IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 27, 2019, accepted October 28, 2019, date of publication November 12, 2019,
date of current version November 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953085

Constrained-Focal-Loss Based Deep Learning for Segmentation of Spores

YAOCHI ZHAO[1,2], FUSHENG LIN[2,3], SHIGUANG LIU[1], ZHUHUA HU[2], (Member, IEEE), HUI LI[2], (Senior Member, IEEE), AND YONG BAI[2], (Senior Member, IEEE)

[1]Division of Intelligence and Computing, Tianjin University, Tianjin 300050, China
[2]School of Computer Science and Cyberspace Security, School of Information and Communication Engineering, Hainan University, Haikou 570228, China
[3]School of Computer Science and Electronic Engineering, Hunan University, Changsha 410012, China

Corresponding author: Shiguang Liu (lsg@tju.edu.cn)

This work was supported in part by the Hainan Province Natural Science Foundation of China under Grant 619QN195, in part by the National Natural Science Foundation of China under Grant 61963012, and in part by the Hainan Province Key Research and Development Project of China under Grant ZDYF2018015.

ABSTRACT The statistics of disease spores is significant for early strategy design of disease control in precision agriculture. To obtain the statistics information of spores in microscopic images, it is crucial to segment spores from images. In this paper, we research a deep learning based method to segment spores, taking anthrax spores as the research objects. We first built an anthrax spore dataset consisting of more than 40,000 spores with accurate labeled spore boundaries to advance the state of the art technology of spore statistics. Then on consideration of the complex class imbalances in actual anthrax spore images, we investigate how class imbalances and hard examples simultaneously influence the loss during training and we discover that hard examples are more likely to appear at the pixels of rare pixels, such as small class pixels and contour pixels. Based on this discovery, we propose Constrained Focal Loss (CFL), which focuses on small class objects, and has a constrained term related to hard examples. In addition, we further propose CFL*, where high importance is put on the pixels surrounding spore contours to improve classification accuracy. The results show that the mean IoU of the DeepLabv3+ trained with CFL* (called as CFL*Net) achieves 91.0%, higher than original DeepLabv3+ with cross-entropy by 8.6 points, and the DeepLabv3+ with Focal Loss by 10.4 points. Moreover, CFLNet* can achieve better performance than original DeepLabv3+, using less than one-third of the training samples and half of the training steps.

INDEX TERMS Image segmentation, class imbalance, focal Loss, hard example, convolutional neural networks (CNN).

I. INTRODUCTION

The diseases during the growth of crops and the storage/transportation of fruits and vegetables are one of the main reasons that cause the reduction of production in agricultural [1]. For example, about 50% of citrus peel is destroyed by diseases every year [2]. In the tropical and subtropical regions, due to the environmental factors, the post-harvest rot rate of fruits is as high as 50% [3]. The diseases are mainly attributed to the pathogenic role of pathogenic fungi, as an example 90% of crop diseases are caused by fungal spores [4]. If the fungal spores causing the diseases of crops, fruits and vegetables can be monitored, detected and counted at the early stages of diseases, and certain preventive measures are taken, it will be more effective. In addition, during the development of drugs for disease control, the statistics information of fungal spores can reveal the degree of resistance and activity of spores [5], thus providing strong technical support for the researches of new biological or chemical drugs. Therefore, the statistics of spores are of great significance for precision agriculture. The most commonly used manual statistics method of spores relies on naked eyes, which is time-consuming, laborious and is easy to make mistakes. With the development of computer vision and artificial intelligence, the automatic information statistics of spores have gradually attracted the attentions of researchers. Generally, they can be divided into two categories: traditional automatic methods and deep learning based methods.

For the traditional automatic methods, researchers need to design a suitable feature extractor to separate spores from

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

background, and based on the segmentation result, perform subsequent counting or recognition of spores. For example, in [6], to detect rice blast spores the authors extract edges of spore images and segment the edges with FCM, followed by morphological operators and shape-based filtering. In [7], morphological reconstruction filter and two-dimensional threshold of intensity are used, and watershed algorithm is employed to separate spores from background. In [8], urediospores are automatically segmented by clustering RGB pixel values with Kmeans, and then watershed algorithm is used for counting. Similarly, in [9], the authors segment spores by clustering pixel values with Kmeans, and isolate touching urediniospores based on their shape and area factors. In addition, Artificial Neural Network (ANN) is used for the detection of spores. For example, in [10], a semi-automated counting method of arbuscular mycorrhizal fungi spores is proposed. In [11], ANN is employed to automatically detect and count powdery mildew spores. These traditional methods are effective in their respective applications. However, due to the poor handcrafted feature extractors, they are successful only when the background is simple and must be supported by complex image acquisition methods [9]. For complex and changeable scenes, traditional handcraft features are no longer effective.

Different from traditional methods, deep learning can learn how to extract the features from a large number of samples. It emerges as an efficient technique, and becomes the mainstream approach in recent years. At present, there are few reports on the deep learning based method for the information statistics of spores. Only in the literature [12], the authors train a deep convolutional network to identify 5 species of spores. This work aims for the identification of spore species, yet the segmentation of spores has not been reported. The segmentation of disease spores is to find the boundaries separating the disease spores from the surrounding environment. Actually, it is crucial to segment spores from microscopic images for some information statistics, such as the density and the number of spores. However, to the best of our knowledge, currently there is a lack of anthrax spore dataset for the network training. So we first construct an anthrax spore dataset consisting of more than 40,000 spores with accurate labeled spore boundaries. Then, we research a deep learning based method for the segmentation of anthrax spores.

In the field of computer vision, the object segmentation is a task of pixel-level classification. For the task, the deep learning based methods have been pioneered by Long et al. [13], who improves the image-level classification to pixel-level classification. Later, many networks for pixel-level segmentation arise, mainly considering two aspects: the integration of local and global knowledge and the corresponding optimization of network structure. In terms of the integration of local and global knowledge, the authors of [14] put forward a post-processing stage using a Conditional Random Field (CRF); Zheng et al. [15] propose the reformulation of the dense CRF as an integral part of network; Pu et al. [16] propose GraphNet, which takes the graph that combines the low-level spatial relation and high-level semantic content as input and learns to predict image masks. In addition, the use of multi-scale predictions is also a possible way to deal with local and global knowledge integration [17], [18]. In terms of network structure, researchers often adopt convolutional-deconvolution and pooling-unpooling structures, with various design details, such as pooled layer index [19], convolution-like operation [20], gate function based structure-aware convolution [21], and so on. Especially, in the famous DeepLab series [14], [22], the use of atrous convolution improves greatly the network. However, whatever from the view of integration of knowledge or network structure, the focus of the researches is on feature extraction at the front of the network, and the problem of classification at the back is less studied.

Specifically, DeepLabv3+ [22] is considered to be the state-of-the-art in pixel-level classification for natural images. The mean IoU of deepLabv3+, trained on our anthrax spore dataset, reaches 83%. Compared with those traditional methods, it is a promising performance but still leaves a lot of room for improvement. For those spores that are adherent together, or close to each other, it is difficult to segment them from background with DeepLabv3+.

In this paper, to address the above problems, we pay attention to how to classify at the back end of network in deep neural network, and we identify class imbalance during training as the main obstacle impeding deepLabv3+ from improving its performance. DeepLabv3+ is designed based on the dataset with balanced class, such as PASCAL VOC and Cityscapes, where the problem of class imbalance is avoided when the dataset is being created. However, our images are taken randomly from real petri dishes, where the problem of class imbalance is serious.

We consider a novel loss function, where the loss value from the spore pixels is magnified. Then a natural question to ask is how to determine the degree of magnification. Generally, the degree of magnification can be calculated based on the ratio of the number of object pixels to that of the background. However, we have verified that these methods are not effective for our dataset. In another hand, in the field of machine learning, the problem of hard example is also worth paying attention to, which is demonstrated by the famous Focal Loss (FL) [23]. In FL, the loss of sample is re-weighted according to the probability value of its output. Using FL, the performance of one-stage detector can be improved to match the two-stage detector.

In this paper, instead of considering class imbalances or hard example alone, we explore the relationship between them. We found that hard examples are more likely to appear at spore pixels. That is to say, the pixels of spores are not only small class, but also hard examples. If we magnify their losses only based on the frequency, the loss of small class will dominate the whole loss, which would lead to over correction. Based on the above observation, we propose a Constrained Focal Loss function (CFL) that simultaneously takes into account class imbalances and hard sample. The loss function

focuses on the loss of the small class, meanwhile considers the fact that samples from small class are also hard examples. In addition to spore pixels, we discover pixels around spore contours are also prone to become hard examples during training. We further propose CFL*, where high importance is put on the pixels surrounding spore contours to improve classification accuracy.

The results show that the mean IoU of the DeepLabv3+ trained with CFL*(called as CFL*Net) achieves 91.0%, higher than the original DeepLabv3+ with cross entropy by 8.6, and the DeepLabv3+ with Focal Loss by 10.4 points. Moreover, CFLNet* can achieve better performance than original DeepLabv3+, using less than one-third of the training samples and half of the training steps.

Our main contributions are as follows:

(1) We propose CFL, where the problems of traditional class imbalance and hard example are taken into account simultaneously. With the CFL, the performance of trained model can be improved significantly.

(2) In the proposed CFL*, the imbalance between the contour pixels and the non-contour pixels are considered, and the losses of pixels surrounding the contours are automatically emphasized in the overall loss, thereby improving the classification rate. Further, with CFL and CFL*, the dependence of network on the number of samples can be significantly reduced and the efficiency of training can be greatly improved.

(3) We built a large dataset consisting of more than 40,000 anthrax spores with accurate spore contours, to advance the state-of-the-art in fungus statistics by placing the question of fungus spores segmentation.

The rest of the paper is organized as follows. Section 2 discusses the related deep learning techniques. Section 3 illustrates the material and anthrax spores dataset built in this paper. Section 4 presents our CFL and CFL* loss used to train the deep neural network. In Section 5, we validate our method by various experiments on the anthrax spores dataset. Conclusions are drawn in Section 6.

II. RELATED WORKS

In the field of deep learning, researchers are more interested in exploring the network structures to improve its performance and there are few work on the research of class imbalance. However, in real datasets for training, class imbalance is widespread [24], [25]. With the extensive application of deep learning to various practical fields, the class imbalance problems will be paid more and more attention.

A. CLASS IMBALANCE

In general, class imbalance refers to a large difference in the amount of data between different categories. Without correction mechanisms, classifier trained with imbalance dataset tends to be bias towards the majority classes. There exist two kinds of methods to correct this bias effect: data-level and algorithm-level [26].

At the data-level, the common way is to re-sample the training data, e.g., by the over-sampling of minority classes or under-sampling of majority classes or their combination [27], [28], [29]. Such approaches are known to change the underlying data distributions and may result in increased computational effort and risk of over-fitting. Another way to re-sample the training data is the ensemble of classifiers, where each classifier is induced by different samples from the original dataset [30], [31]. In order to better learn the distribution of dataset, the Generative Adversarial Network is employed to augment the dataset for the training of model [32], [33]. In addition, other techniques are employed for training the deep neural network, such as iterative sampling [34] and incremental rectification of mini-batches [26].

Compared with the data-based methods, the algorithm-based methods are more intuitive. By introducing weight for each category or each sample, more emphasis is put on the minority classes in the objective, thereby the algorithmic behavior changes. The simplest way is to define the weight parameter as a hyper parameter of category [23], yet it is difficult to obtain an optimal heuristic value. Commonly, weight parameters are set as statistics based cost-sensitive matrices [35], [36], or additional learnable parameters [37], [38]. However, the complexity of the model and the training cost will be increased, and generally it is difficult to optimize the cost matrix [26].

B. HARD EXAMPLES

Hard example mining is another concern in the community of machine learning and has been exploited in many tasks, e.g., object detection [23], [39], [40], face recognition [41], and image categorization [42]. In the hard example mining, it is argued that model can learn much from hard examples than easy examples, so the network can learn better by assigning higher weights to hard examples. The hard example mining can be conducted by solving two problems: (1) how to distinguish hard examples; (2) how to use hard examples to improve network performance. For example, an additional network [43], or additional enhancement images technique [44] are employed to distinguish hard examples. In [39], the authors use the model output probability to select hard examples, and then the network parameters is updated only according to the hard examples. More commonly, in FL [23], a reverse function of sample output probability was used to adjust the weights for each sample loss in the whole loss. Currently, the networks using the technique of hard example mining have been greatly improved, especially in the task of object detection [23], [40]. Yet, the problems of hard example are considered only from the view of the difficulty degree of samples itself.

To the best of our knowledge, the joint influences of class imbalance and hard example on loss value are rarely studied. In [23], [40], the problems of class imbalance and the hard example are both considered, by adding two independent parameters which represent class balance and difficulty degree respectively, into the loss function. However,

the two parameters are independent and the best parameters are obtained through multiple testing the combinations of the two parameters, which causes the parameter debugging difficulty. Actually, the relationship between the hard example problem and the class imbalance problem is also worth attention. During training, samples from small class are always ignored, their feature cannot be well updated into the model through learning and they are likely to become hard samples gradually.

III. MATERIALS AND DATASET

Considering that crops, fruits and vegetables are susceptible to anthrax disease, we take anthrax spores as our research objects. It is well known that a large number of samples are a necessary prerequisite for the effective application of deep learning methods. However, to the best of our knowledge, at present the fungal spore dataset is only reported in the literature [9]. In their dataset, there are 40,800 spores, but the annotations of spore boundaries are missing for training. Moreover, the morphology of their Penicillioides, Restrictus, Versicolor, Cladosporium and Eurotium spores differs greatly from that of anthrax spores.

In order to advance the latest technology of spore statistics, we construct a new anthrax fungal spore dataset by proposing the problem of fungal spore segmentation. We have collected 1600 (resolution:1920*1080) images of anthrax spores and construct a dataset by data augmentation. In our dataset, there are 12800 images and more than 40,000 spores with accurate labeled spore boundaries. In order to enhance the robustness of network to complex conditions, the interferences are included in our dataset, such as hyphae (as shown in the red ellipse of Fig.1(a)), impurities (as shown in the blue ellipse of Fig.1(b)) and stains (as shown in the red ellipses of Fig.1(b)).

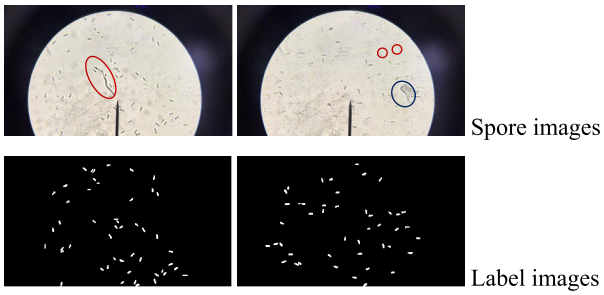
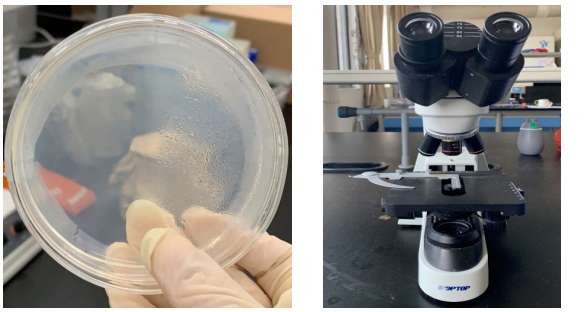


FIGURE 1. Samples of anthrax fungus dataset.

Spore culture vessel and experimental image acquisition device are shown in Fig.2. The process of obtaining anthrax spores images is described as follows. Firstly, 10 strains and blocks of anthrax with a diameter of 6mm are taken by using a hole punker, which are respectively connected to 50mL liquid PD medium and then are cultured for 2 days by shock at 180 RPM. Secondly the bacterial liquid is filtered and we take 5 microliter onto a glass slide. Then the spore images are taken by a microscope (model: SDPTOP E5).



(a) Spore culture vessel (b) Microscope for photography

FIGURE 2. Experimental device.

IV. CONSTRAINED-FOCAL-LOSS (CFL) AND CFL*

The original DeepLabv3+, which takes cross entropy (CE) as the loss function, can reach 83% mIoU. However, the trained network is failure to identify some pixels, such as pixels on or around the spore contours, and those spores pixels that are adherent together, or close to each other. We argue the class imbalance as the main obstacle impeding network from accurately classifying those pixels and we propose novel loss functions to improve the learning strategy of network. Our loss is designed to address the binary segmentation of unbalanced data. For the binary segmentation, CE is widely used, so we introduce our loss function starting from the cross entropy loss.

A. CROSS ENTROPY

The objective of training is to optimize the mean of the sample losses sum, as shown in Eq. (1), and obtain the model parameters when the loss reaches a minimum.

$$loss = \frac{1}{N}\sum_{i} L_i(p, y) \tag{1}$$

In Eq. (1), N is the number of the samples, L_i is the loss of the ith sample, p and y is the predicted value and ground truth class label respectively. In binary classification task, cross entropy loss (CE) is used to represent the loss of sample, as shown in Eq. (2).

$$CE(p, y) = \begin{cases} -\log(p) & if\ y = 1 \\ -\log(1-p) & otherwise \end{cases} \tag{2}$$

where $y \in \{-1, +1\}$ and $p \in [0, 1]$ is the model's predicted probability for the class with label $y = 1$. For notational convenience, we define p_t as:

$$p_t(p, y) = \begin{cases} p & if\ y = 1 \\ 1-p & \text{otherwise} \end{cases} \tag{3}$$

and rewrite $CE(p, y) = CE(p_t) = -\log(p_t)$. It is easy to know p_t is the probability of ground truth class.

B. BALANCED CROSS ENTROPY

To solve the problem of class imbalance, a common method is introducing a weighting factor α_t to balance the loss between

objects and background (called as Balance Cross Entropy, BCE), as shown in Eq. (4).

$$BCE(p, y) = -\alpha_t \log(p_t) \tag{4}$$

In the above equation, α_t is represented as:

$$\alpha_t(y) = \begin{cases} \alpha & \text{if } y = 1 \\ 1 - \alpha & \text{otherwise} \end{cases} \tag{5}$$

where α may be treated as a constant, or the reciprocal of category frequency f (for convenience, we call it as Adaptive Balance Cross Entropy, BCE_A), and the BCE_A can be represented as:

$$BCE\_A(p, y) = -1/f(y) \log(p_t) \tag{6}$$

By balancing, the weight of the pixel in smaller class is higher than the one in larger class, meaning that more importance is put on the small class. However, for the segmentation of anthrax spores, such an up-weighting strategy for small class cannot yield a more optimized model, verified by our experiments.

C. FOCAL LOSS (FL)

In [23], the authors propose FL, in which the loss assigned to well-classified example is down-weighted and thus hard negatives is the focus, as shown in Eq. (7), where γ is the tunable focusing parameter, denoting the degree of focus.

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \tag{7}$$

Further, a α-balanced variant of the FL (called FL$^{\#}$ for convenience in this paper) is proposed to emphasize the small class, as shown in Eq. (8).

$$FL^{\#}(p_t) = -\alpha_t(1 - p_t)^{\gamma} \log(p_t) \tag{8}$$

where α_t is defined by Eq. (5) and α in α_t is a heuristic value. Moreover, the authors of FL point that α interacts with γ making it necessary to select the two together. In general, α should be decreased slightly as γ increases.

D. CONSTRAINED FOCAL LOSS (CFL)

Our method is inspired by the above FL$^{\#}$. Since α and γ in FL$^{\#}$ are to redefine the loss function from the perspective of class unbalance and hard example respectively, and they interact to influence the performance of the trained model, there should be some relationship between the class imbalance and the hard example.

Our conjecture is confirmed, as shown in Fig.3. In the early stages of training, the losses of spore pixels are usually higher than that of background pixels (as shown in Fig.3(a)). As the number of iterations increases, the loss is gradually concentrated on two kinds of pixels: the ones around the contours of spore, and the unusual spore pixels that are strongly disturbed by interferences (as shown in Fig.3 (b)) or adjacent spore (as shown in Fig.3 (c)). At last, the loss mainly derives from those pixels around the spores contours that are extremely rare in dataset. In conclusion, the hard examples with large loss always appear at the pixels of rare pixels.

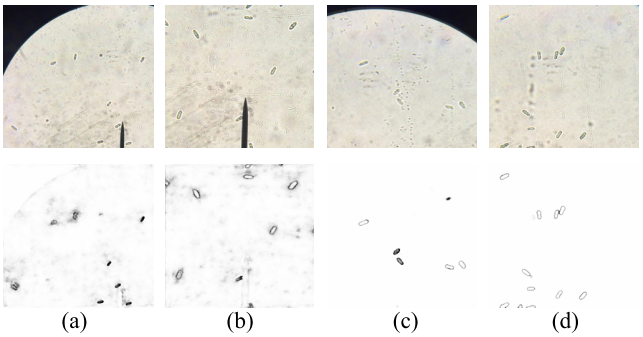


FIGURE 3. Prediction probability maps of ground truth class during the training of DeepLabv3+ ((a) - (d) correspond to the iterations of 60, 1110, 14200 and 17700, respectively).

Further, we randomly test the ratio of spore pixels to the whole hard examples during training (DeepLabv3+). The detail results are given in Table 1, in which p_t is the prediction probability of ground truth class. From Table 1, we can know that pixels from small class overwhelm the whole hard examples.

TABLE 1. The ratio of spore pixels to the overall hard examples.

The number of iterations	60	1000	6000	1000	16000
p_t <=0.5	0.76	0.97	0.77	0.85	0.63
p_t <=0.25	0.93	1.00	0.84	0.95	0.67

From Fig.3 and Table 1, we can explain why the method in [24] does not work well in the segmentation of anthrax spores. Because spore pixels are not only pixels in small class to be amplified in the loss function, but also hard examples with large losses, if the losses of spore pixels are amplified only according to frequency, the losses derived from spore pixels will dominate the overall loss value, leading to over correction.

Based on the above observation, we propose a novel loss function, Constrained Focal Loss (CFL), to consider class imbalance and hard example simultaneously. It can be represented as

$$CFL(p, y) = -\alpha_t^{1/\beta} \log(p_t) \tag{9}$$

$$\alpha_t(y) = \begin{cases} n_b/n_o & \text{if } y = 1 \\ 1 & \textit{otherwise} \end{cases} \tag{10}$$

In the above equations, n_b and n_o are respectively the numbers of background pixel and object pixel, which means that the loss of small class will be amplified and the loss focus on the small class. It is noted that α_t can be calculated from dataset. On the other hand, β is a value greater than 1, meaning that the amplification will be constrained by β, whose function is similar to the role of γ in FL, representing the

difficulty degree to be classified. The harder the classification is, the greater β is.

Unlike the FL [23] with a uniform constant α_t, CFL accurately amplifies the loss of small class based on the statistics. Comparing to the FL$^{\#}$ [23], where there are two parameters α_t and γ in the loss function, CFL has only one parameter to be adjusted. In addition, different from the BCE_A [24], CFL takes account of class imbalances and hard example simultaneously, which avoids the over-amplification of the loss from small class.

E. CFL*

In CFL, the loss focuses on spore pixels, which only considers the traditional class imbalance. However, another imbalance between the number of contour pixels and the one of non-contour pixels is not involved. In actually, except spore pixels, pixels surrounding spore contours account for a large proportion in hard examples, as shown in Table 2. Besides, we can know that as the increasing of the iterations, the proportion tends to increase.

TABLE 2. The ratio of pixels surrounding the spore contours to the overall hard examples.

The number of iterations	60	1000	6000	1000	16000
p_t <=0.5	0.12	0.01	0.19	0.15	0.37
p_t <=0.25	0.07	0.00	0.16	0.05	0.33

Therefore, the pixels surrounding the spore contours are typical hard examples in background. On the basis of CFL, **we further propose to focus the loss on the pixels surrounding** the spores. It is noted that the pixels surrounding the spore contours are easy to be obtained in advance from dataset. So we can add the positions prior of pixels surrounding spore contour into the loss function. We denote the label image as I and its dilated image as I^*. A variant of CFL (CFL*) can be expressed as:

$$CFL^*(p, y, y^*) = -{\alpha^*}_t^{1/\beta} \log(p_t) \tag{11}$$

where $y \in I$, $y^* \in I^*$, and α_t^* is an extended weight function, denoted as Eq. (12). Using CFL*, the loss of pixels surrounding spore contour ($y_i = 0$ and $y_i^* = 1$) will be as important as the inner pixel ($y_i = 1$) in the loss function.

$$\alpha_t^*(y^*) = \begin{pmatrix} n_b/n_o & y^* = 1 \\ 1 & otherwise \end{pmatrix} \tag{12}$$

Table 3 shows the ratio of spore pixels and pixels surrounding spore contours to the whole hard examples. It can be seen that the sum of spore pixels and their surrounding pixels are almost equal to the number of hard examples. Hence, the main imbalance factors that hinder network performance are dealt in CFL*.

TABLE 3. The total ratio of spore pixels and pixels surrounding spore contours to the whole hard examples.

The number of iterations	60	1000	6000	1000	16000
p_t <=0.5	0.88	0.98	0.97	1.00	1.00
p_t <=0.25	0.99	1.00	1.00	1.00	1.00

V. EXPERIMENTS AND DISCUSSIONS

A. IMPLEMENTATION DETAILS

Training Protocol: Our experiment is performed on an Intel(R) Core(TM) I7-7800x CPU @ 3.5GHZ and a NVIDIA Titan XP GPU with 12GBRAM. We use DeepLabv3+ network to extract the features of spore images. Except the hyper-parameters, all the parameters in the model are randomly initialized. In addition, a batch size of 2 is set and a ''poly'' learning rate policy is used, where the initial learning rate is 0.007 and the power is 0.9.

We conduct our experiments in two stages. The two stages aim to test the effect of the network and the training efficiency, respectively. We use the mean IoU (intersection over union) to evaluate the experimental results. Let G is the ground truth set and I is the predicted set, IoU can be denoted as follows.

$$IoU = \frac{|G \cap I|}{|G \cup I|} \tag{13}$$

B. THE EFFECT OF INFERENCE WITH CFL

In the first stage, we use different loss functions to train DeepLabv3+ only on a partial dataset with 400 images, selected randomly from the training dataset. We demonstrate the effectiveness of CFL, comparing it to the state-of-the-art losses: (CE) [22], BCE-A (α is set as Eq.(6)) [24], FL, and FL$^{\#}$ [23]. For FL and FL$^{\#}$, we evaluate the best parameter or the best combination of parameters; the examples of effects are shown in Fig .4.

The third row of Fig.4 shows the segmentation results of network trained with CE. There exist some spores that failure to be classified in image. To show them clearly, we mark them with ellipses in the first row of Fig.5 and magnify them in the second row of Fig .5. It is obvious that for those special spores, such as the contaminated spores (as shown in Fig.5(a)), those that are close together (in Fig.5(c))), and the ones at the edge of the field of microscope view (as shown in Fig.5(b)), the network trained with CE fail to segment them. The reason is that those spores are sparsely distributed, in other words, they are rare objects in the dataset. CE does not pay special attention to this type of spores, so they are neglected during training. While in CFL, the loss focuses on the spore pixels. Hence CFL has a higher focus on the specific spore pixels, and the network trained with CFL (for convenience, we call the network CFLNet later in this paper) can better identify those spores.

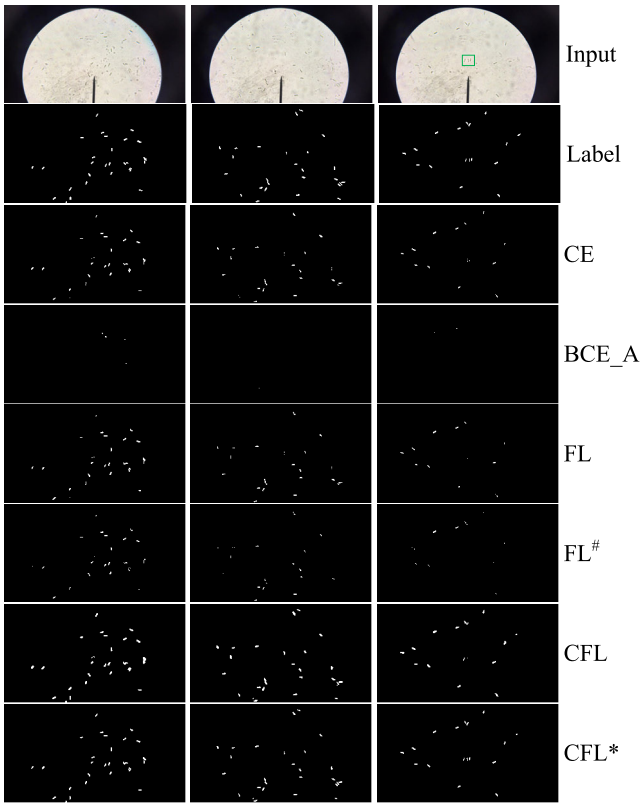


FIGURE 4. Examples of result inferred by network using different losses.

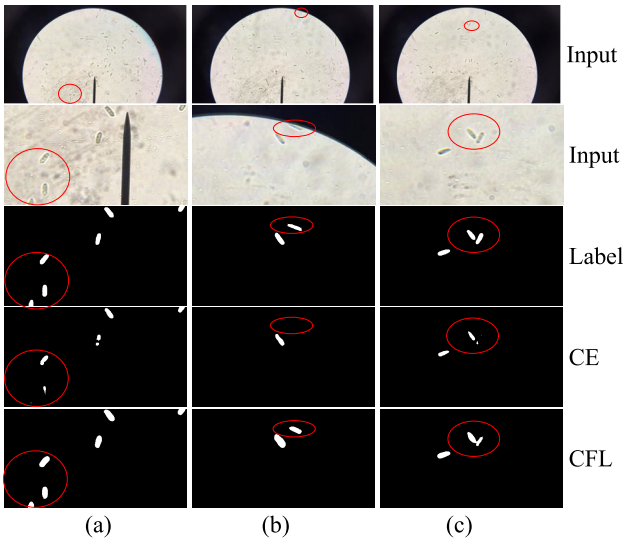


FIGURE 5. Comparisons of segmentation result between CE and CFL.

We compare the performances of DeepLabv3+ trained with different losses as shown in Table 4. From the 4th line of Fig.4 and Table 4, it can be seen that using BCE_A, the segmentation results are bias to the spore class and over correction are very obvious. In our dataset, the mean ratios of the number of spore pixel to that of the background are about 200. Actually, a suitable magnification ratio is much less than the mean ratio. This result is due to that the majority of hard example comes from small classes, and increasing the weight parameters only according to class frequency will lead to the over-amplification of the loss of small classes.

TABLE 4. Quantitative comparisons among different losses in DeepLabv3+.

Method	Mean IoU
CE [22]	82.4
BCE-A [24]	52.5
FL $\gamma = 2$ [23]	80.6
FL# $\gamma = 2.5$ $\alpha = 0.3$ [23]	76.0
CFL $\beta = 2$	89.1
CFL $\beta = 2.5$	87.3

Comparing FL to FL#, the performance of FL is better than FL#. In our experiments, by multiple adjustments of α and β, we obtain the best combination of parameters, but the mean IoU of trained network with FL# is still lower than FL. Comparing FL to our CFL, the mean IoU of CFL is higher than FL by 10.4 points for DeepLabv3+.

In addition, to further verify the effectiveness of our loss, we use different losses to train U-Net [45], another famous network for semantic segmentation, on our dataset. The quantitative comparisons among different losses are shown in Table 5. We can know the comparison results in U-Net are similar to DeepLabv3+.

TABLE 5. Quantitative comparisons among different losses in U-Net.

Method	U-Net
CE [22]	73.2
BCE-A [24]	81.7
FL $\gamma = 2$ [23]	81.1
FL# $\gamma = 2$ $\alpha = 0.3$ [23]	70.0
CFL $\beta = 2$	86.7
CFL $\beta = 2.5$	**88.7**

We tested different values of β in CFL. It can be seen in Fig.6 that the optimal value of β is 2 and 2.5 for DeepLabv3+ and U-Net respectively. For DeepLabv3+, when β is between 2 and 4, the IoU is not sensitive with β, and for U-Net when β is between 2 and 3, the IoU is not sensitive with β as well. Thus it is easy when tuning parameters for CFL.

C. THE EFFECT OF INFERENCE WITH CFL*

As described above, the network trained with CFL can better identify the disturbed spores than with other losses. However, there are still some adjacent spores that cannot be correctly distinguished by the network, such as the spores in the green rectangular in the 1th line of Fig.4(c). In this case, CFL*Net can correctly classify them.

We compare our CFL with CFL*, as shown in Fig.7 and Table 6. To further clearly illustrate their difference, we also

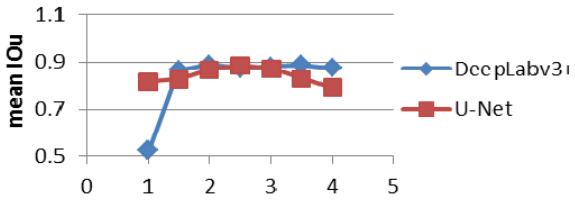


FIGURE 6. The mean IoU of network using CFL with different β.

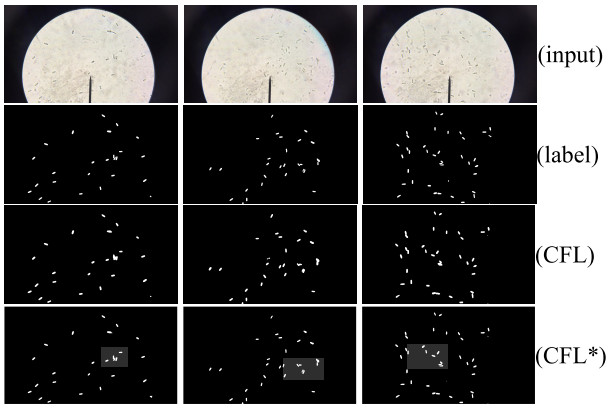


FIGURE 7. Comparison of result with CFL and CFL*.

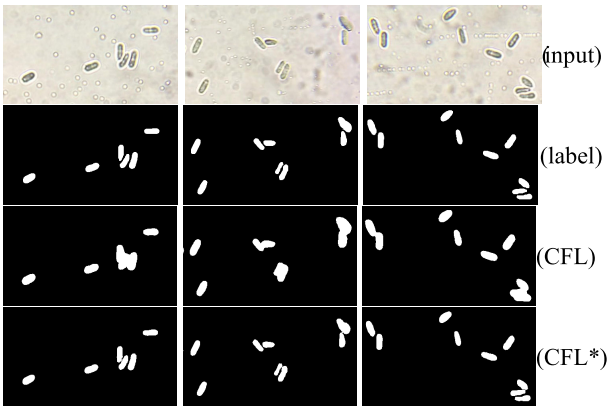


FIGURE 8. Comparison of Local result with CFL and CFL*.

TABLE 6. Quantitative comparisons between CFL and CFL*.

Method (DeepLabv3+)	mean IoU(%)
CFL $\beta=1.5$	86.7
CFL $\beta=2$	89.1
CFL $\beta=2.5$	86.1
CFL* $\beta=1.5$	**89.1**
CFL* $\beta=2$	**91.0**
CFL* $\beta=2.5$	**89.9**

magnify the mask part of Fig.7 in Fig.8. Increasing the weights of losses of pixels surrounding spore contours will improve the network performance. Especially for those pixels near the adhesion of adhering spores, CFLNet is difficult to discriminate them from background, while CFLNet* is easy to classify them.

Table 6 shows the comparison between CFL and CFL*. In CFL*, more importance is put on the pixels surrounding spore contours, and the mean IoU increases by about one to four points.

Comparing DeepLabv3+ to U-Net, DeepLabv3+ generally performs better than U-Net. For DeepLabv3+, among all the methods, the network using CFL* as loss function provides clear segmentation of spores, and its mean IoU achieves 91%, higher than CE by 8.6 points, and FL by 10.4 points. CFLNet* can discriminate the pixels of spores from the pixels of stains in background, and get a clear spore contours for those spores that stick together or close.

Fig.9 illustrates the performance evolution of network with different losses during training. CFL and CFL* outperform all previous loss functions. The superior performances of CFL* are demonstrated from the middle stages of the training (11k training steps). In particular, the performance of CFL* is improving steadily relative to other losses.

TABLE 7. Comparisons of training efficiency.

Method (DeepLabv3+...)	Number of sample	Number of iteration(K)	Mean IoU(%)
CE [22]	1400	35	90.2
FL $\gamma=2$ [23]	1400	35	88.6
FL# $\gamma=2$ α=0.25 [23]	1400	35	79.4
CFL* $\beta=2$	**400**	**18.2**	**91.0**

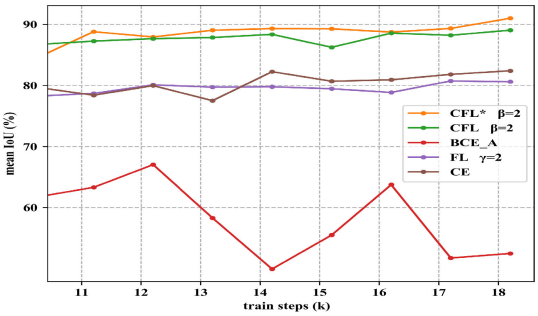


FIGURE 9. Evolution of mean IoU with different loss functions.

D. THE EFFICIENCY OF TRAINING USING CFL*

We compare the efficiency of training among CE, FL, FL*, and CFL* in the second stage. We train DeepLabv3+ with different loss functions on 1400 or 400 samples to achieve approximate performance. The detail mean IoU, the number of iterations and the number of samples are shown in Table 7. Compared to other loss functions, the network using CFL* can obtain better performance with less than one-third of the training samples and about half of the training steps required by other loss functions. Thus it is effective and efficient to choose CFL* as loss function.

It is worth mentioning that we plan to further establish a dataset containing nearly 100 fungal spores. The CFL* reduces the dependence on labeled data, which will reduce the workload of annotation, and facilitate efficient construction of other fungal spores datasets.

VI. CONCLUSION

In this paper, we built an anthrax spore dataset, where there are more than 40,000 spores with accurate spore contours. For the segmentation of anthrax spores, we first identified class imbalance as the main obstacle impeding network from achieving its performance. In addition, we discovered that spore pixels are more likely to become hard examples. We proposed the Constrained Focal Loss, where the weight of samples from small class is increased, and a constrained term is introduced to limit the weight. Furthermore, the position prior of spore contour is incorporated into the loss to accurately classify the pixels surrounding spore contours.

We conducted experiments on a small dataset and found that using our loss function the mean IoU can achieve 91%, much higher than other existing loss functions. Moreover, our method requires less training samples and training iterations. Due to the less demanding on samples, our method can be used for training segmentation network especially on the small dataset. It is worth noting that although our work is to segment spores from microscopic images, the loss functions proposed in this paper, aiming to solve the problem of class imbalance, can be used to guide the training of deep networks for semantic segmentation in other fields. Our future work is to extend the idea of this paper to the semantic segmentation of multiple classes with class imbalances.

APPENDIX

Supplementary data associated with this article can be found, in the online version, at https://drive.google.com/drive/folders/1-Cjy4tkhgBxTip2B_ 3esqw8xWkzofcZX.

ACKNOWLEDGMENT

The authors would like to thank Prof. X. Shi of Hainan University for his help in the data collection process.

REFERENCES

[1] J. G. A. Barbedo, "A review on the main challenges in automatic plant disease identification based on visible range images," *Biosyst. Eng.*, vol. 144, pp. 52–60, Apr. 2016.

[2] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," *Comput. Electron. Agricult.*, vol. 153, pp. 12–32, Oct. 2018.

[3] J. Pitt, A. D. Hocking, *Fungi and Food Spoilage*, vol. 519. New York, NY, USA: Springer, 2009.

[4] C. V. Robinson and A. H. Bishop, "A disclosure gel for visual detection of live Bacillus anthracis spores," *J. Appl. Microbiol.*, vol. 126, no. 6, pp. 1700–1707, 2019.

[5] R. Korsnes, K. Westrum, E. Fløistad, and I. Klingen, "Computer-assisted image processing to detect spores from the fungus Pandora neoaphidis," *MethodsX*, vol. 3, pp. 231–241, Jan. 2016.

[6] L. Qi, Y. Jiang, Z. H. Li, X. Ma, Z. X. Zheng, and W. J. Wang, "Automatic detection and counting method for spores of rice blast based on micro image processing," *Trans. Chin. Soc. Agricult. Eng.*, vol. 31, no. 12, pp. 186–193, 2015.

[7] Y. P. Xiong and A. C. Wang, "Microscopic image recognition method of melampsora larici-populina basidiospore," *J. Northeast Forestry Univ.*, vol. 44, no. 7, pp. 116–119, 2016.

[8] X. L. Li, Z. H. Ma, Z. Y. Sun, and H. G. Wang, "Automatic counting for trapped urediospores of Puccinia striiformis f. sp. tritici based on image processing," *Trans. Chin. Soc. Agricult. Eng.*, vol. 29, no. 2, pp. 199–206, 2013.

[9] Y. Lei, Z. Yao, and D. He, "Automatic detection and counting of urediniospores of Puccinia striiformis f. sp. tritici using spore traps and image processing," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 013647.

[10] W. Danping, W. Botao, and Y. Yue, "The identification of powdery mildew spores image based on the integration of intelligent spore image sequence capture device," in *Proc. 9th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2013, pp. 177–180.

[11] C. A. O. Melo, J. G. Lopes, A. O. Andrade, R. M. P. Trindade, and R. S. Magalhães, "Semi-automated counting of arbuscular mycorrhizal fungi spores using artificial neural network," *IEEE Latin Amer. Trans.*, vol. 15, no. 8, pp. 1566–1573, Jul. 2017.

[12] M. W. Tahir, N. A. Zaidi, A. A. Rao, R. Blank, M. J. Vellekoop, and W. Lang, "A fungus spores dataset and a convolutional neural network based approach for fungus detection," *IEEE Trans. Nanobiosci.*, vol. 17, no. 3, pp. 281–290, Jul. 2018.

[13] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and L. A. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[15] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1529–1537.

[16] M. Pu, Y. Huang, Q. Guan, and Q. Zou, "GraphNet: Learning image pseudo annotations for weakly-supervised semantic segmentation," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 483–491.

[17] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 2650–2658.

[18] A. Roy and S. Todorovic, "A multi-scale cnn for affordance segmentation in RGB images," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 186–201.

[19] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[20] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.

[21] Z. Cheng, J. Li, and C. Yuan, "Gate function based structure-aware convolution for scene semantic segmentation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2017, pp. 253–258.

[22] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 801–818.

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.

[24] M. Mostajabi, P. Yadollahpour, and G. Shakhnarovich, "Feedforward semantic segmentation with zoom-out features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3376–3385.

[25] C. Zhang, J. Bi, S. Xu, E. Ramentol, G. Fan, B. Qiao, and H. Fujita, "Multi-imbalance: An open-source software for multi-class imbalance learning," *Knowl.-Based Syst.*, vol. 174, pp. 137–143, Jun. 2019.

[26] Q. Dong, S. Gong, and X. Zhu, "Imbalanced deep learning by minority class incremental rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1367–1381, Jun. 2019.

[27] D. Kong, J. Tang, Z. Zhu, J. Cheng, and Y. Zhao, "De-biased dart ensemble model for personalized recommendation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2017, pp. 553–558.

[28] 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.

[29] T. Maciejewski and J. Stefanowski, "Local neighbourhood extension of SMOTE for mining imbalanced data," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, Apr. 2011, pp. 104–111.

[30] E. Fernandes, R. L. Rocha, B. Ferreira, E. Carvalho, A. C. Siravenha, A. C. S. Gomes, S. Carvalho, and C. R. B. de Souza, "An ensemble of convolutional neural networks for unbalanced datasets: A case study with wagon component inspection," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2018, pp. 1–6.

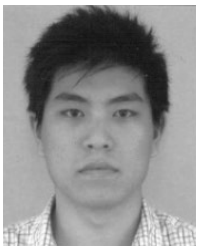
[31] Y. Qian, Y. Liang, M. Li, G. Feng, and X. Shi, "A resampling ensemble algorithm for classification of imbalance problems," *Neurocomputing*, vol. 143, pp. 57–67, Nov. 2014.
[32] M. Rezaei, H. Yang, K. Harmuth, and C. Meinel, "Conditional generative adversarial refinement networks for unbalanced medical image semantic segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2019, pp. 1836–1845.
[33] S. Liu, J. Zhang, Y. Chen, Y. Liu, Z. Qin, and T. Wan, "Pixel level data augmentation for semantic image segmentation using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2019, pp. 1902–1906.
[34] R. R. Morales, D. Domínguez, E. Torres, and J. H. Sossa, "Image segmentation through an iterative algorithm of the mean shift," in *Advances in Image Segmentation*. London, U.K.: IntechOpen, 2012.
[35] J. Xu, A. G. Schwing, and R. Urtasun, "Learning to segment under various forms of weak supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3781–3790.
[36] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 9268–9277.
[37] S. R. Bulò, G. Neuhold, and P. Kontschieder, "Loss max-pooling for semantic image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 7082–7091.
[38] Y. Li and N. Vasconcelos, "REPAIR: Removing representation bias by dataset resampling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 9572–9581.
[39] 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.*, Jun. 2016, pp. 761–769.
[40] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: Towards balanced learning for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 821–830.
[41] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.
[42] H. O. Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep metric learning via lifted structured feature embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4004–4012.
[43] T.-H. Kim and J. Choi, "ScreenerNet: Learning self-paced curriculum for deep neural networks," 2018, *arXiv:1801.00904*. [Online]. Available: https://arxiv.org/abs/1801.00904
[44] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 7340–7351.
[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

YAOCHI ZHAO received the M.S. degree in pattern recognition and intelligent systems from Central South University, Changsha, China, in 2005. She is currently pursuing the Ph.D. degree with Tianjin University, China. She is engaged in teaching and research work at the School of Computer Science and Cyberspace Security, Hainan University, China. She is currently an Associate Professor. Her current research interests include computer vision and artificial intelligence.

FUSHENG LIN received the B.S. degree from Hainan University, China, in 2019. He is currently pursuing the master's degree with Hunan University. His current research interests include artificial intelligence computer networks and data center networks.

SHIGUANG LIU received the Ph.D. degree from the State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou, China. He is currently a Professor with the School of Computer Science and Technology, Division of Intelligence and Computing, Tianjin University, Tianjin, China. His research interests include image/video processing, computer graphics, visualization, and virtual reality.

ZHUHUA HU (M'19) received the B.Eng. degree from Jilin University, in 2002, the M.Eng. degree from Jilin University, in 2005, and the Ph.D. degree from Hainan University, in 2019. He was a Software Engineer with the Ningbo BIRD Research Institute of China, from 2005 to 2006. He was a Senior Engineer with the Nanjing Research Institute of ZTE, from 2006 to 2007. He was the Minister of the Software Department, Shanghai Aoxun Information Technology Company, Ltd., from 2007 to 2009. He has been an Associate Professor with the College of Information and Communication Engineering, Hainan University, since 2014. His current research interests include artificial intelligence and mobile communications. He has acted as a Reviewer for the IEEE INTERNET OF THINGS JOURNAL, IEEE ACCESS, the *Engineering Applications of Artificial Intelligence*, the *International Journal of Electronic*, and the *AEU - International Journal of Electronics and Communications*.



HUI LI (M'06–SM'14) received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology (HIT), China, in 2000, 2002, and 2006, respectively. He held a postdoctoral position and an Associate Professor with HIT, Shenzhen, from 2005 to 2009, and an Associate Professor with Zhejiang University, from 2009 to 2014. He was the General Manager Assistant of the Shanxi Electronic Information Group, China, in 2011, and a Visitor Scholar with the National Institute of Standard and Technology, USA, in 2013. He is currently a Full Professor (D-level talent in Hainan University). He is the Director of the Engineering Research Center of Marine Communication and Networks in Hainan Province. His research interests include artificial intelligence and maritime communications. He received the Best Paper Award in the International Conference of CSPS2018 (July 14-16, Dalian, PRC).

YONG BAI (S'96–M'01–SM'16) received the B.S. degree from Xidian University, China, in 1992, the M.S. degree from the Beijing University of Posts and Telecommunications (BUPT), China, in 1995, and the Ph.D. degree from Rutgers-The State University of New Jersey, in 2001. He was with PacketVideo Corporation, from 2000 to 2002. He was with Motorola, from 2002 to 2004. He was with CEC Wireless, from 2004 to 2005. He was a Senior Researcher with DOCOMO Beijing Communication Labs, from 2005 to 2009. He has been a Professor with the College of Information Science and Technology, Hainan University, since 2010. His current research interests include artificial intelligence and mobile communications. He serves as an Associate Editor for the *EURASIP Journal on Wireless Communications and Networking*, since March 2017.