focal Loss	1	2	64.30%	82.18%	61.27%	64.21%
	Focal Loss	1	0.5	63.97%	81.92%	60.60%	63.65%
	CEFL Loss	*	2	<b>66.39%</b>	<b>82.91%</b>	<b>64.22%</b>	<b>66.23%</b>
	CEFL2 Loss	*	2	<b>66.52%</b>	<b>83.02%</b>	<b>64.14%</b>	<b>65.88%</b>
ResNet-50	Cross-Entropy Loss	*	*	63.85%	82.00%	58.23%	61.75%
	Focal Loss	0.25	2	65.86%	82.65%	61.87%	64.29%
	Focal Loss	0.75	2	65.94%	82.73%	61.99%	64.55%
	Focal Loss	1	2	65.91%	82.69%	61.91%	64.47%
	Focal Loss	1	0.5	65.63%	82.48%	61.24%	63.98%
	CEFL Loss	*	2	<b>68.02%</b>	<b>83.79%</b>	<b>64.77%</b>	<b>66.78%</b>
	CEFL2 Loss	*	2	<b>68.11%</b>	<b>83.88%</b>	<b>64.59%</b>	<b>66.47%</b>
ResNet-101	Cross-Entropy Loss	*	*	63.94%	82.16%	59.17%	62.63%
	Focal Loss	0.25	2	65.97%	82.93%	62.23%	65.19%
	Focal Loss	0.75	2	66.12%	83.13%	62.67%	65.36%
	Focal Loss	1	2	66.05%	83.11%	62.54%	65.26%
	Focal Loss	1	0.5	65.74%	82.67%	61.88%	64.77%
	CEFL Loss	*	2	<b>68.34%</b>	<b>83.90%</b>	<b>65.46%</b>	<b>67.34%</b>
	CEFL2 Loss	*	2	<b>68.43%</b>	<b>84.01%</b>	<b>65.32%</b>	<b>67.02%</b>

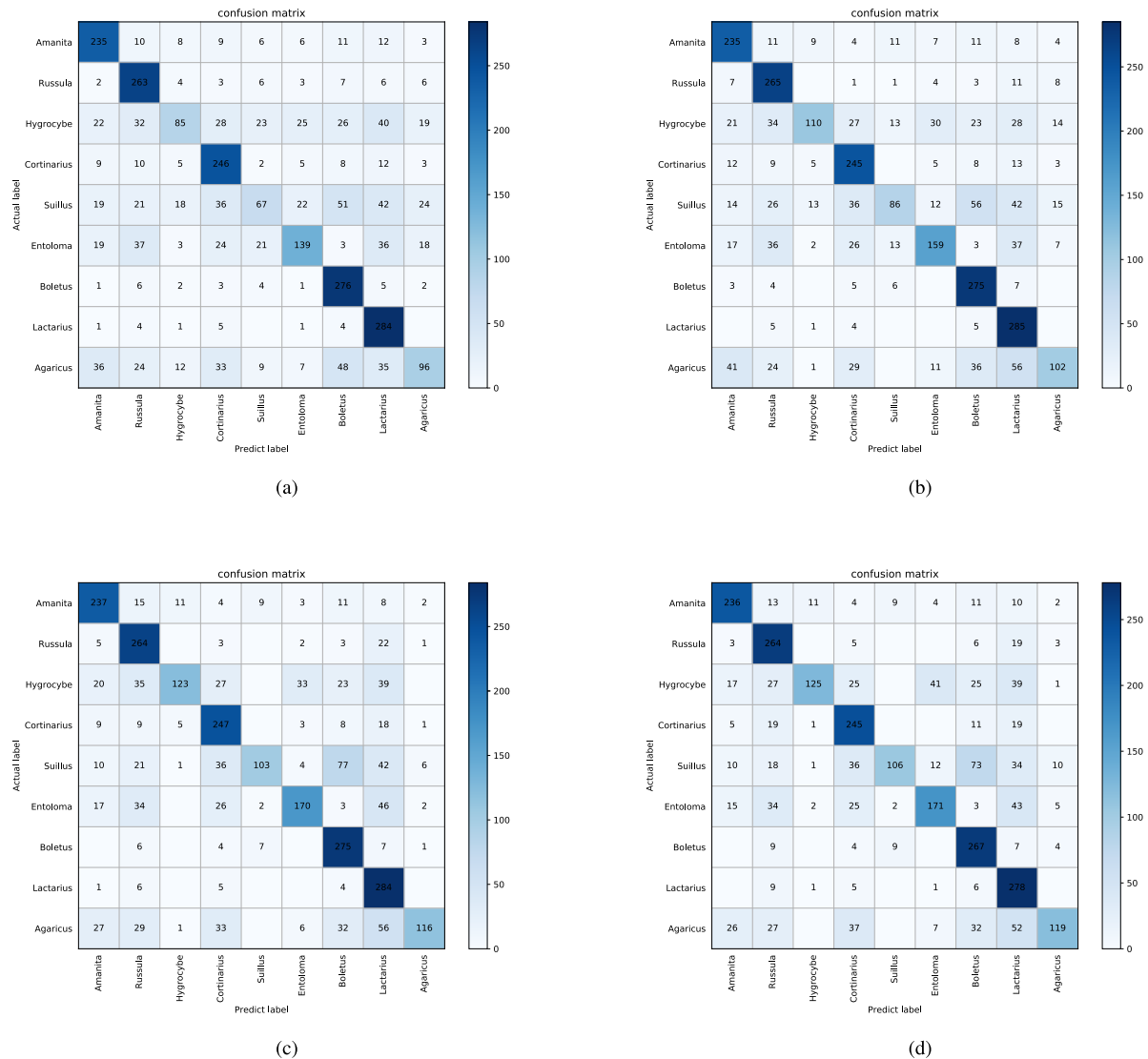


(a)



(b)

**FIGURE 10.** Results of AI Challenger 2018 Crop Disease Detection validation set with different depth of network. Left: Accuracy result. Right:  $F1\_score$  result.

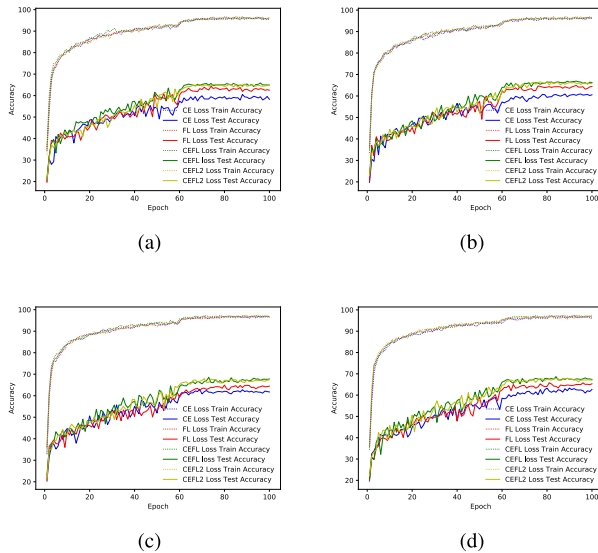


**FIGURE 11.** Confusion matrixes of the Kaggle Mushroom Classification test set training with different loss functions. Upper left: cross-entropy loss. Upper right: focal loss. Lower left: CEFL loss. Lower right: CEFL2 loss.

**TABLE 5.** Classification performance on Kaggle Mushroom test set.

Class	Cross-Entropy Loss			Focal Loss ( $\alpha = 1, \gamma = 2$ )			CEFL Loss ( $\gamma = 2$ )			CEFL2 Loss ( $\gamma = 2$ )		
	Precision	Recall	F1_score	Precision	Recall	F1_score	Precision	Recall	F1_score	Precision	Recall	F1_score
Amanita	68.31%	78.33%	72.98%	67.14%	78.33%	72.30%	72.70%	79.00%	75.72%	75.64%	78.67%	77.13%
Russula	64.62%	87.67%	74.40%	64.01%	88.33%	74.23%	63.01%	88.00%	73.44%	62.86%	88.00%	73.34%
Hygrocybe*	61.59%	28.33%	38.81%	78.01%	36.67%	49.89%	<b>87.23%</b>	<b>41.00%</b>	<b>55.78%</b>	<b>88.65%</b>	<b>41.67%</b>	<b>56.69%</b>
Cortinarius	63.57%	82.00%	71.62%	64.99%	81.67%	72.38%	64.16%	82.33%	72.12%	63.47%	81.67%	71.43%
Suillus*	48.55%	22.33%	30.59%	66.15%	28.67%	40.00%	<b>85.12%</b>	<b>34.33%</b>	<b>48.93%</b>	<b>84.13%</b>	<b>35.33%</b>	<b>49.76%</b>
Entoloma*	66.51%	46.33%	54.62%	69.74%	53.00%	60.23%	<b>76.92%</b>	<b>56.67%</b>	<b>65.26%</b>	<b>72.46%</b>	<b>57.00%</b>	<b>63.81%</b>
Boletus	63.59%	92.00%	75.20%	65.48%	91.67%	76.39%	63.07%	91.67%	74.73%	61.52%	89.00%	72.75%
Lactarius	60.17%	94.67%	73.58%	58.52%	95.00%	72.43%	54.41%	94.67%	69.10%	55.49%	92.67%	69.41%
Agaricus*	56.14%	32.00%	40.76%	66.67%	34.00%	45.03%	<b>89.92%</b>	<b>38.67%</b>	<b>54.08%</b>	<b>82.64%</b>	<b>39.67%</b>	<b>53.61%</b>
Average	61.45%	62.63%	59.17%	66.75%	65.26%	62.54%	<b>72.95%</b>	<b>67.37%</b>	<b>65.46%</b>	<b>71.87%</b>	<b>67.08%</b>	<b>65.32%</b>

\* represents minor class



**FIGURE 12.** Accuracy of the Kaggle Mushroom Classification dataset trained with ResNet at different depths. Upper left: ResNet-18. Upper right: ResNet-34. Lower left: ResNet-50. Lower right: ResNet-101.

ResNet-101. Fig. 11 shows the confusion matrices for the 9 class of Kaggle Mushroom test set when training with ResNet-101.

Table 4 shows *F1\_score* and *accuracy* results of different network depths. It can be seen that CEFL loss function is slightly better than CEFL2 loss function on Kaggle Mushroom Classification dataset. Table 5 presents the classification results with *precision*, *recall* and *F1\_score* for each class. It can be judged that the proposed loss functions generally improve the classification performance by enhancing classification ability of minor classes.

Fig. 12 shows the accuracy curves both in training and test process. It can be concluded that the proposed CEFL loss and CEFL2 loss can improve accuracy on test set considering different network depths.

## V. CONCLUSION

In this paper, we present two theoretical loss functions to address the problem of class imbalance for training neural networks. The idea of the proposed loss functions comes from considering the cross-entropy loss function and focal loss function. Some important considerations are as follows: (1) When the training samples are well classified ( $p > 0.5$ ), the loss approaches the focal loss. (2) When the training samples are poorly classified ( $p \leq 0.5$ ), the loss is close to the cross-entropy loss. If the proposed loss function satisfies both considerations, it will cause the network to learn more from classes with fewer samples. Thus, the trained network will improve the total accuracy of classification. We experiment with cross-entropy loss, focal loss, and the proposed CEFL loss and CEFL2 loss on three different datasets with class imbalance problems, and the results show that the proposed CEFL loss and CEFL2 loss have better classification accuracy than the other two kinds of loss functions.

In addition, the proposed CEFL loss and CEFL2 loss do not add any new hyperparameters and can be regarded as an automatically balanced loss function between the cross-entropy loss function and focal loss function. This strategy allows our loss function to be broadly applied to models of image classification. We believe that our approach can also be extended to target detection to address the imbalance problem of foreground and background in a similar way to that of focal loss, which we will research further in the future.

## REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [5] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," 2019, *arXiv:1904.01169*. [Online]. Available: <http://arxiv.org/abs/1904.01169>
- [6] T. Akilan, Q. J. Wu, and H. Zhang, "Effect of fusing features from multiple DCNN architectures in image classification," *IET Image Process.*, vol. 12, no. 7, pp. 1102–1110, Jul. 2018.
- [7] Q. Shi, Y. Zhang, X. Liu, and K. Zhao, "Regularised transfer learning for hyperspectral image classification," *IET Comput. Vis.*, vol. 13, no. 2, pp. 188–193, Mar. 2019.
- [8] Y. Zhang, L. Shuai, Y. Ren, and H. Chen, "Image classification with category centers in class imbalance situation," in *Proc. 33rd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, May 2018, pp. 359–363.
- [9] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018.
- [10] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 558–567.
- [11] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
- [12] J. C. Pichel and B. Pateiro-Lopez, "Sparse matrix classification on imbalanced datasets using convolutional neural networks," *IEEE Access*, vol. 7, pp. 82377–82389, 2019.
- [13] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," M.S. thesis, Dept. Comput. Sci., Citeseer, Univ. Toronto, Toronto, ON, Canada, 2009.
- [14] B. Y. LeCun and C. Cortes. *The Mnist Database of Handwritten Digits*. Accessed: Jul. 12, 2026. [Online]. Available: <http://yann.lecun.com/exdb/mnist/>
- [15] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9268–9277.
- [16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.
- [17] M. M. Al Rahhal, Y. Bazi, H. Almubarak, N. Alajlan, and M. Al Zuair, "Dense convolutional networks with focal loss and image generation for electrocardiogram classification," *IEEE Access*, vol. 7, pp. 182225–182237, 2019.
- [18] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
- [19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Adv. neural Inf. Process. Syst.*, 2015, pp. 91–99.

- [20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [21] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5375–5384.
- [22] C. Drummond and R. C. Holte, "C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling," in *Proc. ICML Workshop Learn. Imbalanced Datasets*, vol. 11, 2003, pp. 1–8.
- [23] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3573–3587, Aug. 2018.
- [24] K. M. Ting, "A comparative study of cost-sensitive boosting algorithms," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 983–990.
- [25] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, Jan. 2006.
- [26] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
- [27] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1322–1328.
- [28] Y. Zou, Z. Yu, B. V. Kumar, and J. Wang, "Unsupervised domain adaptation for semantic segmentation via class-balanced self-training," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 289–305.
- [29] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation policies from data," 2018, *arXiv:1805.09501*. [Online]. Available: <http://arxiv.org/abs/1805.09501>
- [30] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 761–769.
- [31] C. Elkan, "The foundations of cost-sensitive learning," in *Proc. 17th Int. Conf. Artif. Intell.*, Seattle, WA, USA, Aug. 2001, pp. 973–978.



**LIN WANG** received the B.S. degree from Shanghai DianJi University, Shanghai, China, in 2018. He is currently pursuing the M.E. degree in control science and engineering with the University of Shanghai for Science and Technology, Shanghai. His current research interests include computer vision, image classification, and object detection.



**CHAOLI WANG** (Member, IEEE) received the B.S. and M.Sc. degrees from the Mathematics Department, Lanzhou University, Lanzhou, China, in 1986 and 1992, respectively, and the Ph.D. degree in control theory and engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1999. From 1999 to 2000, he was a Postdoctoral Research Fellow with the Robotics Laboratory, Chinese Academy of Sciences, Shenyang, China. From 2001 to 2002, he was a Research Associate with the Department of Automation and Computer-Aided Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2003, he has been with the Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai, China, where he is currently a Professor with the School of Optical-Electrical and Computer Engineering. His current research interests include nonlinear control, robust control, robot dynamic and control, visual servoing feedback control, and pattern identification.



**ZHANQUAN SUN** received the B.S. degree in mechanical design and theory from Shenyang Jianzhu University, Shenyang, China, in 2000, the M.Sc. degree in mechanical design and theory from the Dalian University of Technology, Dalian, China, in 2003, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007. From 2007 to 2018, he was a Researcher with the Shandong Computer Science Center, Jinan, China. He is currently an Associate Professor with the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China, where he has been with the Department of Control Engineering, since 2018. His current research interests include artificial intelligent, big data, and machine learning.



**SHUQUN CHENG** received the B.S. degree from Wenzhou Medical University, Wenzhou, China, in 1989, and the master's and Ph.D. degrees from the Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, China. Then, he entered to the Eastern Hepatobiliary Surgery Hospital for postdoctoral training. He is currently the Director of the Department of Hepatic Surgery VI, Eastern Hepatobiliary Surgery Hospital, The Second Military Medical University. As a Surgeon, he did almost 400 operations for liver cancer resection every year. Besides these, he has engaged in doing researches on basic science of liver cancer surgery and clinical problems, such as the mechanism of HCC tumor thrombus forming and comprehensive treatment for liver cancer.



**LEI GUO** received the B.S. degree in medicine from Shanghai Jiao Tong University, Shanghai, China, in 2010. He currently works at Eastern Hepatobiliary Surgery Hospital, Shanghai, as a Clinically Experienced Liver and Bile Surgeon. He has been supported by the Shanghai Science and Education Commission to carry out a project on deep learning of liver cancer CT scan images.

...