2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Semi-Supervised Object Detection via Dynamic Reweighting of Localization Error

Huajie Xu[1,2] (✉) and Ganxiao Nong[1]

[1] College of Computer and Electronic Information, Guangxi University, Nanning 530004, China
[2] Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China

Abstract. Semi-supervised object detection (SSOD) leverages limited labeled data alongside abundant unlabeled data to improve detection performance. Existing SSOD methods based on teacher-student framework tend to neglect localization error within pseudo-labels, detrimentally affecting the student model's bounding box regression and classification. To address this issue, a novel SSOD method based on dynamic localization error reweighting is proposed. In the method, predicted bounding boxes are modeled using Gaussian distribution to derive a localization quality score quantifying localization error. This score underpins a strategy of Localization Error reweighting in Regression (LER), which dynamically adjusts the unsupervised regression loss to prioritize accurately localized pseudo-labels. Simultaneously, a strategy of Proposal Reliability reweighting in Classification (PRC) is proposed, utilizing teacher predictions to assess student proposal reliability. PRC combines class probabilities and localization quality scores to dynamically reweight the unsupervised classification loss, thereby mitigating interference from misassigned labels. Extensive experiments on the MS COCO and PASCAL VOC datasets demonstrate the effectiveness and superiority of our approach.

Keywords: Semi-supervised object detection, pseudo-label, localization error, loss reweighting.

1 Introduction

In recent years, object detection methods based on deep learning have developed rapidly, achieving significant performance improvements through supervised learning on large-scale annotated datasets. However, accurately labeling such datasets is both time-consuming and costly. In contrast, collecting unlabeled data is easier and less costly. As a result, semi-supervised object detection (SSOD) methods have received increasing attention from researchers. By leveraging a small amount of labeled data alongside a large volume of unlabeled data, SSOD enhances model performance while reducing reliance on labeled data.

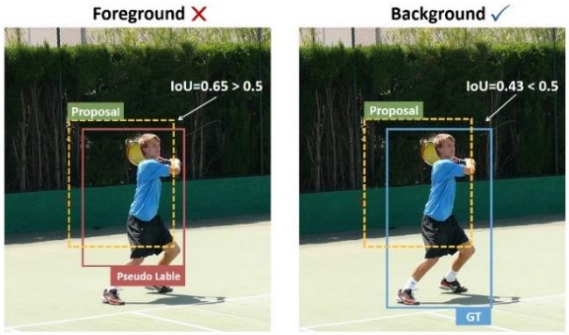

Fig. 1. An example to demonstrate that noisy pseudo-labels can mislead the label assignment. The left figure shows the assignment using noisy pseudo-labels, while the right figure shows the assignment using ground-truth.

At present, most of the SSOD methods adopt the teacher–student framework [1], the teacher model generates pseudo-labels for unlabeled data, while the student model is trained jointly using labeled data and unlabeled data with pseudo-labels. During training, the teacher model's parameters are gradually updated using the exponential moving average (EMA) of the student model's parameters to ensure stable pseudo-label generation. Due to the limited amount of labeled data, the teacher model may still generate a number of inaccurate pseudo-labels. To ensure the quality of pseudo-labels, existing SSOD methods [2-4] follow the practice in semi-supervised image classification (SSIC) [5], selecting predicted bounding boxes with foreground scores above a certain threshold as pseudo-labels. However, unlike SSIC, pseudo-labels in SSOD consist of both category label and bounding boxes. Although category labels can be ensured to be accurate via setting a high foreground scores threshold, the localization quality of pseudo-label fails to be measured and guaranteed. Incorporating pseudo-labels with localization errors (i.e., noisy pseudo-labels) into training hinders model optimization, ultimately limiting SSOD performance gains. Specifically, most existing SSOD methods employ Faster R-CNN [6] as the object detector for both the teacher and student models, which involves both bounding box regression task and classification task. For the bounding box regression task, noisy pseudo-labels can mislead the model into learning inaccurate localization information, causing localization errors to accumulate during training. For the classification task, since the IoU-based label assignment strategy adopted by Faster R-CNN relies on localization information, noisy pseudo-labels can mislead label assignment. As shown in Fig. 1, when the IoU between a proposal and a noisy pseudo-label exceeds the threshold (0.5), a proposal that actually belongs to the background is mistakenly assigned as foreground.

To address aforementioned issues, we propose a novel SSOD approach based on dynamic reweighting of localization errors. First, we model predicted bounding boxes as Gaussian distributions to derive a new metric named localization quality score that quantifies localization error. Then, to mitigate the adverse effects of localization errors in pseudo-labels on both bounding-box regression and classification tasks, we employ the localization quality score to guide the model training in these two tasks, so as to

propose a strategy of Localization Error reweighting in Regression (LER) and a strategy of Proposal Reliability reweighting in Classification (PRC). In LER, by introducing a localization-aware branch into the object detector and trains it using a KL divergence-based regression loss, the key information for calculating the localization quality score can be obtained. The localization quality score of the pseudo-label is then used to dynamically reweight the unsupervised regression loss, enabling the model to focused on pseudo-labels with higher localization quality during training. In RPC, since low-quality proposals are more likely to be misled by noisy pseudo-labels, we use the more stable teacher model to evaluate the quality of the proposals generated by the student model. The reliability scores for both foreground and background proposals are computed by combining the teacher's predicted class probabilities with the localization quality score, and then used to dynamically reweight the unsupervised classification loss, ensuring that more reliable proposals contribute more effectively to classification learning. The main contributions of this paper are as follows:

1. We analyze the negative effects brought by noisy pseudo-labels on bounding box regression and classification tasks, and propose a localization quality score that effectively quantifies localization errors.

2. We propose LER and PRC strategies, which take localization errors into account during the trainings of teacher and student models. These two strategies alleviate the negative impact of pseudo-label noise on model training and make better use of unlabeled data to enhance model performance.

3. Extensive experiments on the object detection benchmark datasets MS COCO [7] and PASCAL VOC [8] demonstrate that our approach achieves significant performance gains compared with the supervised baseline, and surpasses the mainstream state-of-the-art SSOD methods in detection accuracy.

2 Related works

2.1 Semi-supervised image classification

In the field of computer vision, semi-supervised learning has been widely applied to image classification tasks, existing methods can be categorized into two groups: consistency regularization based methods and pseudo-label based methods [9]. Recent studies [10-12] have combined consistency regularization with pseudo-labeling, significantly improving the performance of SSIC. These methods generate pseudo-labels on weak augmented unlabeled images, and then train the model on strong augmented unlabeled images to ensure that the model has consistent predictions for different augmentations. Although SSIC has made significant progress, applying semi-supervised learning to the more complex task of object detection remains challenging. Therefore, this paper focuses on the problem of SSOD.

2.2 Semi-supervised object detection

The current mainstream SSOD methods draw on the success of SSIC. STAC [1] first proposed the SSOD framework that combines pseudo-label and data augmentation,

using a pre-trained teacher model to generate pseudo-labels offline for unlabeled data. Subsequent studies were inspired by Mean Teacher [13], utilized the EMA mechanism to continuously update the teacher model and generate pseudo-labels online after each training iteration, achieving an end-to-end framework. For instance, Instant-teaching [14] and ISMT [15] improved the quality of pseudo-labels by aggregating the model's predictions under the EMA mechanism. CST [16] proposed a cyclic self-training method to overcome the coupling between the teacher and student models caused by EMA. Additionally, RPL [17] and USD [18] dynamically adjusted the thresholds used to generate pseudo-labels. MUM [3], Robust Teacher [19] and Elaborate Teacher [20] introduced new data augmentation strategies to better exploit information in unlabeled images. However, most of these methods overlook the impact of localization errors in pseudo-labels. Therefore, Unbiased Teacher [21] directly discards the regression loss calculation of unlabeled data, but it wastes potentially valuable localization information. Soft Teacher [2] employs a bounding box jittering technique to select reliable pseudo-labels for regression, but this approach requires tuning several hyperparameters. In contrast, our method dynamically adjusts the contribution of different samples in training via loss reweighting. This ensures the retention of effective localization information while avoiding the introduction of additional hyperparameters.

3 Method

3.1 Overview

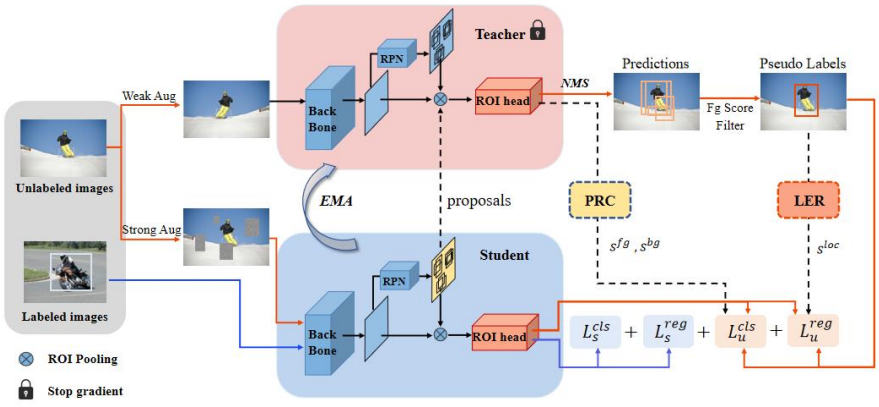

Fig. 2. The overall framework of our method. Both the teacher and the student models use Faster R-CNN as the detector, which consists of a backbone network, a region proposal network (RPN), and a region of interest (ROI) head.

Following the Soft Teacher [2], our method adopts the teacher–student framework which is trained in an end-to-end manner, which is shown as Fig. 2. In each training

iteration, we randomly sample a fixed ratio of labeled and unlabeled images from the dataset to form the training batch. Subsequently, the labeled images are fed into the student model for training, and the supervised loss is computed using the ground-truth labels:

$$L_s = L_s^{cls} + L_s^{reg} \tag{1}$$

where L_s^{cls} and L_s^{reg} denote the supervised classification loss and supervised regression loss, respectively.

Meanwhile, the unlabeled images are processed through strong and weak data augmentation, respectively. The weak augmented images are fed into the teacher model to generate predicted bounding box, which are then filtered using Non-Maximum Suppression (NMS) and a foreground score threshold to produce pseudo-labels; the strong augmented images are sent to the student model for training, and the pseudo-labels generated by the teacher are used to compute the unsupervised loss:

$$L_u = L_u^{cls} + L_u^{reg} \tag{2}$$

In this process, the LER strategy computes a localization quality score s^{loc} for each pseudo-label and uses it to reweight the unsupervised regression loss L_u^{reg}, enhancing the contribution of accurately localized pseudo-labels to the bounding box regression task. Then, proposals generated by the RPN of student model are fed to the ROI head of teacher model for prediction. The PRC strategy computes the reliability scores s^{fg} and s^{bg} for foreground and background proposals, respectively, and uses them to reweight the unsupervised classification loss L_u^{cls}, reducing the negative impact of low-quality proposals on classification.

Finally, the student model's parameters are updated via backpropagation, while the teacher model's parameters are updated using an EMA of the student model's parameters. The overall loss is defined as a weighted sum of the supervised loss and the unsupervised loss:

$$L = L_s + \beta L_u \tag{3}$$

where β is the weight of the unsupervised loss, which controls its contribution to the overall training process.

3.2 Localization quality score

Existing SSOD methods retain the predicted bounding boxes with a foreground score higher than the threshold as pseudo-labels. However, as shown in Fig. 3(a)**,** our visualization experiment reveals the lack of correlation between the foreground score and the localization quality, which implies that this selection criterion may result in pseudo-labels with significant localization errors. To address this issue, we propose a new metric to assess the localization error of predicted bounding boxes, i.e. localization quality score.

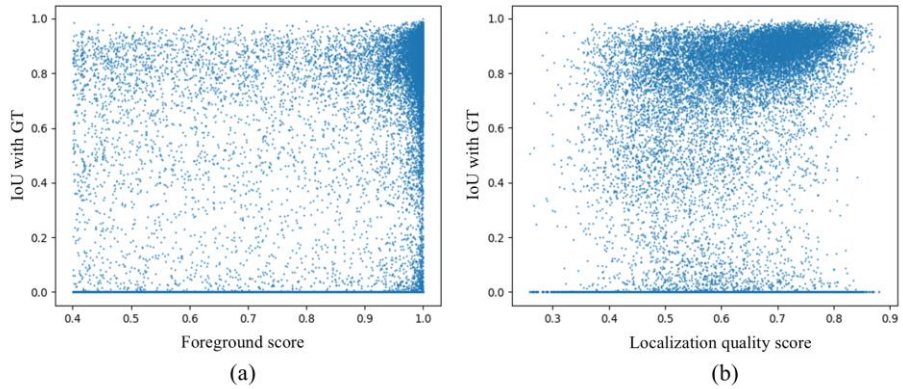

Fig. 3 Visualization of the correlation between IoU with ground-truth (GT) and different criteria. Each data point represents a predicted bounding box, where a higher IoU indicates better localization quality. (a) the correlation between the IoU with ground-truth and foreground score. (b) the correlation between the IoU with ground-truth and localization quality score.

Specifically, in the Faster R-CNN detector adopted in this work, the predicted bounding box coordinates are modeled as a simple Dirac delta distribution [22], which only represents the exact position and fails to reflect localization error. Therefore, we instead model the bounding box coordinates as Gaussian distributions to extract information indicative of localization error. For computational simplicity, we assume independence among coordinates and adopt univariate Gaussian distributions, which are defined as follows:

$$G_\Theta(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-t^e)^2}{2\sigma^2}} \tag{4}$$

where $t \in \{x_1, y_1, x_2, y_2\}$ denote the four bounding box coordinates (top-left and bottom-right). Each coordinate is optimized individually as a random variable of the Gaussian distribution. Θ is the set of learnable parameters of the model. t^e represents the predicted coordinates of the bounding box, which serves as the mean of the Gaussian distribution.σ^2 denotes the coordinate variance (as detailed in Section 3.3), which serves as the variance of the Gaussian distribution.

After modeling the predicted bounding box as a Gaussian distribution, its coordinates are transformed from definite values to probability representations, with larger σ^2 indicates lower localization reliability and greater error. Therefore, to assess the overall localization error of a predicted bounding box, we compute the average variance across its four coordinates and apply sigmoid normalization to obtain the localization quality score:

$$s^{loc} = 1 - Sigmoid(\frac{\sum_{k=1}^{4} \sigma_k^2}{4}) \tag{5}$$

As shown in Fig. 3(b), compared with the foreground score, the proposed localization quality score demonstrates a stronger correlation with the actual localization quality, validating its effectiveness in measuring localization error.

3.3 Localization Error reweighting in Regression

Incorporating pseudo-labels with localization errors into training can hinder the optimization of the bounding box regression task, resulting in decreased localization accuracy. To address this issue, we propose a strategy of Localization Error reweighting in Regression (LER), which utilize the proposed localization quality score to assess the localization error of pseudo-labels and guide the student's model training.

Specifically, the calculation of the localization quality score depends on the coordinate variance. Therefore, we introduce a localization-aware branch in the ROI head of Faster R-CNN, designed to predict coordinate variances, as depicted in Fig. 4. To jointly train the original localization branch and the newly added localization-aware branch, allowing the model to simultaneously estimate bounding box coordinates and coordinate variances, we reformulate the regression loss based on a Gaussian distribution.

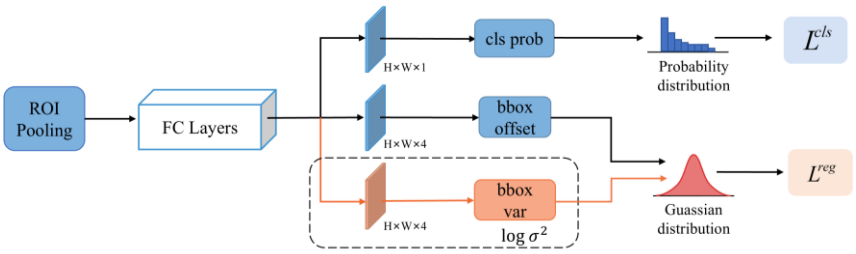

Fig. 4 The structure of improved ROI head. The dashed part represents the localization-aware branch, which is constructed by a fully connected layer. To avoid gradient explosion during training, this branch predicts $log(\sigma^2)$ instead of σ^2 directly.

First, similar to the predicted bounding boxes, the ground-truth box coordinates t^g are also modeled as Gaussian distributions. In this case, as $\sigma^2 \to 0$, the Gaussian distribution degenerates into a Dirac delta distribution:

$$D(t) = \delta(t - t^g) \tag{6}$$

Then, the objective of the bounding box regression task is defined as learning a set of parameters $\hat{\theta}$ that minimizes the distance between the predicted bounding box distribution $G_\theta(t)$, as defined in Eq. (4), and the ground-truth box distribution $D(t)$ over N samples:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum \mathcal{M}_{KL}(G_\theta(t) \| D(t)) \tag{7}$$

Finally, we adopt the KL divergence as the metric for distribution distance to calculate the regression loss, which is called the KL regression loss:

$$L_{KL} = \mathcal{M}_{KL}\big(G_\Theta(t) \parallel D(t)\big) = e^{-log\,(\sigma^2)}\frac{(t^g - t^e)^2}{2} + \frac{1}{2}log\,(\sigma^2) \tag{8}$$

To reduce the interference of outliers on backpropagation and improve training stability, when $|t^l - t^e| > 1$, the loss is smoothed in a manner similar to the Smooth L1 loss:

$$L_{KL} = e^{-log(\sigma^2)}\left(|t^g - t^e| - \frac{1}{2}\right) + \frac{1}{2}log(\sigma^2) \tag{9}$$

When the error between t^e and t^g is large, increasing σ^2 can reduce L_{KL}. Therefore, we replace the traditional Smooth L1 regression loss in Faster R-CNN with the KL regression loss, aiming to encourage the model to output a larger σ^2 when the localization prediction is inaccurate by minimizing L_{KL}. This ensures that the localization quality score calculated based on σ^2 can effectively reflect the localization error of the predicted bounding box.

In the teacher-student framework, the student model is jointly trained with labeled and unlabeled images. Since the noisy pseudo-labels on the unlabeled images can interfere with the training of the localization-aware branch, we only calculate L_{KL} when training the student model with labeled images. Therefore, the supervised regression loss is expressed as:

$$L_s^{reg} = \frac{1}{N_s^{fg}}\sum_{i=1}^{N_s^{fg}} L_{KL}(t_i^e, t_i^g) \tag{10}$$

where N_s^{fg} is the number of foreground proposals generated by the student model on labeled images, t_i^e and t_i^g represent the predicted bounding box coordinates and the corresponding ground-truth box coordinates of the *i*-th foreground proposal, respectively.

To help the model focus on pseudo-labels with high localization quality during training, the teacher model predicts the coordinate variance of each pseudo-label through the localization-aware branch. The resulting localization quality score, is then used as the weight for the unsupervised regression loss:

$$L_u^{reg} = \frac{1}{N_u^{fg}}\sum_{j=1}^{N_u^{fg}} s_j^{loc} L_{smoothL1}\,(\hat{t}_j^e, \hat{t}_j^u) \tag{11}$$

where N_u^{fg} is the number of foreground proposals generated by the student model on unlabeled images. $\hat{t}_j^e$ and $\hat{t}_j^u$ represent the predicted bounding box coordinates and the corresponding pseudo-label coordinates of the *j*-th foreground proposal, respectively. s_j^{loc} denotes the localization quality score of the pseudo-label. $L_{smoothL1}$ refers to the Smooth L1 loss adopted by default in Faster R-CNN.

By reweighting the loss, the contribution of each pseudo-label is dynamically adjusted during training according to its localization quality. This guides the model to learn reliable localization information from noisy pseudo-labels and effectively enhances localization accuracy.

3.4 Proposal Reliability reweighting in Classification

The Faster R-CNN adopted in this work uses an IoU-based label assignment strategy, where a proposal is assigned as foreground if its IoU with a ground-truth box exceeds a predefined threshold; otherwise, it is treated as background. However, when training with pseudo-labels that contain localization errors (i.e., noisy pseudo-labels), the deviation between the pseudo-labels and the ground-truth boxes can lead to incorrect assignments. These misassigned proposals will confuse the decision boundary between foreground and background, ultimately degrading the model's classification accuracy.

To mitigate this issue, we propose a strategy of Proposal Reliability weighting in Classification (PRC). Since low-quality proposals are more susceptible to be misled by noisy pseudo-labels, we design a reliability score to assess the quality of each proposal and use it to reweight the unsupervised classification loss. Specifically, the proposals generated by the student model within the RPN are fed into the ROI head of the teacher model for re-prediction. For the *i*-th foreground proposal, the foreground class probability $p_i^{fg}(T)$ predicted by the teacher model is used to estimate the likelihood of being a real foreground; the predicted coordinate variance is then used to compute the localization quality score s_i^{loc}, as defined in Eq. (5), which evaluate its localization error. Thus, the reliability score for a foreground proposal is defined as:

$$s_i^{fg} = max(p_i^{fg}(T)) \cdot s_i^{loc} \tag{12}$$

For the *j*-th background proposal, we directly use the background probability $p_j^{bg}(T)$ predicted by the teacher model as its reliability score, which reflects the likelihood of being a real background. It is defined as follows:

$$s_j^{bg} = p_j^{bg}(T) \tag{13}$$

It should be noted that the unlabeled images fed into the teacher and the student model undergo different data augmentations. Geometric transformations such as flipping, scaling, and translation may cause spatial misalignment between the proposals generated by the two models. Therefore, before feeding the student model's proposals into the teacher model, we apply the corresponding inverse transformations to align them spatially.

To mitigate the adverse effects of incorrectly assigned low-quality proposals during training, we reweight the unsupervised classification loss using the proposal reliability scores:

$$L_u^{cls} = \frac{1}{N_u^{fg}} \sum_{i=1}^{N_u^{fg}} s_i^{fg} L_{CE}(p_i^{fg}(S), \hat{c}_i^{fg}) + \frac{1}{N_u^{bg}} \sum_{j=1}^{N_u^{bg}} s_j^{bg} L_{CE}(p_j^{bg}(S), \hat{c}_j^{bg}) \quad (14)$$

where N_u^{fg} and N_u^{bg} denote the numbers of foreground and background proposals generated by student model on unlabeled images, respectively. L_{CE} represents the cross-entropy loss. $p_i^{\mathrm{fg}}(S)$ is the foreground probability predicted by the student model, $\hat{c}_i^{fg}$ is the pseudo-label category. Similarly, $p_j^{\mathrm{bg}}(S)$, $\hat{c}_j^{bg}$ and s_j^{bg} denote the background predicted probability, pseudo-label, and reliability score for the background proposal, respectively. As for the supervised classification loss L_s^{cls}, the default classification loss of Faster R-CNN is directly used.

Compared with the student model, the teacher model updated by EMA has stronger generalization ability, resulting in more stable predictions. Therefore, using the reliability scores provided by the teacher model as loss weights enhances the contribution of reliable proposals to the classification task, which leads to improved classification performance.

4 Experiments

4.1 Dataset and metrics

To evaluate the effectiveness of the proposed method, we conduct experiments on two widely used benchmarks datasets MS COCO [7] and PASCAL VOC [8]. MS COCO dataset contains 80 object categories, with 118k images in the train2017 set, 123k unlabeled images in the unlabeled2017 set, and 5k images in the val2017 set. The PASCAL VOC dataset includes 20 categories, with 5,011 images in VOC07 trainval set, 11,540 in VOC12 trainval set, and 4,952 in VOC07 test set.

The model performance is measured by mean Average Precision (mAP) which is the standard metric in object detection. Specifically, we report AP_{50} (IoU threshold of 0.5), AP_{75} (IoU threshold of 0.75), and $AP_{50:95}$ (averaged over IoU thresholds from 0.5 to 0.95 in steps of 0.05). The larger the IoU threshold, the stricter the requirement for the predicted localization. Thus, AP_{50} mainly reflects classification performance, AP_{75} better reflects the localization accuracy, and $AP_{50:95}$ provides a comprehensive measure of overall performance.

4.2 Implementation details

All experiments were conducted on a hardware environment equipped with two NVIDIA RTX 3090 GPUs and implemented using the MMDetection toolbox. For fair comparison, we adopt Faster R-CNN with a Feature Pyramid Network (FPN) as the base detector and use ResNet-50 as the backbone network. Following the experimental settings in related works [2][20][21], for the MS COCO datasets, we set the batch size to 10 per GPU, with a labeled to unlabeled image ratio of 1:4. The unsupervised loss

weight β is set to 4, and the model is trained for 180k iterations. For the PASCAL VOC dataset, we set the batch size to 8 per GPU, with a labeled to unlabeled image ratio of 1:1. The unsupervised loss weight β is set to 2, and the model is trained for 72k iterations. The data augmentation strategy is consistent with those used in related work.

4.3 Performance Comparisons

To evaluate the performance of our method, we compare it with Faster R-CNN trained on limited labeled data only (supervised baseline) and several mainstream state-of-the-art SSOD methods. All the compared methods are implemented within the teacher-student framework and use Faster R-CNN as the detector. Following the practices of most SSOD methods, we evaluate our method under three experimental settings: COCO-standard, COCO-addition and PASCAL VOC. The experimental results are presented in Table 1 and Table 2.

Table 1. Comparison results on COCO-standard and COCO-addition. The $AP_{50:95}$ (%) is used as the evaluation metric. Bold numbers indicate the best performance. — means that the results are missing in the source paper. * denotes the results we reproduced.

Methods	COCO-standard			COCO-addition
	1%	5%	10%	100%
Supervised [1]	9.05	18.47	23.86	37.63
STAC [1]	13.97	24.38	28.64	39.21
Instant-Teaching [14]	18.05	26.75	30.40	40.20
ISMT [15]	18.88	26.37	30.53	39.64
Unbiased Teacher [21]	20.75	28.27	31.50	41.30
Soft Teacher* [2]	19.21	29.70	32.08	42.50
MUM [3]	21.88	28.52	31.87	42.11
CST [16]	22.20	29.75	32.65	42.05
RPL [17]	19.02	28.40	32.23	41.00
Elaborate Teacher [20]	22.65	30.05	32.90	—
Ours	20.10	**30.70**	**33.52**	**43.20**

COCO-standard. We randomly sample 1%, 5%, and 10% of the COCO train2017 set as labeled data, using the remaining images as unlabeled data. The method's performance is evaluated on the COCO val2017 set. As shown in Table 1, our method achieves improvements of 11.05%, 12.23%, and 9.66% in $AP_{50:95}$ over the supervised baseline at 1%, 5%, and 10% labeling ratios, respectively, demonstrating significant performance gains. Compared with other state-of-the-art SSOD methods, the proposed method achieves the best performance on the 5% and 10% labeled ratios. At the 1% labeling ratios, our method underperforms some SSOD methods. This is likely because, to effectively assess the localization error, we train the localization-aware branch on the labeled images only. As a result, the model fails to fully capture the localization information in the case of extremely scarce labeled data. Nevertheless, when the proportion of labeled data increases to a certain extent (5%, 10%), our method achieves superior detection accuracy compared to these methods. These results demonstrate that

our method effectively leverages a small amount of labeled data along with abundant unlabeled data to boost detection performance.

COCO-addition. We use the entire COCO train2017 set (100%) as labeled data and COCO unlabeled2017 as additional unlabeled data. The method's performance is evaluated on the COCO val2017 set. As shown in Table 1, our method improves $AP_{50:95}$ by 5.57% over the supervised baseline and outperforms other state-of-the-art SSOD methods. This improvement can be attributed to the method's ability to make full use of the prediction information on unlabeled data to guide the model training, which enhances the utilization efficiency of unlabeled data. The aforementioned results indicate that when the amount of labeled data is relatively large, our method can still effectively utilize the additional unlabeled data to achieve better performance.

Table 2. Comparison results on PASCAL VOC. The AP_{50} (%), AP_{75} (%) and $AP_{50:95}$ (%) is used as the evaluation metric. Bold numbers indicate the best performance. — means that the results are missing in the source paper. * denotes the results we reproduced.

Method	AP_{50}	AP_{75}	$AP_{50:95}$
Supervised*	72.20	45.10	42.40
STAC [1]	77.40	—	44.60
Instant-Teaching [14]	78.30	52.00	48.70
ISMT [15]	77.23	—	46.23
Unbiased Teacher [21]	77.40	—	48.70
Soft Teacher* [2]	79.00	57.40	51.47
MUM [3]	78.90	—	50.20
CST [16]	78.70	—	51.50
RPL [17]	76.90	57.90	52.40
Elaborate Teacher [20]	78.30	—	50.20
Ours	**80.40**	**58.80**	**52.81**

PASCAL VOC. In addition to the MS COCO dataset, we also conducted experiments on the PASCAL VOC dataset. We use VOC07 trainval set as labeled data, VOC trainval set as the unlabeled data. The method's performance is evaluated on the VOC07 test set. As shown in Table 2, our method outperforms other state-of-the-art SSOD methods across all evaluation metrics, further demonstrating its superiority. Notably, in terms of the metric for higher localization accuracy (AP_{75}), our method achieves a significant 1.4% improvement over Soft Teacher [2], which also considers the localization errors in pseudo-labels. This result indicates that our approach improves the localization quality for SSOD.

4.4 Ablation studies

Ablations of LER and PRC. To evaluate the effectiveness of the proposed LER and PRC strategy, we conducted ablation studies under the 10% labeled COCO-standard setting. As shown in Table 3, compared to the teacher-student framework baseline (No. 1), applying LER or PRC individually improves $AP_{50:95}$ by 0.7% and 1.0%, respectively

(No. 2 and No .3). Meanwhile, LER increases AP_{75} by 1.2%, indicating that LER is beneficial to improving the localization performance. This is attributed to the proposed localization quality score, which helps the model focus on pseudo-labels with higher localization accuracy. PRC increases AP_{50} by 0.7%, suggesting a positive effect on classification performance. This improvement comes from the teacher model's ability to better assess proposal reliability and enhance the impact of reliable proposals on classification. Furthermore, combining LER and PRC (No. 4) yields the best performance across all metrics. This result indicates that LER and PRC are compatible and can work synergistically to achieve maximum performance gains.

Table 3. Ablation study on the effects of LER and PRC.

No.	LER	PRC	AP_{50}	AP_{75}	$AP_{50:95}$
1			50.80	33.80	31.50
2	✓		50.60	35.00	32.20
3		✓	51.50	34.90	32.50
4	✓	✓	**53.40**	**35.70**	**33.50**

Analysis of foreground score threshold. We also investigated the impact of different foreground score thresholds (τ) on the proposed method. As shown in Table 4, when τ gradually increases from 0.6 to 0.8, $AP_{50:95}$ gradually improves and reaches the optimum at 0.8. This is because a higher threshold helps ensure the quality of pseudo-labels. However, when τ is further increased to 0.9, $AP_{50:95}$ decreases by 0.9%, primarily due to an insufficient number of pseudo-labels, which leads to the loss of valuable training samples. Notably, despite the lower pseudo-label quality at $\tau = 0.8$ compared to 0.9, the model achieves superior performance, which can be attributed to the effectiveness of LER and PRC in suppressing pseudo-label noise. Therefore, $\tau = 0.8$ is selected as the final setting to achieve an optimal balance between pseudo-label quality and quantity.

Table 4. Ablation study on the effects of different foreground score threshold.

τ	AP_{50}	AP_{75}	$AP_{50:95}$
0.6	52.70	35.10	32.90
0.7	53.30	35.70	33.30
0.8	**53.40**	**35.70**	**33.50**
0.9	51.90	34.90	32.60

4.5 Visualization

To better observe and compare the detection performance of our method, we visualize the detection results under the 10% labeled COCO-standard setting. our method is compared against the supervised baseline and the Soft Teacher [2] method, which also

considers localization errors. Three images were randomly selected from the COCO val2017 set for detection, as shown in Fig. 5. In the first row of images, our method demonstrates more accurate object localization than the other two methods. In the second row of images, the supervised baseline and Soft Teacher show false positives and missed detections, respectively, whereas our method correctly detects all objects. In the third row, the detected objects from our method exhibit higher foreground scores. These results highlight the superior detection performance of our method. This can be attributed to the dynamic reweighting strategy we employed in the bounding box regression and the classification tasks, which effectively enhances the model's localization precision and classification accuracy.

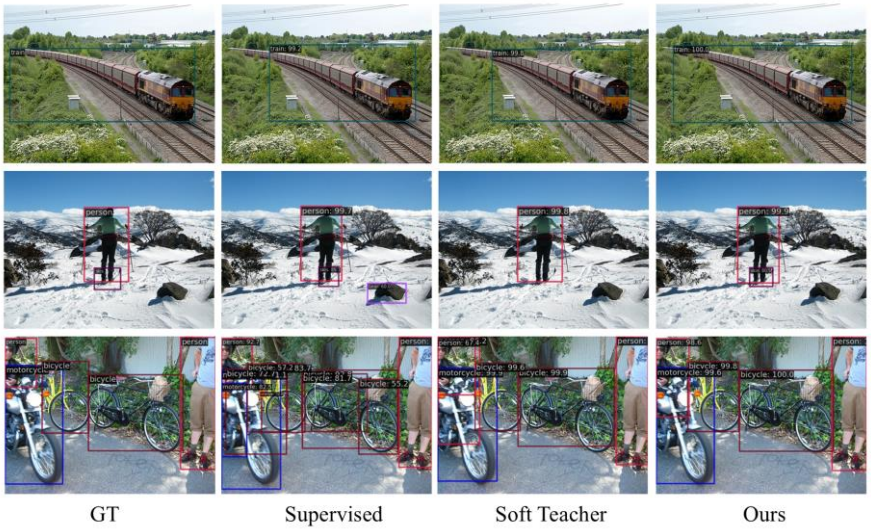

Fig. 5 Visualization of detection results. The first column shows the ground-truth, while the second to fourth columns display the detection results from the supervised baseline, Soft Teacher method, and our method, respectively.

5 Conclusion

A novel semi-supervised object detection (SSOD) method based on dynamic reweighting of localization errors is proposed in this paper. A localization quality score is first introduced to estimate the localization error of predicted bounding boxes, and then a Localization Error reweighting in Regression (LER) strategy is proposed to reweight the unsupervised regression loss according to the localization quality of pseudo-labels, thereby driving the model to prioritize accurately localized pseudo-labels during regression. Furthermore, a Proposal Reliability reweighting in Classification (PRC) strategy is proposed to adjust the unsupervised classification loss based on the reliability scores of proposals, amplifying the contribution of reliable proposals to the

classification task. Collectively, these strategies effectively mitigate the detrimental impact of localization noise within pseudo-labels on both the regression and classification tasks, enabling the model to fully leverage unlabeled data for performance enhancement. Extensive experiments on the MS COCO and PASCAL VOC datasets demonstrate that our method consistently surpasses the other mainstream state-of-the-art approaches. Our future work will focus on how to take advantage of unlabeled data more effectively to further improve the method's performance.

References

1. Sohn, k., Zhang, Z., Li, CL., Zhang, H., Lee, CY., Pfister, T.: A simple semi-supervised learning framework for object detection. arXiv preprint arXiv:2005.04757 (2020)
2. Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., Liu, Z.: End-to-end semi-supervised object detection with soft teacher. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3060–3069. IEEE, Montreal, Canada (2021)
3. Kim, J.M., Jang, J., Seo, S., Jeong, J., Na, J., Kwak, N.: Mum: Mix image tiles and unmix feature tiles for semi-supervised object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14492–14501 (2022)
4. Vandeghen, R., Louppe, G., Van Droogenbroeck, M.: Adaptive self-training for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 914–923 (2023)
5. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: Advances in neural information processing systems, vol. 33 (2020)
6. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, vol. 1, pp. 91–99. MIT Press, Montreal, Canada (2015)
7. 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: Computer Vision ECCV 2014: 13th European Conference, pp. 740–755. Springer, zurich, Switzerland (2014)
8. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. International journal of computer vision 88, 303-338 (2010)
9. Yang, X., Song, Z., King, I., Xu, Z.: A survey on deep semi-supervised learning. IEEE transactions on knowledge and data engineering 35(9), 8934–8954 (2022)
10. Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., Shinozaki, T.: FlexMatch: Boosting Semi-Supervised Learning with Curriculum Pseudo Labeling. In: Advances in neural information processing systems, vol. 34 (2021)
11. Li, D., Liu, Y., Song, L.: Adaptive weighted losses with distribution approximation for efficient consistency-based semi-supervised learning. IEEE Transactions on Circuits and Systems for Video Technology, 32(11), 7832–7842 (2022)
12. Chen, Y., Tan, X., Zhao, B., Chen, Z., Song, R., Liang, J., Lu, X.: Boosting semi-supervised learning by exploiting all unlabeled data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7548–7557 (2023)
13. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in neural information processing systems, vol. 30 (2017)

14. Zhou, Q., Yu, C., Wang, Z., Qian, Q., Li, H.: Instant-teaching: An end-to-end semi-supervised object detection framework. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4081–4090 (2021)
15. Yang, Q., Wei, X., Wang, B., Hua, X.S., Zhang, L.: Interactive self-training with mean teachers for semi-supervised object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5941–5950 (2021)
16. Liu, H., Chen, B., Wang, B., Wu, C., Dai, F., Wu, P.: Cycle Self-Training for Semi-Supervised Object Detection with Distribution Consistency Reweighting. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 6569–6578 (2022)
17. Li, H., Wu, Z., Shrivastava, A., Davis, L.S.: Rethinking pseudo labels for semi-supervised object detection. In: Proceedings of the AAAI conference on artificial intelligence, vol. 36, no. 2, pp. 1314–1322 (2022)
18. Chun, D., Lee, S., Kim, H.: USD: Uncertainty-based one-phase learning to enhance pseudo-label reliability for semi-supervised object detection. IEEE Transactions on Multimedia 26, 6336–6347 (2024)
19. Li, S., Liu, J., Shen, W., Sun, J., Tan, C.: Robust Teacher: Self-correcting pseudo-label-guided semi-supervised learning for object detection. Computer Vision and Image Understanding 235, 103788 (2023)
20. Yang, X., Zhou, Q., Wei, Z., Liu, H., Wang, N., Gao, X.: Elaborate Teacher: Improved semi-supervised object detection with rich image exploiting. IEEE Transactions on Multimedia 26, 11345–11357 (2024)
21. Liu, Y.C., Ma, C.Y., He, Z., Kuo, C.W., Chen, K., Zhang, P., Wu, B., Kira, Z., Vajda, P.: Unbiased teacher for semi-supervised object detection. In: Proceedings of the International Conference on Learning Representations (2021)
22. Zheng, H., Ye, R., Hou, Q., Ren, D., Wang, P., Zuo, W., Cheng, M.: Localization distillation for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 45(8), 10070–10083 (2023)