The Visual Computer (2022) 38:1051–1063
https://doi.org/10.1007/s00371-021-02067-9

ORIGINAL ARTICLE

Class-discriminative focal loss for extreme imbalanced multiclass object detection towards autonomous driving

Guancheng Chen[1] · Huabiao Qin[1]

Accepted: 7 January 2021 / Published online: 20 January 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Currently, modern object detection algorithms still suffer the imbalance problems especially the foreground–background and foreground–foreground class imbalance. Existing methods generally adopt re-sampling based on the class frequency or re-weighting based on the category prediction probability, such as focal loss, proposed to rebalance the loss assigned to easy negative examples and hard positive examples for single-stage detectors. However, there are still two critical issues unresolved. In practical applications, such as autonomous driving, the class imbalance will become more extreme due to the increased detection field and target distribution characteristics, needing a more effective way to balance the foreground–background class imbalance. Besides, existing methods typically employ the sigmoid or softmax entropy loss for classification task, which we believe is not capable to realize the foreground–foreground class balance. In this paper, we propose a new form of focal loss by re-designing the re-weighting scheme that can calculate the weight according to the probability as well as widen the weight difference of the examples. Besides, we introduce the extended focal loss to multi-class classification task by reformulating the standard softmax cross-entropy loss for better utilizing the discriminant difference of foreground categories, thereby yielding a class-discriminative focal loss. Comprehensive experiments are conducted on the KITTI and BDD dataset, respectively. The results show that our approach can easily surpass focal loss with no more training and inference time cost. Besides, when trained with the proposed loss function, current state-of-the-art object detectors no matter in one-stage or two-stage paradigms can achieve significant performance gains.

Keywords Object detection · Focal loss · Class-discriminative · Class imbalance

1 Introduction

Modern object detection algorithms are developed based on convolutional neural networks (CNNs) and can be roughly divided into two categories, two-stage detectors [13] and one-stage detectors [24,40]. Compared with the classical object detector, the modern object detector has evolved from the traditional manual feature extractor (e.g. LBP [25], Haar [19,27,38], or HOG [7,17,23]) to the semantic feature extractor based on CNN, but inherits the two-stage and proposal-driven mechanism. As popularized in the R-CNN framework [13], an appropriate amount of candidate target locations is generated in the first stage, and then finely classified into foreground or background classes in the second stage. Since then, a series of advanced two-stage detectors [6,12,14,18,20,33] have been proposed, and have achieved a constant improvement in accuracy on the challenging PASCAL VOC [8] and COCO benchmark [22].

Although the two-stage detector can achieve high detection accuracy, it has the disadvantage of tacking too long time due to the need to carry out two stages of training and inference. In contrast, the one-stage detector is proposed, such as YOLO series [1,5,30–32], SSD [10,24], RetinaNet [21] and FCOS [36], to complete the target classification and location regression only in a single stage. The one-stage detector greatly reduces the inference time while achieves considerable detection accuracy, and is considered to be a more efficient and elegant target detection method especially for autonomous driving application, which requires a high tradeoff between accuracy and speed.

✉ Huabiao Qin
eehbqin@scut.edu.cn

Guancheng Chen
eechengc@mail.scut.edu.cn

[1] School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China

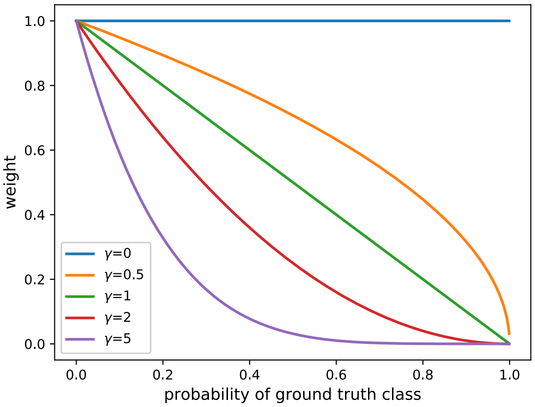


Fig. 1 Illustration of focal weight. Focal loss has introduced a scheme of weighting the loss of the examples based on the predicted probability, which we call focal weight. It is determined by the probability of the ground truth class as well as the parameter γ. As γ increases, the contribution of the easy examples decreases

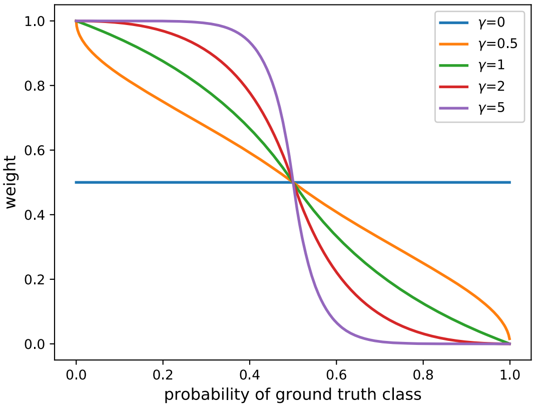


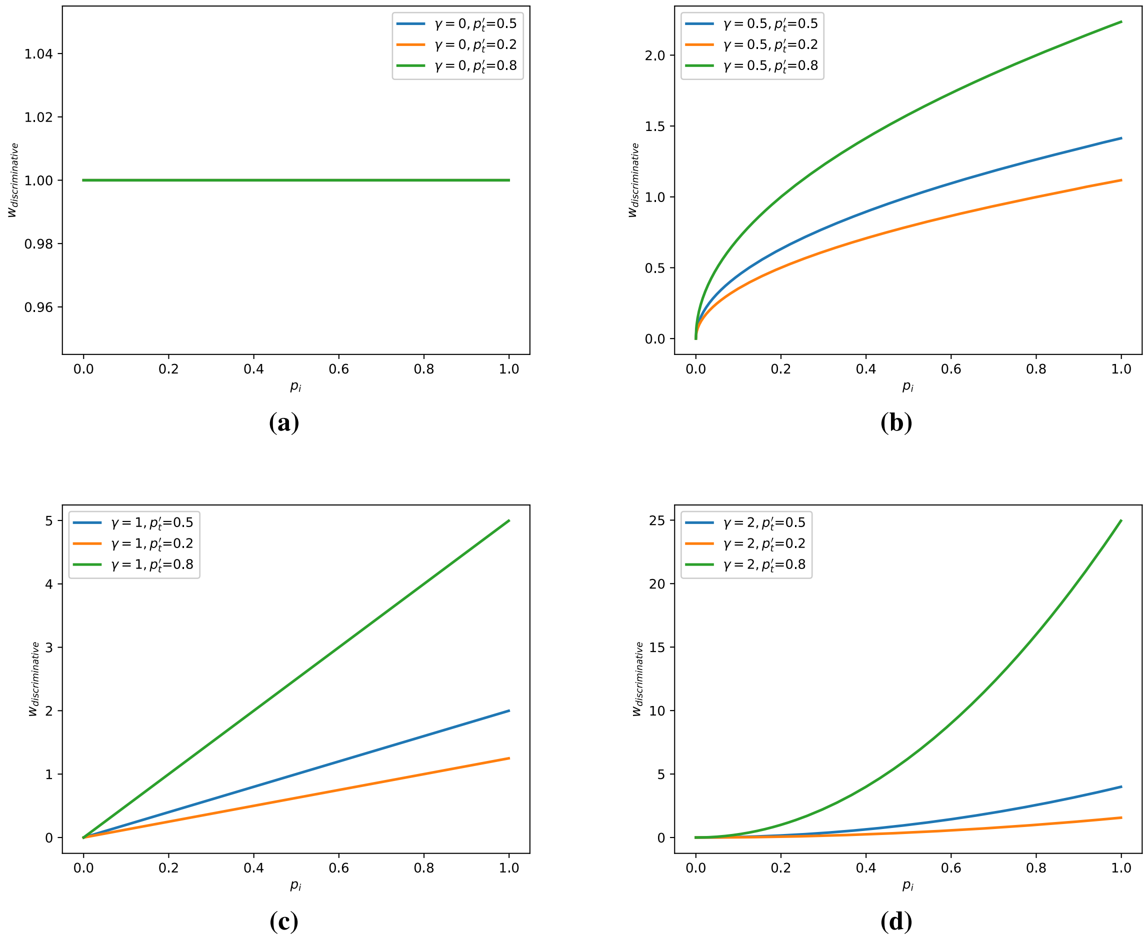
Fig. 2 Illustration of extended focal weight. Inspired by the original focal weight, we proposed the extended focal weight which adaptively calculates the loss weight according to the probability, but further widens the weight difference of the examples. In addition to reducing the loss from easy examples, the loss from hard examples is attached a higher weight

While the two-stage detector and the one-stage detector are structurally different, they are all subject to the class imbalance problems [26]. The two-stage framework typically applies region proposal mechanisms (e.g. Selective Search [37], Edge Boxes [42], RPN [33], DeepMask [28,29]) to screen a large number of candidate target locations in the first stage. In the second stage, sampling methods like fixed ratio of foreground to background [33] or online hard example mining (OHEM) [35] are used to obtain a reasonable balance between foreground and background. But for one-stage detectors, solving the class balance problem is a bigger challenge because it regularly adopts dense sampling of target locations, aspect ratios, and scales and all the possible candidates need to be learned during training. To improve the training efficiency, techniques like data enhancement [31], hard examples mining [35,38], and loss function design [2,16,21] have been proposed. The recent work focal loss [21] has received rising spotlight. It tried to reduce the class imbalance by modifying the sigmoid cross-entropy loss to down-weight the loss assigned to the easy negative examples.

However, in practical applications, such as automatic driving, it is necessary to perform multiclass object detection in a wide viewing angle scene. The anchor mechanism will create more extreme imbalance in foreground and background candidates in high-resolution images, increasing the difficulty of balancing the quantity of negative and positive examples. Therefore, the form of focal loss needs a further improvement to accommodate more imbalanced application scenarios. On the other hand, focal loss only considers the balance between foreground and background. There is no use of discriminative information between foreground classes, which is helpful for improving the discrimination of foreground categories.

In this paper, firstly we explore the form of the weighting factor in the focal loss, which we call it focal weight, as shown in Fig. 1. It can be intuitively seen that the loss weight of the current example is determined by the prediction probability. The greater the probability, the smaller the weight of loss. Inspired by that, we propose the extended function form of focal weight as shown in Fig. 2. Compared with the focal weight, it has the same function of adaptively calculating the weight according to the probability, but can further widen the weight difference of the examples. Secondly, we investigate the function form of focal loss. We believe that the original focal loss cannot dig into the constraint relationship between foreground classes in the form of sigmoid cross-entropy loss, so we apply the extended focal weight to softmax cross-entropy loss. Finally, we propose a new loss function called class-discriminative focal loss aiming for achieving the foreground–foreground class imbalance. On the one hand, we reformulate the standard softmax cross-entropy loss to calculate the negative logarithmic loss of the prediction probability for both the ground-truth class and wrong categories. On the other hand, as shown in Fig. 3, we define a weighting factor called discriminative weight in order to adjust the loss of the wrong prediction probability according to its similarity with ground truth. In short, the contributions of our research are as follows:

1. We propose a new form of focal loss, namely extended focal loss, that is capable to further mitigate the extreme class imbalance.
2. We propose the class-discriminative focal loss by introducing the extended focal loss to multi-class classification task as well as reshaping the standard softmax cross-entropy loss, which can improve the discriminability of



(a) **(b)**

(c) **(d)**

Fig. 3 Illustration of discriminative weight. In addition to the parameter γ, it will adjust adaptively based on the prediction probability for both the ground-truth class $p_t{}'$ and wrong categories p_i

foreground categories so as to reduce the foreground–foreground class imbalance.

3. Our proposed loss function can easily surpass the state-of-the-art method, focal loss, by nearly 1.1 mAP with no more cost of training as well as inference time. It is easy to generalize and apply to other detection models.
4. When trained with our proposed loss function, the network can achieve significant performance gains, outperforming other state-of-the-art methods on two major datasets of autopilot detection tasks, KITTI and Berkeley deep drive (BDD).

The rest of the paper is organized as follow. Section 2 reviews related works. Section 3 details the proposed method. Experimental results are given in Sect. 4, and the conclusions are presented in Sect. 5.

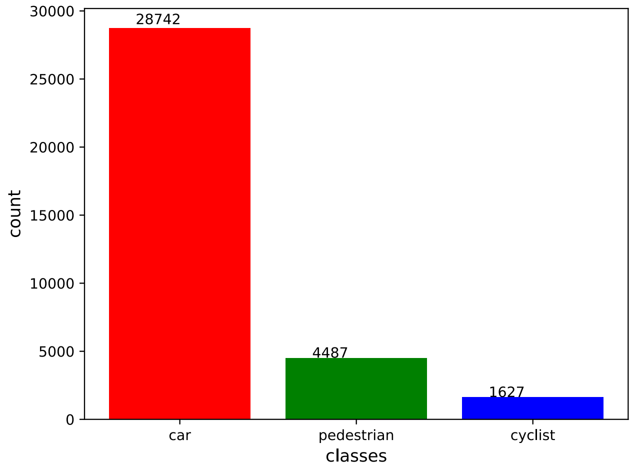


Fig. 4 Histogram of target category statistics in KITTI dataset

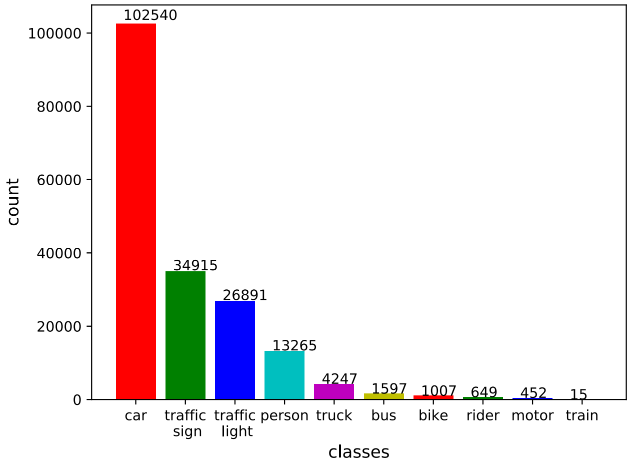


Fig. 5 Histogram of target category statistics in BDD dataset

2 Related works

2.1 CNN-based object detection

Since the application of CNNs, the accuracy of object detection methods have been greatly improved. Especially after the impressive work called AlexNet done by Krizhevsky et al. in 2012 [15], deep neural network has begun to dominate the object detection and other various tasks in computer vision. With the development of neural network structure, the object detection algorithm is also progressing, and it is gradually divided into two main directions: two-stage detectors and one-stage detectors.

The two-stage framework, applied on classical object detection methods, has a long history. The two-stage detector has adopted this framework into CNN architectures. R-CNN [13] was the pioneer to use the CNN as the feature extractor in the first stage following by the support vector machine (SVM) for the classification task in the second stage. After that, Fast R-CNN [12] upgraded the classifier to a convolutional neural network in the second stage largely improving the accuracy. Faster R-CNN [33] creatively proposed the region proposal mechanism making the object detection system an entire neural network structure. Numerous extensions to this structure have been proposed, e.g. [6,18,20].

One-stage detectors typically finish the feature extraction, object localization, and object classification in a convolutional neural network. OverFeat [34] was one of the first one-stage detectors. SSD [10,24] and YOLO [30–32] drew on many ideas such as anchor boxes and feature pyramid from two-stage detectors. The recent work, RetinaNet [21], has received great attention for its elegant architecture and high efficiency.

2.2 Class imbalance

Imbalance problems in object detection have received significant attention, especially class imbalance [26]. For two-stage detectors, owing to the region proposal mechanism [33], this problem was solved more satisfactorily by some common sampling schemes [33,35]. While these sampling heuristic can be applied on one-stage detectors, they are still inefficient due to the domination of the easily classified background examples in the training process [21]. Despite that, kinds of hard negative mining [24,35,38] that excavate the hard examples are proposed to improve the training efficiency. Another influential approach is to modify the loss function. Bulo et al. [2] put forward a loss function called Loss Max-Pooling to eliminate the influence of dataset with long tail distribution on training. Liu et al. [24] integrated the so-called α-balance into the cross-entropy loss to weight the losses of different classes according to their frequency. Lin et al. [21] brought up the focal loss for down-weighting the easy negatives, while the hard examples are unaffected. Weber et al. [39] introduced a focal loss variant called automated focal loss, which can greatly reduce the training convergence time. The above methods hold the opinion that examples of minor classes should have higher losses than those of major classes as the feature learned from the minor classes is poorer. While the focal loss focuses addressing inliers (easy examples), the Huber Loss [9] is designed to reduce the contribution of outliers (hard examples). The recent work Gradient Harmonizing Mechanism [16] also considers the harmfulness of the very hard examples, but it bases on the statistical distribution of the gradient, not the statistical distribution of the loss. Meanwhile, as discussed in [16], the optimal distribution of gradient is unclear. In our work, we also take the idea of reshaping the loss function. However, in addition to reducing the class imbalance, our proposed class-discriminative focal loss is also capable to utilize the interrelationship between foreground classes so as to increase the discriminability of foreground categories and improve the accuracy.

2.3 Objective function design

The loss function of the object detection system usually combines two parts, one for object classification, the other for object location regression. In general, softmax cross-entropy [10,31,34] or sigmoid cross-entropy [21] is adopted for the classification loss. In [21], the function form of focal loss is sigmoid cross-entropy. The work presented in [3] introduced focal loss to softmax cross-entropy and demonstrated that sigmoid cross-entropy is more stabled for training with a variety of aspect ratios and scales, while softmax cross-entropy can get higher performance. We also talk about the difference of them in our work, and our class-discriminative focal loss bases on softmax cross-entropy in consideration of its ability to generate category prediction probability with constraints.

For box regression loss, usually the L_2 loss [30,34], the smooth L_2 loss [24] or the similar smooth L_1 loss [10] are used. The modification of regression loss is not our aim, and we follow the RetinaNet to adopt the smooth L_1 loss.

3 Class-discriminative focal loss

The focal loss introduced by [21] tried to eliminate the training inefficiency caused by the imbalanced data distribution for one-stage detectors. However, while focal loss achieves competitive results on the COCO benchmark [22], it has slightly worse performance on the much more imbalanced dataset like KITTI [11] and BDD [41]. The reason is that in these autopilot datasets, the resolution of the images is higher than those in COCO, which is closer to the practical applications such as automated driving, leading to more extreme imbalance between foreground and background with the anchor mechanism. Besides, as discussed in [3], extending the focal loss to multi-class task works better. When the focal loss is applied on the binary classification, the sigmoid operation is utilized to compute the probability of the targets with the loss computation in the loss layer. But for multi-class classification, the softmax operation is adopted. The former performs in greater numerical stability, while the latter performs in higher accuracy. Based on the above considerations, we extend the form of focal weight to further widen the weight difference of the examples, forming the extended focal loss. In this case, the hard positive examples can get more contributions in the loss, adapting to the extreme imbalanced situations. In addition, we apply the extended focal loss on multi-class classification. In contrast to the previous work [3], we utilize the softmax operation in the loss layer aiming to get the classes prediction probability with constraints. With the help of the constraint category probability, we furthermore propose the class-discriminative focal loss to increase the difference in loss weight between foreground categories, which helps to improve the discriminability of foreground categories, especially similar categories.

To clearly introduce our class-discriminative focal loss, a normal definition of focal loss is required. Focal loss was first applied on sigmoid cross-entropy:

$$\mathrm{CE}_{\mathrm{sigmoid}}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise.} \end{cases} \tag{1}$$

where p is the prediction probability of the class, generated by the sigmoid function:

$$p = \sigma(z) = \frac{1}{1+\exp(-z)} \tag{2}$$

where z is the output of the network. For notational convenience, the probability that the network assigned to the positive example or the negative example can be unified as:

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

and the sigmoid cross-entropy can be simplified as:

$$\mathrm{CE}_{\mathrm{sigmoid}}(p, y) = \mathrm{CE}_{\mathrm{sigmoid}}(p_t) = -\log(p_t) \tag{4}$$

The main contribution of focal loss is the adaptive weight w formulated as Eq. 5:

$$w = (1-p_t)^\gamma \tag{5}$$

In the above, w is determined by two variable, p_t and γ. The former is the probability of the ground truth class estimated by the model and the latter is a modulating parameter. According to [21], since the range of p_t is [0,1], it is used to quantify the classification difficulty of the examples. When p_t is big enough ($p_t \gg 0.5$), the corresponding example is well-classified. In this case, $1-p_t$ is near 0, down-weighting the loss. In contrast, $1-p_t$ is near 1 when p_t is small, keeping the loss for the hard examples unaffected. Besides, the modulating factor γ is used for smooth adjustment. We called w as focal weight and plotted it with $\gamma \in [0, 5]$ as shown in Fig. 1.

In addition to reducing the imbalance between hard examples and easy examples, focal loss also integrates a weighting factor α_t for addressing the class imbalance between negative examples and positive examples:

$$\alpha_t = \begin{cases} \alpha & \text{if } y = 1 \\ 1-\alpha & \text{otherwise.} \end{cases} \tag{6}$$

In the above, α is the weighting factor for positive examples while $1-\alpha$ for negative examples, and α can be set by the inverse class frequency.

Finally, the focal loss can be defined as:

$$\begin{aligned} \mathrm{FL}(p_t) &= -\alpha_t \cdot w \cdot \mathrm{CE}_{\mathrm{sigmoid}}(p_t) \\ &= -\alpha_t (1-p_t)^\gamma \log(p_t) \end{aligned} \tag{7}$$

3.1 Extended sigmoid focal loss

Since the class imbalance in the practical applications is even greater, we try to improve the form of focal weight and propose the extended focal weight w_{extended} as Eq. 8:

$$w_{\mathrm{extended}} = \begin{cases} 1 - \frac{1}{2}\left(\frac{p_t}{1-p_t}\right)^\gamma & \text{if } p_t < 0.5 \\ \frac{1}{2}\left(\frac{1-p_t}{p_t}\right)^\gamma & \text{if } p_t \geq 0.5 \end{cases} \tag{8}$$

As shown in Fig. 2, w_{extended} is a piecewise symmetric function. We assume that the example is easy to classify when the corresponding probability is greater than 0.5, in which the assigned loss weight should be small. (We have also experimented other probabilities like 0.3, 0.4, 0.6, and 0.7, but we found 0.5 to work best in our experiments. Further discussion are presented in Sect. 4.3.6.)

Figure 2 shows the graph of the extended focal weight. Intuitively, the extended focal weight reduces the loss contribution from the well-classified examples just like the focal weight. However, for difficult examples, the extended focal weight endows them a higher loss weight compared to focal weight. In this case, the difference between the hard examples and easy examples is widen to adapt to the more imbalanced situations.

With the extended focal weight, the extended focal loss for binary classification can be formulated as:

$$\begin{aligned}\text{EFL-B}(p_t) &= -\alpha_t \cdot w_{\text{extended}} \cdot \text{CE}_{\text{sigmoid}}(p_t)\\ &= \begin{cases} -\alpha_t \left(1 - \frac{1}{2}\left(\frac{p_t}{1-p_t}\right)^{\gamma}\right) \log(p_t) & \text{if } p_t < 0.5 \\ -\alpha_t \frac{1}{2}\left(\frac{1-p_t}{p_t}\right)^{\gamma} \log(p_t) & \text{if } p_t \geq 0.5 \end{cases}\end{aligned} \tag{9}$$

3.2 Extended softmax focal loss

In order to investigate the relationship between the foreground categories, we introduce the extended focal weight to the softmax cross-entropy loss:

$$\begin{aligned}&\text{CE}_{\text{softmax}}(\mathbf{p}, \mathbf{y}) = -\sum_{i=1} y_i \log(p_i) = -\log(p_t{}')\\ &p_i = p_t{}'(y_i = 1)\end{aligned} \tag{10}$$

In the above, $\mathbf{p}$ is a vector meaning the estimated probability of the network for multiclass prediction and $\mathbf{y}$ is the one-hot ground-truth label. Since $\mathbf{y}$ is one-hot label, we define p_t' for the ground-truth class. The element of $\mathbf{p}$ is p_i, generated by the softmax operation:

$$p_i = \text{softmax}(\mathbf{z}) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{11}$$

Similar with Eq. 6, we define a weighting factor α_t' for rebalancing the loss assigned to foreground and background:

$$\alpha_t{}' = \begin{cases} \alpha' & \text{if } y_i = 1 \\ 1 - \alpha' & \text{otherwise.} \end{cases} \tag{12}$$

But in the above α' is for the ground-truth foreground class while $1-\alpha'$ for the other foreground classes and background.

With the extended focal weight and the softmax cross-entropy as well as the weighting factor α_t', the extended focal loss for multiclass classification can be formulated as:

$$\begin{aligned}\text{EFL-M}(p_t{}') &= -\alpha_t{}' \cdot w_{\text{extended}} \cdot \text{CE}_{\text{softmax}}(p_t{}')\\ &= \begin{cases} -\alpha_t{}' \left(1 - \frac{1}{2}\left(\frac{p_t{}'}{1-p_t{}'}\right)^{\gamma}\right) \log(p_t{}') & \text{if } p_t{}' < 0.5 \\ -\alpha_t{}' \frac{1}{2}\left(\frac{1-p_t{}'}{p_t{}'}\right)^{\gamma} \log(p_t{}') & \text{if } p_t{}' \geq 0.5 \end{cases}\end{aligned} \tag{13}$$

3.3 Class-discriminative focal loss

The traditional softmax cross-entropy loss only calculates the negative logarithmic loss of the ground-truth class $-log(p_t)'$ due to the one-hot label $\mathbf{y}$ ignoring the other prediction probability for the wrong categories $p_i(y_i \neq 1)$. In our opinion, the same p_t' may be generated with different $p_i(y_i \neq 1)$ implying that the similarities between the predicted classes and the ground-truth class are different, which helps to improve the discriminability of the foreground categories. Therefore, we reshape the original softmax cross-entropy to calculate the negative logarithmic loss on the ground-truth class as well as the wrong classes $\sum_i -I\,(y_i \neq 1) \log(1 - p_i)$. Besides, we also define a weighting factor called discriminative weight $w_{\text{discriminative}}$:

$$w_{\text{discriminative}} = \begin{cases} \left(\frac{p_i}{1-p_t{}'}\right)^{\gamma} & \text{if } p_t{}' < 0.5 \\ \left(\frac{p_i}{p_t{}'}\right)^{\gamma} & \text{if } p_t{}' \geq 0.5 \end{cases} \tag{14}$$

In the above, both $\frac{p_i}{1-p_t'}$ and $\frac{p_i}{p_t'}$ quantify the difference between the predicted wrong classes and the ground-truth class. The former calculates the ratio of each wrong predicted probability to the total wrong predicted probability, weighting for the hard examples. The latter calculates the ratio of each wrong predicted probability to the ground-truth probability, weighting for the easy examples. We plotted $w_{\text{discriminative}}$ with $\gamma \in [0, 2]$ as shown in Fig. 3.

With the reshaped softmax cross-entropy and the extended focal weight as well as the discriminative weight, we define our class-discriminative focal loss as Eq. 15:

$$\begin{aligned}\text{CDFL}(p_i, p_t{}') = &-\alpha_t{}' I\,(y_i = 1)\, w_{\text{extended}} \log\left(p_t{}'\right)\\ &-\alpha_t{}' \sum_i \left(I\,(y_i \neq 1)\, w_{\text{discriminative}} \log(1 - p_i)\right)\end{aligned} \tag{15}$$

4 Experiments

To compare with the original focal loss and other rebalance strategies, we choose RetinaNet [21] as the detection network and adopt ResNet-50 as backbone with feature pyramid network (FPN) architecture for ablation study. Furthermore, to better demonstrate the effectiveness of our methods, we conducted horizontal study by improving current state-of-the-art detection networks with our methods. For comprehensive evaluation, mean of average precision (mAP) is reported.

4.1 Datasets

We present experimental results on the challenging detection tasks of KITTI and BDD since these public datasets are collected from real application environment which has an extreme imbalanced distribution.

KITTI consists of 7481 images, containing three object categories of car, pedestrian and cyclist. These object categories are in a great imbalance, in which the ratio of the ground-truth numbers is 17.7:2.7:1, as shown in Fig. 4. Besides, separating pedestrian from cyclist is quite difficult for its similarities. We divide the dataset into two parts, 90% for training and 10% for validation.

BDD has a larger amount of data, where 70K is the training set, 10K is the validation set, and 20K is the test set. Similarly, the target category distribution in the BDD dataset is highly uneven, and it contains ten target categories, more than KITTI. The target quantity distribution is shown in Fig. 5. We only use the 10K validation set for algorithm research since both the training set and validation set of BDD have the same uneven category distributions. In the same way, we divide the dataset into two parts, 90% for training and 10% for validation.

4.2 Implementation details

For the models except YOLOv3 [32] and YOLOv4 [1], we make the implementation based on the Open MMLab Detection Toolbox [4] and make the implementation for YOLOv3 and YOLOv4 based on their pytorch implementation. All studies are trained using the default settings in the original code of each algorithm and adaptively conducted on an NVIDIA GTX 1080Ti.

4.3 Ablation study on KITTI dataset

Comprehensive experiments are conducted on KITTI and BDD dataset, respectively. In this section, we mainly show the ablation study on KITTI dataset for hyperparameter tuning and validating the performance of our proposed methods, since the experiments on BDD have the similar results.

4.3.1 Sigmoid focal loss

We first train the network with the original sigmoid focal loss as the baseline and the results are presented in Table 1. According to [21], the parameter γ cannot be set too large, and the parameter α usually ranges from 0.25 to 0.9. As shown in Table 1, the original focal loss achieved a best mAP of 85.3, in which the AP of car or cyclist is much higher than pedestrian. The parameter setting of $\gamma = 1.0$ and $\alpha = 0.5$ achieved the highest AP of car and cyclist. Besides, the AP of pedestrian and cyclist greatly declined while the AP of car was unaffected when γ was set as 0 and α was set as 0.75. In this case only the α-balance strategy was implemented, indicating the effectiveness of focal loss to improve the detection accuracy of hard examples.

Table 1 Varying γ for FL (w. optimal α)

γ	α	Car	Pedestrian	Cyclist	mAP (%)
1.0	0.25	89.5	78.8	82.4	83.55
2.0	0.25	89.2	77.8	81.8	82.93
1.0	0.5	**89.6**	80.3	**86.2**	**85.34**
2.0	0.5	89.5	79.1	84.6	84.39
1.0	0.75	**89.6**	80.0	86.1	85.22
2.0	0.75	89.4	80.9	85.0	85.08
1.0	0.8	89.5	80.3	85.3	85.04
2.0	0.8	89.3	**81.2**	84.5	85.01
1.0	0.9	89.4	80.6	85.1	85.04
2.0	0.9	89.0	78.6	84.6	84.10
0	0.75	**89.6**	78.3	84.0	83.99

Bold values indicate the best performance

Table 2 Varying γ for EFL-B (w. optimal α)

γ	α	Car	Pedestrian	Cyclist	mAP(%)
1.0	0.25	89.6	78.1	83.6	83.76
2.0	0.25	89.4	76.2	81.2	82.35
1.0	0.5	**89.7**	79.9	83.6	84.39
2.0	0.5	**89.7**	78.8	82.2	83.56
1.0	0.75	89.5	**81.5**	**86.6**	**85.85**
2.0	0.75	89.6	79.7	86.3	85.18
1.0	0.8	89.3	80.9	85.5	85.30
2.0	0.8	89.5	80.0	**86.6**	85.38
1.0	0.9	89.3	80.2	85.3	84.92
2.0	0.9	89.3	80.2	85.6	85.04
0	0.75	89.6	76.3	85.4	83.73

Bold values indicate the best performance

4.3.2 Extended sigmoid focal loss

Results using our extended sigmoid focal loss are shown in Table 2. The extended sigmoid focal loss achieved a best mAP of 85.9 with the parameter setting of $\gamma = 1.0$ and $\alpha = 0.75$. In this case, the APs of pedestrian and cyclist are highest while the AP of car is high enough, showing that hard positive examples have gotten more attention and the loss distribution is more balanced. We can see our extended

Table 3 Varying γ for EFL-M (w. optimal α')

γ	α'	Car	Pedestrian	Cyclist	mAP(%)
1.0	0.25	89.5	78.4	85.0	84.29
2.0	0.25	89.4	76.3	84.0	83.26
1.0	0.5	**89.6**	79.7	85.7	85.03
2.0	0.5	89.5	79.5	85.1	84.70
1.0	0.75	89.5	79.9	86.9	85.44
2.0	0.75	89.5	80.0	86.3	85.25
1.0	0.8	89.4	**80.7**	87.1	**85.72**
2.0	0.8	89.4	78.5	85.3	84.37
1.0	0.9	88.9	79.3	84.9	84.36
2.0	0.9	88.9	**80.7**	86.3	85.29
0	0.75	89.4	76.8	**87.9**	84.71

Bold values indicate the best performance

Table 4 Varying γ for CDFL (w. optimal α')

γ	α'	Car	Pedestrian	Cyclist	mAP(%)
1.0	0.25	–	–	–	–
2.0	0.25	–	–	–	–
1.0	0.5	89.8	78.2	82.4	83.49
2.0	0.5	**90.0**	79.6	86.1	85.23
1.0	0.75	89.9	81.1	87.1	86.06
2.0	0.75	89.9	80.5	86.2	85.54
1.0	0.8	89.6	**82.3**	87.1	**86.35**
2.0	0.8	89.8	79.4	87.3	85.52
1.0	0.9	88.7	81.3	85.3	85.43
2.0	0.9	88.5	80.2	**87.4**	85.69
0	0.75	89.8	81.2	86.9	85.95

Bold values indicate the best performance

sigmoid focal loss has slightly better performance than the original focal loss.

4.3.3 Extended softmax focal loss

Table 3 shows the results using our extended softmax focal loss. For multi-class classification task, the weighting factor α' is only applied on the ground-truth class. We can see the best mAP of the extended softmax focal loss is 85.7 with $\gamma = 1.0$ and $\alpha' = 0.8$, which is higher than that of the original focal loss. When $\gamma = 0$, our loss is equivalent to the softmax cross-entropy with α-balance scheme, which outperforms the sigmoid cross entropy. When $\gamma = 1.0$ and $\alpha' = 0.75$, the performance of the extended softmax focal loss is almost equal to that of the extended sigmoid focal loss. However, a small increase in α' to 0.8 brought a considerable promotion of the mAP, demonstrating the better performance and less numerical stability of softmax cross-entropy. We also found that the best mAP of the extended softmax focal loss is slightly lower than that of the extended sigmoid focal loss. We blame this deficiency on the traditional softmax cross-entropy method that only calculates the negative logarithmic loss of the ground-truth class.

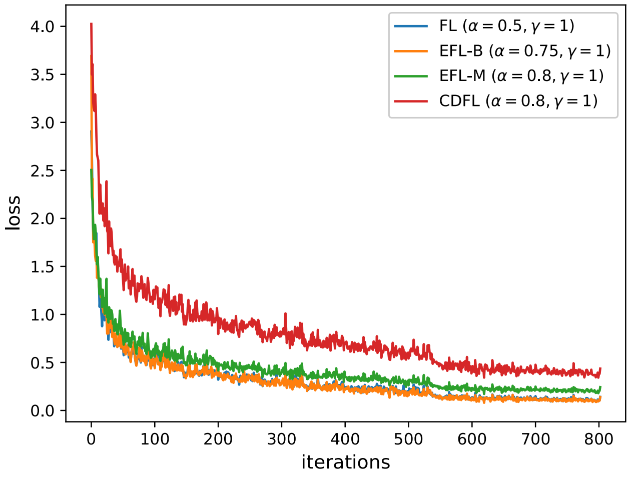


Fig. 6 Loss curves of RetinaNet variants with various focal loss functions. Models trained with original focal loss or our proposed variants of focal loss can converge in the same amount of time

Table 5 Ratio of the loss of negative examples to positive examples

	FL	EFL-B	EFL-M	CDFL
ratio	5.60:1	5.31:1	473.25:1	260.25:1

4.3.4 Class-discriminative focal loss

Results using our class-discriminative focal loss are given in Table 4. The class-discriminative focal loss achieved a best mAP of 86.4 with $\gamma = 1.0$ and $\alpha' = 0.8$, surpassing the original focal loss by 1.1 mAP. The results are better than that of the extended softmax focal loss except the results with $\alpha' = 0.5$ and $\gamma = 1.0$, demonstrating the effectiveness of our approach. When $\alpha' = 0.25$, the model cannot converge due to the excessive imbalanced distribution of the examples. However, we can easily avoid this situation by setting α' based on the inverse class frequency. When $\gamma = 0$ and $\alpha' = 0.75$, the loss function adopting only the α-balance strategy achieved a mAP of 86.0, which outperforms other α-balance variants of sigmoid cross-entropy as well as softmax cross-entropy. Focusing on the best mAP, we found that the AP of pedestrian, the most difficult class to classify, is the highest, and the gap of the APs has been narrowed down. These results show that our class-discriminative focal loss is capable to fully exploit the relationship between foreground classes as well as mitigate the problem of imbalanced data distribution.

Table 6 Experiments based on RetinaNet+CDFL with different threshold

	KITTI (%)	BDD (%)
0.3	86.29	39.97
0.4	86.09	39.18
0.5	**86.35**	**41.72**
0.6	86.27	40.06
0.7	86.11	40.07

Bold values indicate the best performance

4.3.5 Analysis of the various focal loss

For an in-depth understanding of the various focal loss functions, we plotted the loss curves as shown in Fig. 6. We can see that our proposed focal loss variants converge within the same number of iterations as the original focal loss, demonstrating its effectiveness. Besides, since the loss calculation is not required in the inference stage, our method has the same inference speed compared to focal loss. Furthermore, we made statistics on the loss contribution of negative examples and positive examples in the training process and obtained the ratio of the loss of negative examples to positive examples as shown in Table 5. These ratios also reflect the loss proportion of easy examples to hard examples, because most

Table 7 Performance comparison with different rebalance schemes on KITTI validation set

	Backbone	Car	Pedestrian	Cyclist	mAP (%)
Sigmoid + α balance ($\alpha = 0.75$)	ResNet-50-FPN	**89.6**	78.3	84.0	83.99
Softmax + α balance ($\alpha = 0.75$)	ResNet-50-FPN	89.4	76.8	**87.9**	84.71
FL ($\alpha = 0.5, \gamma = 1$) [21]	ResNet-50-FPN	**89.6**	80.3	86.2	85.34
GHM ($M = 30$) [16]	ResNet-50-FPN	89.3	78.6	81.7	83.20
EFL-B ($\alpha = 0.8, \gamma = 1$, ours)	ResNet-50-FPN	89.5	81.5	86.6	85.85
CDFL ($\alpha = 0.8, \gamma = 1$, ours)	ResNet-50-FPN	**89.6**	**82.3**	87.1	**86.35**

Bold values indicate the best performance

Table 8 Performance comparison with different rebalance schemes on BDD validation set

	Car	Bus	Person	Bike	Truck	Motor	Train	Rider	Traffic sign	Traffic light	mAP (%)
FL ($\alpha = 0.75, \gamma = 1$) [21]	70.3	**47.2**	54.5	**38.4**	44.0	19.5	0.00	28.5	55.5	49.0	40.71
GHM ($M = 30$) [16]	**71.0**	40.3	53.6	31.6	44.0	21.6	0.00	32.0	54.8	48.1	39.65
EFL-B ($\alpha = 0.9, \gamma = 1$, ours)	70.6	44.9	54.9	37.5	**46.3**	**25.8**	0.00	30.8	54.9	48.3	41.39
CDFL ($\alpha = 0.8, \gamma = 1$, ours)	70.4	44.7	**55.8**	35.2	44.7	25.5	0.00	**35.4**	**56.5**	**49.1**	**41.72**

Bold values indicate the best performance

Table 9 Performance comparison with other state-of-the-art methods on KITTI validation set

	Backbone	Car	Pedestrian	Cyclist	mAP (%)
Two-stage methods					
Faster R-CNN [33]	ResNet-50	**89.9**	79.4	88.1	85.82
Faster R-CNN + CDFL (ours)	ResNet-50	89.5	**81.2**	**88.6**	**86.44**
One-stage methods					
SSD512 [24]	VGG-16	87.0	51.6	64.2	67.60
YOLOv3 [32]	DarkNet-53	89.9	75.8	75.2	80.28
GHM [16]	ResNet-50-FPN	89.3	78.6	81.7	83.20
RetinaNet [21]	ResNet-50-FPN	89.6	80.3	86.2	85.34
RetinaNet + EFL-B (ours)	ResNet-50-FPN	89.5	81.5	86.6	85.85
RetinaNet + CDFL (ours)	ResNet-50-FPN	89.6	82.3	87.1	86.35
FCOS [36]	ResNet-50-FPN	89.7	79.8	87.0	85.54
FCOS + EFL-B (ours)	ResNet-50-FPN	89.8	82.0	86.6	86.10
YOLOv4m [1]	CSPResNext50	**98.1**	**90.6**	**96.4**	**95.02**
YOLOv4m + EFL-B (ours)	CSPResNext50	**98.1**	89.4	96.0	94.49

Bold values indicate the best performance

Table 10 Performance comparison with other state-of-the-art methods on BDD validation set

	Car	Bus	Person	Bike	Truck	Motor	Train	Rider	Traffic sign	Traffic light	mAP (%)
Two-stage methods											
Faster R-CNN [33]	**72.1**	46.9	**55.3**	36.7	**51.2**	32.0	0.00	**38.5**	**59.0**	**49.7**	44.14
Faster R-CNN + CDFL (ours)	**72.1**	**51.9**	54.7	**37.2**	49.9	**33.9**	0.00	36.4	58.5	48.8	**44.35**
One-stage methods											
SSD512 [24]	58.3	19.5	16.2	10.6	28.4	0.00	0.00	0.00	32.0	24.3	18.94
YOLOv3 [32]	59.2	37.0	35.5	19.9	39.5	7.6	0.00	15.4	34.8	26.6	27.54
GHM [16]	71.0	40.3	53.6	31.6	44.0	21.6	0.00	32.0	54.8	48.1	39.65
RetinaNet [21]	70.3	**47.2**	54.5	**38.4**	44.0	19.5	0.00	28.5	55.5	49.0	40.71
RetinaNet + EFL-B (ours)	70.6	44.9	54.9	37.5	46.3	25.8	0.00	30.8	54.9	48.3	41.39
RetinaNet + CDFL (ours)	70.4	44.7	**55.8**	35.2	44.7	25.5	0.00	**35.4**	56.5	49.1	41.72
FCOS [36]	71.6	45.4	54.5	34.4	47.2	27.4	0.00	28.2	58.2	51.1	41.79
FCOS + EFL-B (ours)	72.0	46.1	54.0	36.1	**47.8**	26.9	0.00	32.7	59.2	**51.9**	42.68
YOLOv4m [1]	73.7	45.0	53.2	30.4	47.2	28.8	0.00	28.4	63.6	59.1	42.90
YOLOv4m + EFL-B (ours)	**73.8**	44.8	54.0	32.4	46.2	**34.9**	0.00	29.1	**64.9**	58.5	**43.98**

Bold values indicate the best performance

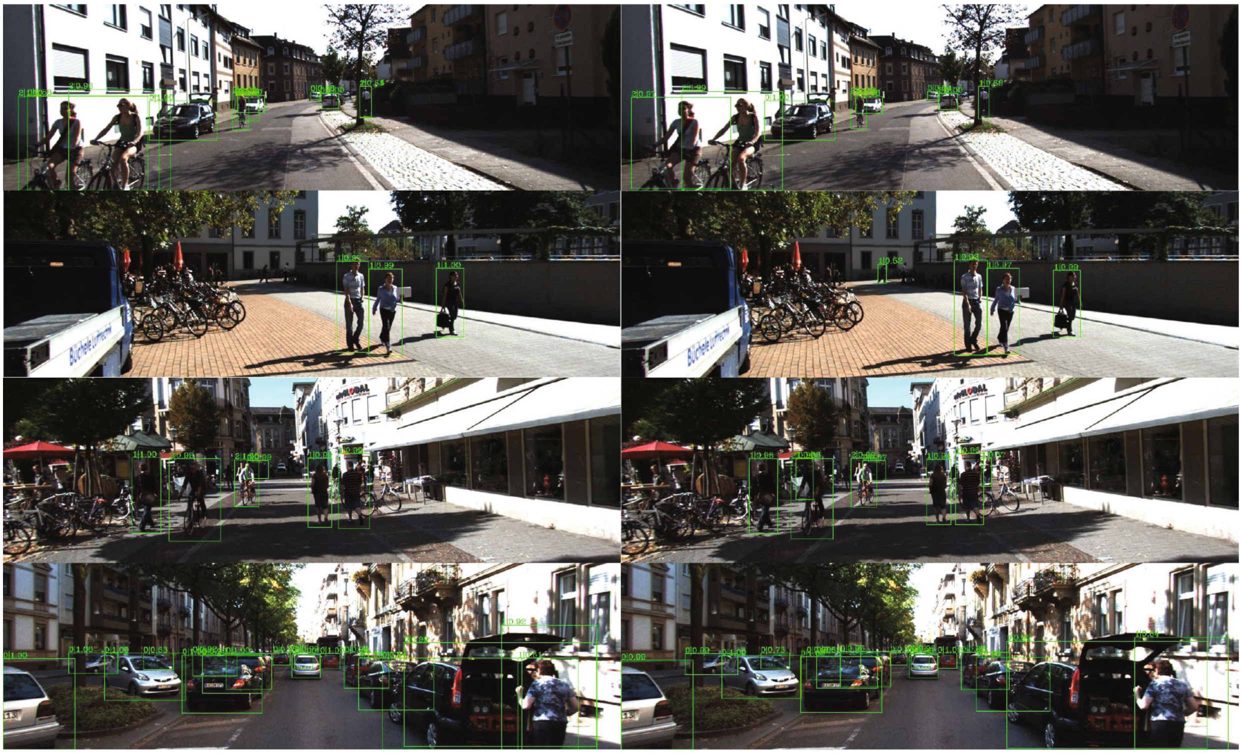


Fig. 7 Detection results of the baseline and proposed algorithms on the KITTI validation set. The first column shows the detection results of RetinaNet, whereas the second column shows the results of RetinaNet+CDFL

of easy examples are negative examples. For sigmoid cross-entropy, the ratio of EFL-B is 5.31:1, slightly smaller than that of FL. For softmax cross-entropy, the ratio of CDFL is 260.25:1, greatly smaller than that of EFL-M. These results confirm that our proposed loss functions, especially CDFL, can better achieved the rebalancing of categories.

4.3.6 Analysis of different threshold setting

According to Focal Loss [21], the threshold of p_t is defined as 0.5 without in-depth analysis. In this paper, we have further explored the impact of different threshold setting.Based on RetinaNet+CDFL, we implemented experiments under different thresholds between 0.3 to 0.7 on both KITTI and BDD

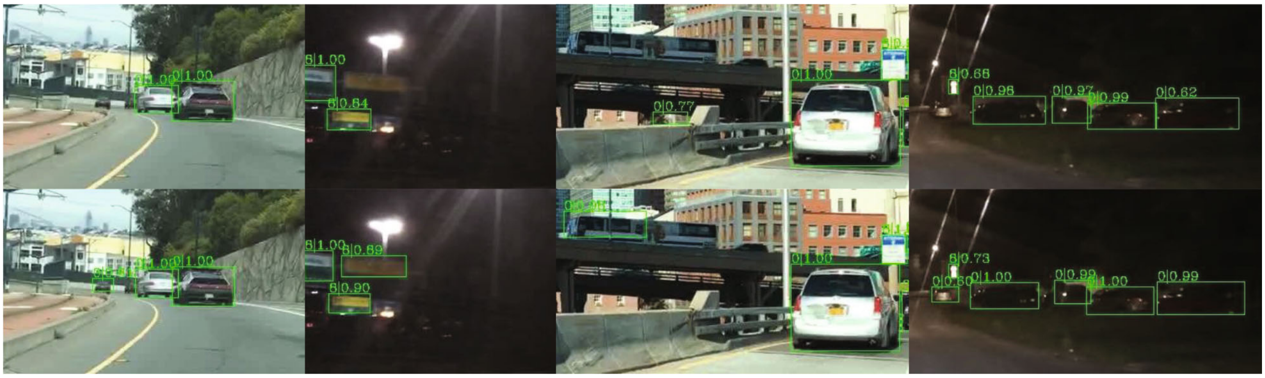


Fig. 8 Detection results of the baseline and proposed algorithms on the BDD validation set. The first row shows the detection results of RetinaNet, whereas the second row shows the results of RetinaNet+CDFL

datasets, and the results are shown as Table 6. The results show that 0.5 is the best threshold, which is consistent with focal loss [21].

4.3.7 Qualitative results

Some qualitative results are shown in Figs. 7 and 8. As shown in the figure, our proposed method can detect more difficult targets that have high similarity such as pedestrian and cyclist. Besides, our method will get more accurate bounding box locations.

4.4 Horizontal study on BDD and KITTI datasets

In this section, experiments are conducted on both KITTI and BDD datasets. We first compare our methods with other rebalance schemes and the results are shown in Tables 7 and 8. The CDFL outperforms all current state-of-the-art methods for weakening the damage of the class imbalance in object detection. In both Tables 7 and 8, it achieves a ∼ 1.1 point mAP gap (86.4 vs. 85.3, 41.72 vs. 40.71) with the closest competitor, focal loss [21]. Compared to GHM [16], we can see a gain of 2–3.2 mAP based on CDFL.

Furthermore, to further verify the effectiveness of our proposed methods, we conduct thorough ablation experiments to compare the proposed mechanisms with current state-of-the-art detectors. Except RetinaNet, we also employ our proposed methods to the main stream two-stage detector Faster R-CNN, one-stage anchor-free detector FCOS and one-stage anchor-based detector YOLOv4. For faster analysis in our ablation experiments, we implement the simplified version of YOLOv4, namely YOLOv4m, which is all the same with YOLOv4 except the model depth and width. The results are shown in Tables 9 and 10. It shows that when trained with the proposed mechanisms, the baseline network can achieve significant performance gains, 1.01/1.01 for RetinaNet, 0.62/0.21 for Faster R-CNN, 0.56/0.89 for FCOS in KITTI/BDD. Although YOLOv4m+EFL-B has a slight performance degradation in KITTI, it has a significant performance improvement in BDD, which has a more serious category imbalance. As shown in Table 11, it is obvious that the mAPs of minor classes such as pedestrian and cyclist in KITTI as well as rider and motor in BDD have a remarkable improvement. These results confirm that our proposed methods, namely EFL-B and CDFL, can effectively improve the performance of main stream one-stage as well as two-stage detectors in imbalance application scenarios.

Table 11 Performance improvement on minor classes in KITTI and BDD validation set compared with each baseline network

	KITTI(%)		BDD(%)	
	Pedestrian	Cyclist	Rider	Motor
Faster R-CNN + CDFL	+1.8	+0.5	− 2.1	+1.9
RetinaNet + EFL-B	+0.8	+0.4	+2.3	+6.3
RetinaNet + CDFL	+2.0	+0.9	+6.9	+6.0
FCOS + EFL-B	+2.2	− 0.4	+4.5	− 0.5
YOLOv4m + EFL-B	− 1.2	− 0.4	+0.7	+6.1

5 Conclusions

In this work, we analyse the limitation of existing rebalance schemes for object detection in consideration of the practical extreme imbalanced scenarios and multi-class classification task. To address this, we propose a extended focal loss to further mitigate the foreground-background class imbalance. Moreover, we propose the class-discriminative focal loss by introducing the extended focal loss to multi-class classification task and reformulating the standard softmax cross-entropy loss, which can improve the discriminability of foreground categories so as to reduce the foreground-foreground class imbalance. Extensive experiments conducted on KITTI and BDD datasets show that our approach can easily surpass the state-of-the-art method, focal loss, with no more training and inference time cost. Besides, our method is easy to generalize and apply to current state-of-the-art one-stage or two-stage object detectors and achieve the best performance.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
2. Bulo, S.R., Neuhold, G., Kontschieder, P.: Loss max-pooling for semantic image segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 7082–7091. IEEE (2017)
3. Chen, C., Song, X., Jiang, S.: Focal loss for region proposal network. In: Pattern Recognition and Computer Vision—First Chinese Conference, pp. 368–380. Springer, Berlin (2018)
4. Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al.: Mmdetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 (2019)
5. Chen, W., Huang, H., Peng, S., Zhou, C., Zhang, C.: Yolo-face: a real-time face detector. Visual Comput., 1–9 (2020)
6. Dai, J., Li, Y., He, K., Sun, J.: R-fcn: Object detection via region-based fully convolutional networks. In: Advances in Neural Information Processing Systems, pp. 379–387 (2016)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893 (2005)
8. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
9. Friedman, J., Hastie, T., Tibshirani, R.: The Elements of Statistical Learning: Springer Series in statistics. Springer, Berlin (2001)
10. Fu, C.Y., Liu, W., Ranga, A., Tyagi, A., Berg, A.C.: Dssd: Deconvolutional single shot detector. arXiv preprint arXiv:1701.06659 (2017)
11. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361. IEEE (2012)
12. Girshick, R.: Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
13. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
14. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1106–1114 (2012)
16. Li, B., Liu, Y., Wang, X.: Gradient harmonized single-stage detector. In: The Thirty-Third AAAI Conference on Artificial Intelligence, vol. 33, pp. 8577–8584 (2019)
17. Li, T., Ye, M., Ding, J.: Discriminative hough context model for object detection. Visual Comput. **30**(1), 59–69 (2014)
18. Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., Sun, J.: Light-head R-CNN: In defense of two-stage object detector. arXiv preprint arXiv:1711.07264 (2017)
19. Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: Proceedings of the 2002 International Conference on Image Processing, pp. 900–903. IEEE (2002)
20. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 936–944 (2017)
21. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: IEEE International Conference on Computer Vision, pp. 2999–3007 (2017)
22. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: 13th European Conference on Computer Vision, pp. 740–755. Springer, Berlin (2014)
23. Liu, B., Wu, H., Su, W., Zhang, W., Sun, J.: Rotation-invariant object detection using sector-ring hog and boosted random ferns. Visual Comput. **34**(5), 707–719 (2018)
24. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single shot multibox detector. In: 14th European Conference on Computer Vision, pp. 21–37. Springer, Berlin (2016)
25. Ojala, T., Pietikäinen, M., Mäenpää, T.: Gray scale and rotation invariant texture classification with local binary patterns. In: 6th European Conference on Computer Vision, pp. 404–420. Springer, Berlin (2000)
26. Oksuz, K., Cam, B.C., Kalkan, S., Akbas, E.: Imbalance problems in object detection: a review. IEEE Trans. Pattern Anal. Mach. Intell. (2020)
27. Papageorgiou, C.P., Oren, M., Poggio, T.: A general framework for object detection. In: Proceedings of the Sixth International Conference on Computer Vision, pp. 555–562. IEEE (1998)
28. Pinheiro, P.O., Collobert, R., Dollár, P.: Learning to segment object candidates. In: Advances in Neural Information Processing Systems, pp. 1990–1998 (2015)
29. Pinheiro, P.O., Lin, T.Y., Collobert, R., Dollár, P.: Learning to refine object segments. In: 14th European Conference on Computer Vision, pp. 75–91. Springer, Berlin (2016)
30. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
31. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517–6525 (2017)
32. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
33. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
34. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229 (2013)
35. Shrivastava, A., Gupta, A., Girshick, R.: Training region-based object detectors with online hard example mining. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 761–769 (2016)
36. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE international conference on computer vision, pp. 9627–9636 (2019)

37. Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. Int. J. Comput. Vis. **104**(2), 154–171 (2013)
38. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
39. Weber, M., Fürst, M., Zöllner, J.M.: Automated focal loss for image based object detection. arXiv preprint arXiv:1904.09048 (2019)
40. Wei, L., Cui, W., Hu, Z., Sun, H., Hou, S.: A single-shot multi-level feature reused neural network for object detection. Visual Comput. 1–10 (2020)
41. Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2636–2645 (2020)
42. Zitnick, C.L., Dollár, P.: Edge boxes: locating object proposals from edges. In: 13th European Conference on Computer Vision, pp. 391–405. Springer, Berlin (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Guancheng Chen received his B.Eng. degree in Communication Engineering and M.Eng. degree in Electronic and Communication Engineering from South China Normal University, Guangzhou, China in 2016 and 2018, respectively. He is currently pursuing his Ph.D. degree in Information and Communication Engineering at South China University of Technology, Guangzhou, China. His research interests lie in the areas of computer vision and pattern recognition.

Huabiao Qin Ph.D. graduated from School of Electronics and Information, South China University of Technology, professor of South China University of Technology, Ph.D. supervisor, director of Intelligent information processing, wireless communication network and embedded system.