Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 1, No. 1, 2026

Semi-Supervised Learning Method for Drill Pipe Detection

Dongyu Zhang

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, Henan, China

Abstract: In underground coal mine drilling operations, complex environmental factors such as low illumination, dust interference, and severe occlusion significantly degrade the performance of visual perception systems. Meanwhile, large-scale surveillance video data generated in mining environments are difficult to fully annotate, leading to insufficient scene coverage in labeled datasets and consequently limiting the generalization capability of detection models. To address these challenges, this paper proposes a semi-supervised object detection method with a multi-strategy pseudo-label optimization mechanism based on a teacher–student framework, aiming to improve drill-related object detection performance under low-labeling conditions. Specifically, in the pseudo-label generation and filtering stage, a dual-teacher consistency verification mechanism is introduced, along with a scene-aware noise evaluation factor to adaptively adjust pseudo-label selection thresholds, thereby reducing the impact of noisy labels in complex environments. Furthermore, a pseudo-label localization optimization module is designed from the perspective of regression stability, where the spatial consistency of multiple regression predictions is exploited to evaluate localization quality and eliminate unreliable pseudo-labels with large deviations. In addition, an uncertainty-aware weighting strategy is proposed to dynamically adjust training loss by jointly modeling pseudo-label confidence, environmental noise, and localization quality, enabling more effective utilization of high-quality pseudo-labels while suppressing the influence of low-quality samples. Experiments are conducted on a self-constructed underground drilling dataset under different labeling ratios. The results demonstrate that the proposed method significantly improves detection accuracy in low-label scenarios and exhibits strong robustness and generalization capability in complex underground environments. This study provides an effective solution for leveraging large-scale unlabeled data in intelligent visual monitoring for coal mine applications and offers valuable insights for semi-supervised object detection in complex industrial scenarios.

Keywords: Semi-supervised learning; Object detection; Pseudo-label optimization; Underground coal mine; Teacher–student framework.

1. Introduction

Underground coal mine drilling is a critical process in gas extraction and geological exploration, and its operational efficiency and safety are directly related to the stability of mine production. In practical scenarios, accurate identification and status monitoring of key components, including the drill body, drill pipe, drill bit, and drill tail, are essential for achieving intelligent and refined management of underground operations. However, underground environments are typically characterized by low illumination, heavy dust interference, severe occlusion, and complex backgrounds. These factors significantly degrade the performance of visual perception systems, making traditional manual inspection or rule-based image processing methods insufficient to meet the requirements of high accuracy and real-time applications.

With the rapid development of deep learning, object detection methods have achieved remarkable progress in complex visual scenarios. One-stage detection algorithms represented by the YOLO[1] series have been widely applied in underground equipment monitoring tasks due to their high detection accuracy and real-time performance. Meanwhile, region proposal-based and multi-scale feature modeling methods also demonstrate strong robustness in complex environments [2,3].These approaches are capable of automatically learning multi-scale feature representations through end-to-end training, enabling effective object representation under challenging conditions. However, most existing deep learning-based methods rely heavily on large-scale, high-quality annotated datasets for supervised training [4].In underground mining scenarios, long-term monitoring systems generate massive amounts of video data, but manual annotation is labor-intensive and time-consuming, making it difficult to cover diverse operational conditions. As a result, the distribution of training data is often insufficient to represent real-world scenarios, thereby limiting model generalization and practical performance.

To alleviate the dependence on labeled data, semi-supervised learning has emerged as an important research direction in object detection [6].By incorporating unlabeled data into the training process, such methods can reduce annotation costs while improving model performance. Among them, teacher–student frameworks have been widely adopted, where the teacher model generates pseudo-labels for unlabeled data to guide the student model. Although these methods have shown promising results on general datasets, they still face significant challenges in underground drilling detection scenarios. First, complex environmental conditions may lead to false positives and missed detections, introducing substantial noise into pseudo-labels and affecting training stability. Second, existing approaches typically rely on confidence thresholds or simple consistency constraints for pseudo-label filtering, lacking effective mechanisms to evaluate localization accuracy, which makes it difficult to ensure spatial precision. Furthermore, conventional pseudo-label utilization strategies often adopt fixed weights or rely solely on confidence scores, failing to capture uncertainty variations under complex conditions, thus limiting the effective use of unlabeled data.

To address these issues, this paper focuses on drill object detection in complex underground environments and proposes a semi-supervised object detection method with a multi-strategy pseudo-label optimization mechanism based on the teacher–student framework. Specifically, a dual-teacher consistency verification mechanism is introduced to reduce noise caused by occasional misdetections through multi-model prediction agreement. In addition, a scene-aware noise evaluation factor is incorporated to adaptively adjust pseudo-label filtering thresholds, enabling the model to dynamically adapt to varying environmental conditions. Furthermore, from the perspective of bounding box regression stability, a pseudo-label localization optimization method is designed, which evaluates spatial consistency through multiple regression predictions and removes low-quality pseudo-labels with large localization deviations. Moreover, to address the varying reliability of pseudo-labels under complex environments, an uncertainty-aware weighting strategy is proposed, which integrates confidence scores, scene noise levels, and localization quality to dynamically reweight training loss. This allows high-quality pseudo-labels to contribute more effectively while suppressing the influence of noisy samples, thereby improving overall training stability and detection accuracy.

In experiments, a dedicated dataset for underground drilling scenarios is constructed, and systematic evaluations are conducted under different labeling ratios. Through ablation and comparative studies, the effectiveness of the proposed method is verified. The results demonstrate that, under low-label conditions, the proposed semi-supervised approach significantly improves detection accuracy and exhibits stronger robustness and generalization in complex environments. These findings indicate that multi-dimensional modeling and optimization of pseudo-label quality can effectively enhance the utilization of unlabeled data, providing a practical solution for intelligent visual monitoring in underground coal mines.

2. Related-work

With the rapid advancement of deep learning, object detection methods have achieved significant success in complex industrial visual scenarios. One-stage detection algorithms, represented by YOLOv5[5], have been widely applied in underground equipment recognition and operation monitoring due to their end-to-end training paradigm and high efficiency. These methods rely on convolutional neural networks to perform multi-scale feature modeling, demonstrating a certain level of robustness under challenging conditions such as complex backgrounds and low illumination. However, supervised detection methods typically require large amounts of high-quality annotated data, and in underground applications, the high cost and long cycle of annotation make it difficult to achieve sufficient scene coverage, thereby limiting model performance.

To reduce dependence on labeled data, semi-supervised learning has become an important research direction in object detection. Among existing approaches, teacher–student frameworks have attracted considerable attention. Representative methods such as STAC [7]utilize weak and strong data augmentations for unlabeled data and leverage teacher-generated pseudo-labels to guide student training, effectively exploiting unlabeled data. Building upon this, Unbiased Teacher[8] introduces a class-balanced mechanism to mitigate pseudo-label distribution bias and improve performance on long-tail categories. Subsequently, Soft Teacher[9] enhances pseudo-label representation by incorporating soft label information, further improving detection accuracy. These studies demonstrate that teacher–student frameworks can significantly boost performance under low-label conditions and have become the mainstream approach in semi-supervised object detection.

Despite their success on general datasets, these methods still suffer from unstable pseudo-label quality in complex industrial environments, especially in underground coal mine scenarios. To address this issue, some studies focus on improving pseudo-label filtering mechanisms[10]. For example, confidence-based thresholding filters low-confidence predictions but is sensitive to environmental variations, often discarding useful samples or retaining incorrect ones under noisy conditions. To improve reliability, several methods introduce multi-model consistency constraints[11], where prediction agreement across different models or augmentation branches is used to assess pseudo-label reliability, thereby reducing noise. However, such approaches typically rely on fixed thresholds and lack adaptability to dynamic noise variations, resulting in unstable performance across different conditions.

In addition to filtering strategies, pseudo-label localization accuracy is another critical factor affecting semi-supervised detection performance. Most existing methods primarily focus on classification confidence or bounding box overlap, while lacking comprehensive modeling of regression stability. In complex environments with background clutter or occlusion, even when multiple models agree on the category, their localization results may exhibit systematic deviations, which can negatively affect training. Recently, some studies have attempted to evaluate localization stability using multiple predictions or temporal information, but these approaches are still in the early stage and lack unified and effective solutions.

Furthermore, in terms of pseudo-label utilization, most methods adopt confidence-based weighting strategies, treating confidence as the primary measure of pseudo-label importance. However, in complex environments, confidence is influenced not only by object characteristics but also by image quality and background noise, making it insufficient to fully reflect pseudo-label reliability. Some studies attempt to incorporate uncertainty modeling or noise estimation to assign adaptive weights, but how to achieve stable and effective weighting under multi-factor conditions remains an open problem.

In summary, although existing semi-supervised object detection methods have made progress in pseudo-label generation and utilization, they still face challenges in suppressing noise, constraining localization errors, and effectively exploiting pseudo-labels in complex industrial environments. To address these issues, this paper proposes a systematic optimization framework based on the teacher–student paradigm, focusing on pseudo-label filtering, localization quality evaluation, and training weight assignment. By introducing a dual-teacher consistency mechanism with adaptive thresholds, a regression stability-based localization optimization method, and an uncertainty-aware weighting strategy, a more robust semi-supervised detection framework tailored for underground coal mine environments is developed.

3. Method

3.1. Semi-Supervised Drill Pipe Detection Framework

The overall architecture of the proposed semi-supervised object detection framework is illustrated in Fig. 1. The framework is built upon a teacher–student paradigm and incorporates multiple pseudo-label optimization strategies to improve the utilization efficiency of unlabeled data.

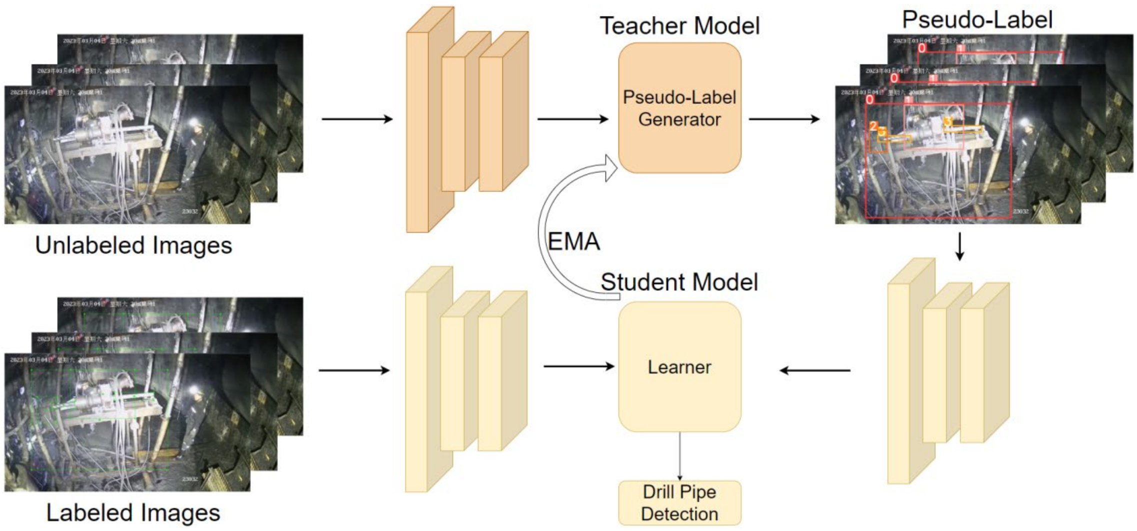

Fig.1 Basic Framework of Semi-Supervised Detection

The overall training process consists of the following steps:

(1) Supervised Training Stage: The student model is first initialized using a small amount of labeled data to acquire basic object detection capability.

(2) Pseudo-Label Generation Stage: During training, the teacher model performs inference on unlabeled images to generate candidate bounding boxes along with corresponding class probabilities. High-confidence predictions are retained as pseudo-labels based on a confidence threshold.

(3) Pseudo-Label Optimization Stage: To mitigate the negative impact of noisy pseudo-labels, three optimization strategies are introduced:

DTCV: Dual-Teacher Consistency Verification with adaptive thresholds to suppress noisy labels.

PLOM: Pseudo-Label Localization Optimization based on regression stability to improve localization quality

UWL: Uncertainty-aware weighting that dynamically assigns training weights based on confidence, noise level, and regression consistency.

(4) Student Model Training: The student model is trained jointly on labeled data and high-quality pseudo-labeled data, while the teacher model is updated via Exponential Moving Average (EMA).

Through these strategies, the proposed framework not only enhances detection accuracy using unlabeled data but also enables adaptive regulation of environmental noise, thereby improving overall detection performance.

3.2. Dual-Teacher Consistency Verification Mechanism

In semi-supervised object detection tasks, the quality of pseudo labels generated by teacher models directly affects the training performance of the student model. However, in underground drill pipe detection scenarios, factors such as dust interference, low illumination, occlusion, and equipment vibration often cause the teacher model to produce false detections or localization deviations, thereby increasing pseudo-label noise.

To mitigate the impact of noisy pseudo labels on model training, inspired by Chen[12], this paper proposes a Dual-Teacher Consistency Verification (DTCV) mechanism. Compared with single-teacher pseudo-label generation, the dual-teacher framework performs predictions from different parameter spaces on the same image and reduces accidental false detections through prediction consistency constraints.

Furthermore, conventional dual-teacher methods typically adopt fixed thresholds to determine pseudo-label validity, which cannot adapt to significant variations in noise intensity across different underground frames (e.g., dust-heavy frames versus clear frames). This may lead to either excessive filtering of valid labels or retention of numerous noisy labels.

To address this limitation, this study enhances the standard dual-teacher architecture by introducing an adaptive threshold mechanism, where thresholds dynamically adjust according to scene noise. Additionally, the derivation of the scene noise evaluation factor and the collaborative workflow of the dual teachers are explicitly modeled to improve interpretability and engineering feasibility. The overall structure is illustrated in Fig. 2.

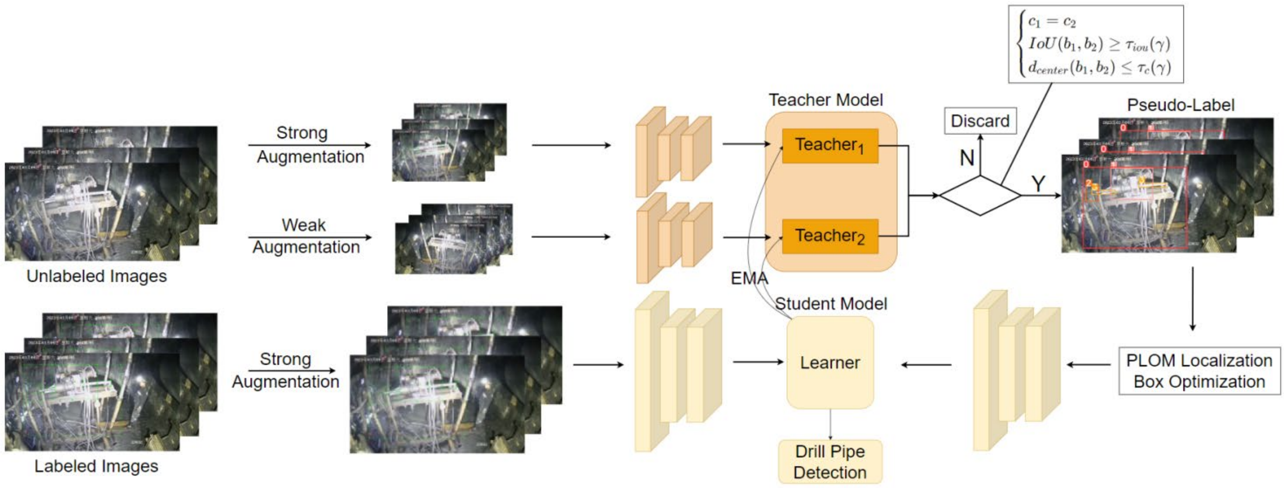

Fig.2 Dual-teacher consistency verification framework

Specifically, two teacher models with identical architectures but different initial parameters are constructed, denoted as $Teacher_1$ and $Teacher_2$. Both share the same network structure as the student model and are initialized by copying the student parameters while introducing diversity via different random initialization strategies.

During training, teacher models are only used for forward inference to generate pseudo labels and do not participate in backpropagation. Their parameters are updated using the Exponential Moving Average (EMA) of the student model, enhancing prediction stability and pseudo-label reliability. The update rules are defined as:

$$\alpha(e)=\alpha_0+(\alpha_{max}-\alpha_0)\cdot\left(1-\exp\left(-\frac{ce}{E_{max}}\right)\right) \quad (1)$$

$$\theta_t^{(k)}=\alpha(e)\theta_t^{(k)}+(1-\alpha(e))\theta_s \quad (2)$$

Where e denotes the current training epoch, $\alpha_0=0.05$is the initial smoothing coefficient, $\alpha_{max}=0.99$ is the maximum coefficient, c=0.01 is the decay factor, and E_{max} is the total number of training epochs.$\theta_t^{(k)}$ and θ_s represent the parameters of the k-th teacher and the student model, respectively.

The exponential increase of $\alpha(e)$ ensures that the teacher model quickly adapts to the student in early training stages while maintaining stable predictions in later stages, preventing pseudo-label degradation caused by parameter fluctuations.

Unlike conventional methods, this work introduces a scene noise evaluation factor γ\gammaγ to dynamically adjust pseudo-label filtering thresholds. The computation is as follows:

First, compute the average confidence of candidate pseudo labels in the current frame:

$$s_{avg}=\frac{1}{N}\sum_{i=1}^{N} s_i \quad (3)$$

where N is the number of candidate pseudo labels and $s_i\in[0,1]$ is the confidence score of the i-th pseudo label. A lower s_{avg} indicates poorer overall pseudo-label quality and stronger noise interference.

Next, image clarity is quantified using the metric β:

$$\beta=\frac{1}{M\times N}\sum|\nabla^2 I(x,y)| \quad (4)$$

where M and N denote image width and height, and $\nabla^2 I(x,y)$ represents image gradient variation, reflecting edge intensity and thus image clarity. $\beta\in[0,1]$, where smaller values indicate blurrier images and stronger interference.

Finally, the scene noise factor is defined as:

$$\gamma=\lambda\cdot(1-s_{avg})+(1-\lambda)\cdot(1-\beta) \quad (5)$$

where λ balances confidence and image quality. $\gamma\in[0,1]$, and a larger γ indicates stronger noise interference in the current frame, requiring stricter filtering thresholds; conversely, the filtering criteria can be relaxed when γ small.

After obtaining the noise evaluation factor γ, the pseudo-label filtering criteria are adaptively adjusted. Specifically, the final pseudo labels must satisfy the following three constraints:

Category consistency constraint: the predicted categories from the two teacher models must be identical:

$$c_1=c_2 \quad (6)$$

where c_1 and c_2 denote the predicted categories from $Teacher_1$ and $Teacher_2$, respectively.

Spatial overlap constraint: the Intersection over Union (IoU) between the two predicted bounding boxes must exceed an adaptive threshold $\tau_{IoU}(\gamma)$:

$$\tau_{IoU}(\gamma)=\tau_0\cdot(1+0.3\gamma) \quad (7)$$

$$IoU(b_1,b_2)\geq\tau_{IoU}(\gamma) \quad (8)$$

where τ_0 is the base threshold (e.g., $\tau_0=0.5$, thus $\tau_{IoU}(0)=0.5$ and $\tau_{IoU}(1)=0.65$). In this way, as the noise evaluation factor γ increases, the filtering strictness is correspondingly strengthened.

Center deviation constraint: the distance between the centers of the two predicted bounding boxes must be smaller than an adaptive threshold $\tau_c(\gamma)$:

$$\tau_c(\gamma)=\tau_{c0}\cdot(1-0.2\gamma) \quad (9)$$

$$d_{center}(b_1,b_2)\leq\tau_c(\gamma) \quad (10)$$

where τ_{c0} is the base threshold (e.g., $\tau_{c0}=5$, thus $\tau_c(0)=5$ and $\tau_c(1)=4$ .As γ increases, the threshold gradually decreases, imposing stricter constraints on localization errors under high-noise conditions.

In summary, the DTCV module achieves adaptive pseudo-label selection through the joint effect of scene noise estimation, dynamic threshold adjustment, and dual-teacher consistency verification. Compared with traditional dual-teacher methods with fixed thresholds, the proposed approach can dynamically adjust filtering criteria according to actual underground noise conditions (e.g., variations in dust concentration or illumination), thereby significantly improving pseudo-label reliability and training stability. Experimental results demonstrate that the proposed method effectively improves pseudo-label accuracy in underground drilling detection tasks and significantly enhances detection performance in complex environments.

3.3. Pseudo-Label Localization Quality Optimization Method

After the adaptive dual-teacher consistency-based coarse filtering, although obvious false-positive pseudo labels can be removed and classification accuracy is improved to a certain extent, when the two teacher models share similar architectures or feature representations, they may still produce consistent but incorrect predictions. For example, in underground drill pipe detection scenarios with complex backgrounds, both teacher models may generate similar false detections. Even for real targets, similar localization offsets may occur, resulting in consistent yet inaccurate pseudo labels. Therefore, relying solely on prediction consistency is insufficient to guarantee the localization quality of pseudo labels. Such inaccurate pseudo labels may introduce erroneous regression supervision signals, thereby affecting the model's ability to learn accurate object localization features and ultimately degrading detection performance and generalization ability.

To address this issue, a Pseudo-Label Localization Optimization Module (PLOM) is proposed as a post-processing step following dual-teacher consistency verification. This module evaluates pseudo-label localization quality based on the stability of bounding box regression. The core idea is that real targets tend to exhibit high localization consistency across multiple regression predictions, whereas noisy pseudo labels often show significant instability. Therefore, by analyzing the spatial consistency of bounding boxes across multiple regression processes, localization noise can be effectively identified and filtered out.

In the dual-teacher framework, the weakly augmented branch typically preserves more stable structural information, while the strongly augmented branch enhances robustness against perturbations. Thus, pseudo labels are first generated from weakly augmented images using the teacher model and then verified through strong augmentation consistency. On this basis, the PLOM module is introduced for further localization refinement. The overall process is illustrated in Fig. 3.

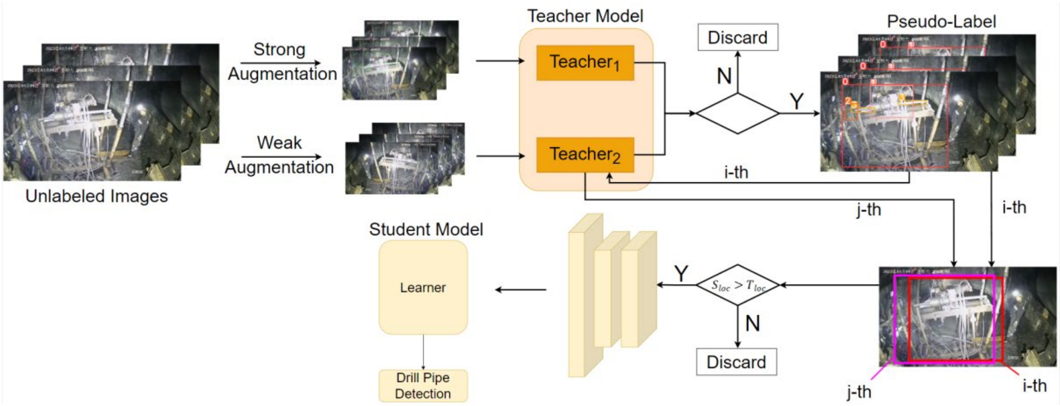

Fig.3 Pseudo-Label Localization Quality Optimization Pipeline

Let the pseudo-label bounding box after dual-teacher consistency filtering be defined as:

$$B_0=\{x_{tl},y_{tl},x_{br},y_{br}\} \quad (11)$$

where (x_{tl},y_{tl}) and (x_{br},y_{br}) denote the top-left and bottom-right coordinates of the bounding box, respectively.

To evaluate localization stability, the weakly augmented teacher model is used to perform regression re-estimation on the pseudo-label region. Specifically, slight random perturbations (e.g., small translations or scale variations) are applied to the region while preserving the overall image structure. The perturbed inputs are then forwarded through the teacher model multiple times, yielding K regression results:

$$B_1,B_2,\ldots,B_K \ (K\geq 2) \quad (12)$$

where B_i represents the bounding box obtained in the i-th inference.

After obtaining multiple regression outputs, spatial overlap between different bounding boxes is used to measure localization consistency. Let $IoU(B_i,B_j)$ denote the Intersection over Union between the i-th and j-th bounding boxes. The localization consistency score S_{loc} is defined as:

$$S_{loc}=\frac{2}{K(K-1)}\sum_{i<j} IoU(B_i,B_j) \quad (13)$$

This metric reflects the spatial stability of pseudo labels across multiple regression predictions. A higher S_{loc} indicates stronger localization consistency, while a lower value suggests significant drift and potential low-quality pseudo labels.

Based on the consistency score, a threshold-based filtering strategy is applied. Let T_{loc} denote the consistency threshold.A pseudo label is retained if $S_{loc}>T_{loc}$, and discarded otherwise (i.e., when $S_{loc}\leq T_{loc}$).

The final high-quality pseudo-label set is expressed as:

$$P_{final}=PLOM(P_{consist}) \quad (14)$$

where $P_{consist}$ denotes the pseudo-label set after dual-teacher consistency filtering, and P_{final} is the refined set after localization optimization.

By introducing the PLOM module, pseudo-label quality is further improved from the perspective of regression stability. Compared with traditional confidence-based or cross-model consistency-based filtering strategies, PLOM evaluates pseudo-label quality from the viewpoint of localization consistency, forming a task-specific validation mechanism for object detection.

3.4. Uncertainty-Aware Weighting Learning Strategy

After dual-teacher consistency filtering and PLOM-based localization refinement, most low-quality pseudo labels are effectively removed. However, due to complex underground conditions such as dust interference, illumination variation, and occlusion, some noisy pseudo labels may still remain. Meanwhile, conventional pseudo-label training strategies typically rely solely on detection confidence for weighting, which fails to distinguish between low-confidence labels caused by environmental noise and those representing ambiguous but informative targets, thereby affecting training stability.

To address this issue, an Uncertainty-Aware Weighting Learning (UWL) strategy is proposed. This method extends conventional confidence-based weighting by incorporating scene noise uncertainty and localization quality scores from the PLOM module. It evaluates pseudo-label reliability from multiple dimensions and dynamically adjusts training weights, enabling high-quality pseudo labels to contribute more while suppressing the influence of low-quality ones, thus improving robustness and detection accuracy in complex environments.

The UWL strategy constructs a joint weighting function combining confidence and noise factors. The weight distribution is illustrated in Fig. 4. As shown, higher weights are assigned to pseudo labels with high confidence and low noise, whereas weights are significantly reduced under low confidence or high noise conditions. This mechanism enables differentiated utilization of pseudo labels and enhances training stability.

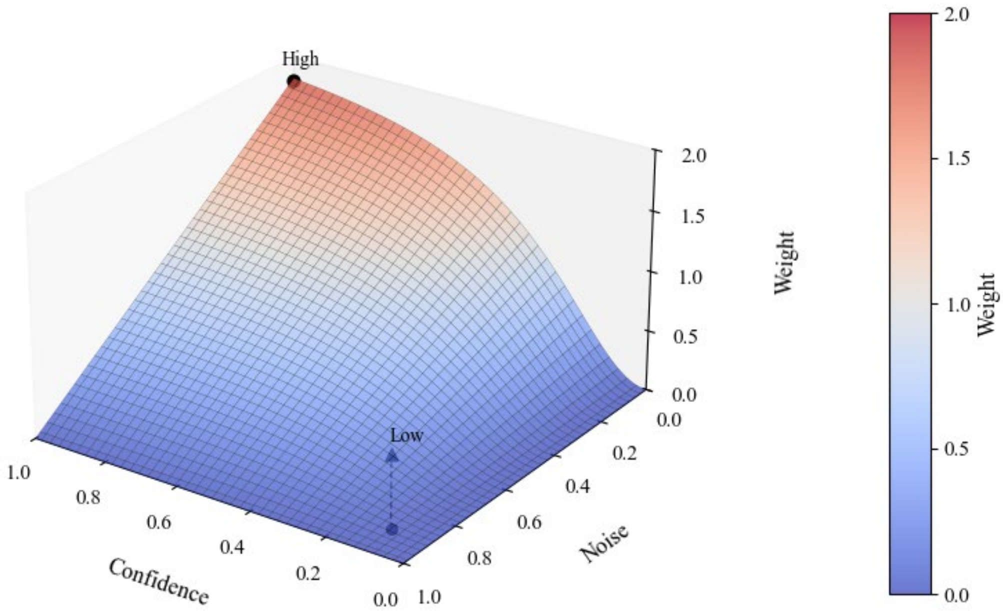

Fig.4 Weight Distribution in Confidence-Noise Space

n the implementation, the final weight of the iii-th pseudo-label sample is jointly determined by three components: the pseudo-label confidence, scene noise uncertainty, and bounding box localization quality.

First, the confidence score s_i reflects the reliability of the prediction and is directly used as the base weight:

$$w_i^{(1)}=s_i \tag{15}$$

Second, to mitigate the influence of complex environmental noise, the scene noise factor γ from the DTCV module is introduced to construct a noise suppression term:

$$w_i^{(2)}=1-\kappa\cdot\gamma \tag{16}$$

where κ\kappaκ is the noise tolerance coefficient. As the scene noise increases, w_{i2} decreases accordingly, thereby reducing the impact of noisy samples during training.

Furthermore, to enhance the contribution of high-quality pseudo labels, the localization quality score S_{loc} from the PLOM module (denoted as Q) is incorporated to refine the weight:

$$w_i=(w_i^{(1)}\times w_i^{(2)})\times(1+Q) \tag{17}$$

where Q represents the localization quality score of the predicted bounding box. The term 1+Q provides a gain for high-quality predictions while preventing the weight from degenerating to zero.

To ensure training stability, all weights are normalized:

$$w_i^{norm}=\frac{w_i}{\Sigma_{i=1}^{N} w_i} \tag{18}$$

Based on the normalized weights, the overall training loss is defined as:

$$L_{total}=L_{sup}+w_i^{norm}\cdot L_{pseudo} \tag{19}$$

where L_{sup} denotes the supervised loss and L_{pseudo} denotes the pseudo-label loss.

With this loss formulation, the model is trained jointly using both labeled and pseudo-labeled data. Meanwhile, the teacher model is continuously updated via the Exponential Moving Average (EMA) mechanism, forming a stable semi-supervised training loop. By modeling uncertainty from multiple dimensions and dynamically adjusting weights, the UWL strategy enhances the learning of high-quality pseudo labels while suppressing low-quality ones, thereby improving localization accuracy and robustness in underground drill pipe detection.

4. Experimental Results and Analysis

4.1. Drill Pipe Detection Dataset Construction

The experimental data are collected from underground drilling surveillance videos. A dataset named Drill Scene Dataset (DSD) is constructed through frame extraction. The dataset is designed following the principles of diversity, representativeness, and independence, with systematic procedures including data collection, filtering, partitioning, and annotation to ensure high quality and strong generalization capability.

The raw videos are captured from multiple mines under different tunnel layouts, working shifts, and lighting conditions. They include complex scenarios such as uneven illumination, dust interference, equipment occlusion, and cluttered backgrounds, ensuring environmental diversity.

During frame extraction, Python scripts are used to sample frames at adaptive intervals of 20–30 frames based on motion dynamics. Structural Similarity Index (SSIM) is applied to remove redundant frames. Subsequently, 200–300 representative frames are selected from each drilling stage to ensure adequate coverage of key actions and state transitions. The final DSD dataset contains 2,657 high-quality images from 14 typical drilling scenarios.

For dataset partitioning, a scene-independent splitting strategy is adopted to evaluate generalization to unseen environments. Among the 14 scenarios, 10 are used for training, 2 for validation, and 2 for testing, ensuring no overlap in spatial layout, equipment configuration, or operation stages. Images are annotated using LabelImg, with four categories: drill head, drill tail, drill rod, and drill body. Annotations are stored in XML format, including class labels and horizontal bounding box coordinates. The dataset contains 2,657 training images, 483 validation images, and 355 test images. Annotation examples are shown in Fig. 5.

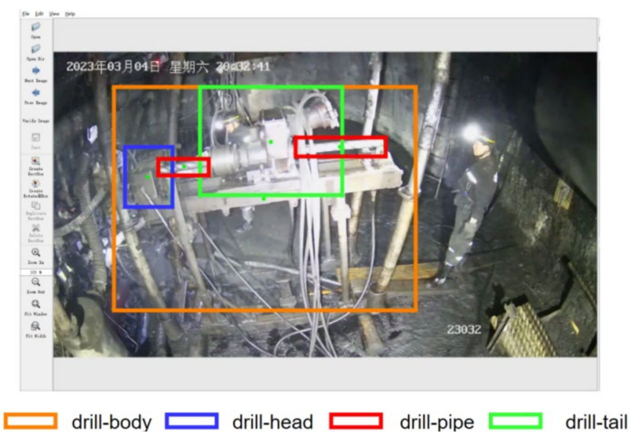

Fig.5 DSD Annotation Examples and Label Categories

To evaluate the effectiveness of the proposed semi-supervised method under low annotation ratios, a semi-supervised dataset is constructed based on DSD. Ten representative scenarios are selected as labeled training data, covering diverse operation stages, equipment poses, and background conditions. Four challenging scenarios are used for validation and testing (2 for validation and 2 for testing). Additionally, 41 underground drilling videos are collected from different time periods and viewpoints. After frame extraction and preprocessing, 41,116 unlabeled images are obtained, forming the unlabeled dataset S-DSD, which simulates a realistic "low-label, high-unlabeled" setting. Experiments are conducted under labeling ratios of 5%, 10%, and 15%.

4.2. Experimental Environment

To ensure reproducibility, all experiments are conducted under a unified hardware and software environment. The hardware configuration is listed in Table 1, and the software environment and training parameters are provided in Table 2.

Table 1. Hardware Configuration

Item	Details
Operating System	Ubuntu 24.04 LTS
CPU	Intel Xeon Gold 5218R (40 cores)
RAM	192 GB
GPU	Dual NVIDIA RTX A6000 (48 GB each)

Table 2. Software Environment and Training Parameters

Item	Details
Programming Language	Python 3.8.2
Framework	PyTorch
Environment	Conda (S-YOLO)
Batch Size	64
Input Size	640 × 640
Initial Learning Rate	0.01
Epochs	300

4.3. Evaluation Metrics

Based on common practices in semi-supervised object detection and the practical requirements of underground drilling equipment detection, the following evaluation metrics are adopted:

Mean Average Precision (mAP@0.5): A core metric in object detection, computed as the average precision across all categories (drill head, drill tail, drill rod, and drill body) at an IoU threshold of 0.5, reflecting overall detection accuracy.

Inference Speed (FPS): The number of image frames processed per second during inference, reflecting the practical efficiency of the method.

Pseudo Label Precision (PLP):

$$PLP=\frac{TP_{pseudo}}{TP_{pseudo}+FP_{pseudo}} \tag{20}$$

where TP_{pseudo} and FP_{pseudo} denote the number of correct and incorrect pseudo labels, respectively.

Pseudo Label Recall (PLR):

$$PLR=\frac{TP_{pseudo}}{TP_{pseudo}+FN_{pseudo}} \tag{21}$$

where FN_{pseudo} denotes the number of true objects missed by pseudo labels. This metric reflects the coverage ability of the teacher model.

Pseudo Label Utilization Rate (PLU):

$$PLU=\frac{N_{train}}{N_{pseudo}} \tag{22}$$

where N_{pseudo} is the total number of generated pseudo labels and N_{train} is the number of pseudo labels actually used for training. This metric evaluates the efficiency of the pseudo-label filtering strategy.

4.4. Ablation Studies

(1) Detection Performance Ablation Study

To verify the effectiveness of each core component in the proposed method, ablation experiments are conducted based on a baseline semi-supervised detection framework, where YOLOv5 is adopted as the detector. Each module is progressively introduced to evaluate its contribution to performance improvement. All experiments are conducted on the S-CMGED semi-supervised dataset.

To ensure fairness and eliminate confounding factors, all ablation experiments are performed under a unified 10% labeled data setting. Under this configuration, the impact of each module on model performance is analyzed incrementally. The results are shown in Table 3.

Table 3. Detection Performance Ablation Results (10% Labeled Ratio)

YOLOv5s	S-YOLO	+DTCV	+PLOM	+UWL	mAP@0.5/%	FPS/frame·s⁻¹
√					76.3	115.2
	√				78.8	112.6
	√	√			80.1	110.9
	√		√		79.6	111.7
	√			√	79.3	112.1
	√	√	√		81.2	109.5
	√	√		√	80.7	110.3
	√		√	√	80.2	110.8
	√	√	√	√	82.4	108.9

As shown in Table 3, compared with the fully supervised YOLO baseline, introducing the semi-supervised framework significantly improves detection performance. When using the basic semi-supervised framework (S-YOLO), the mAP@0.5 increases from 76.3% to 78.8%, achieving a gain of 2.5 percentage points. This demonstrates that pseudo labels generated via the teacher–student paradigm can effectively exploit unlabeled data.

After incorporating the Dual-Teacher Consistency Verification (DTCV) module, the mAP further improves to 80.1%. This indicates that enforcing prediction consistency between two teacher models helps filter erroneous pseudo labels and enhances their reliability, thereby reducing the impact of noisy supervision.

When only the Pseudo-Label Localization Optimization Module (PLOM) is introduced, the mAP reaches 79.6%. This result suggests that improving localization quality via regression stability verification effectively reduces the negative impact of low-quality pseudo labels and stabilizes the training process.

With the addition of the Uncertainty-Aware Weighting Learning (UWL) module, the mAP improves to 79.3%. Although the standalone improvement is relatively modest, UWL dynamically adjusts the contribution of pseudo labels during training, assigning higher weights to reliable samples and suppressing low-quality ones, thereby enhancing robustness.

When multiple modules are combined, further performance gains are observed. In particular, the combination of DTCV and PLOM achieves an mAP of 81.2%, indicating that temporal consistency and spatial consistency constraints are complementary. Finally, integrating all three modules yields the best performance, with an mAP of 82.4%, representing an improvement of 6.1 percentage points over the fully supervised baseline.

In terms of inference efficiency, as more modules are introduced, the FPS decreases slightly from 115.2 to 108.9 due to additional computations for pseudo-label filtering and localization optimization. However, the overall reduction is limited, and the model still maintains strong real-time performance, meeting the requirements of underground drilling monitoring. These results demonstrate that the proposed modules effectively collaborate to enhance pseudo-label generation, localization quality, and training weighting, leading to significant improvements in semi-supervised detection performance.

(2) Pseudo-Label Quality Ablation Study

To further evaluate the effectiveness of the proposed method in improving pseudo-label quality, a comparative analysis is conducted using different module combinations. Three metrics defined in Section 3.3.3 are adopted: Pseudo Label Precision (PLP), Pseudo Label Recall (PLR), and Pseudo Label Utilization Rate (PLU). The results are presented in Table 4.

Table 4. Pseudo-Label Quality Ablation Results

Method	PLP/%	PLR/%	PLU/%
S-YOLO	72.1	68.3	84.1
+DTCV	78.6	67.2	76.2
+DTCV+PLOM	82.4	66.9	71.6
+DTCV+PLOM+UWL	82.3	67.0	71.5

From Table 4, it can be observed that although the baseline S-YOLO achieves a relatively high utilization rate (84.1%), its precision is only 72.1%, indicating the presence of noisy pseudo labels.

After introducing the DTCV module, the PLP improves to 78.6%, demonstrating that consistency constraints effectively filter false positives. However, PLU decreases to 76.2%, indicating that stricter filtering reduces the number of pseudo labels used for training.

With the addition of PLOM, PLP further increases to 82.4%, confirming that evaluating regression stability can effectively eliminate localization-inconsistent pseudo labels. Although PLR decreases slightly, the overall change is minor, indicating that PLOM mainly improves localization accuracy without significantly harming recall.

After incorporating UWL, PLP remains high (82.3%) while PLR slightly increases to 67.0%. This shows that dynamic weighting enables better utilization of reliable pseudo labels while suppressing noisy ones, thereby improving training stability.

Overall, as DTCV, PLOM, and UWL are progressively introduced, pseudo-label precision consistently improves, while utilization rate moderately decreases. This indicates that the proposed method effectively filters noisy labels and retains high-quality ones, providing more reliable supervision for student model training.

4.5. Comparative Experiments

(1) Detection Performance Comparison

To validate the superiority of the proposed method in semi-supervised object detection, the proposed S-YOLO is compared with several state-of-the-art methods, including STAC, Unbiased Teacher, Soft Teacher, Dense Teacher, and Humble Teacher.

All methods are evaluated on the same S-CMGED dataset under 5%, 10%, and 15% labeled data settings, using the same YOLO detector and evaluation metrics for fairness. The results are shown in Table 5 and illustrated in Fig. 6.

Table 5. Detection Performance Comparison

Method	Type	5% mAP@0.5/%	10% mAP@0.5/%	15% mAP@0.5/%
YOLOv5s	Fully Supervised Learning	76.3	78.9	80.6
STAC	Semi-supervised	79.5	81.1	82.1
Unbiased Teacher	Semi-supervised	79.7	81.5	82.6
Soft Teacher	Semi-supervised	80.3	81.9	83.2
Dense Teacher[13]	Semi-supervised	79.9	81.8	82.9
Humble Teacher[14]	Semi-supervised	80.1	82.0	83.0
S-YOLO(Ours)	Semi-supervised	82.4	84.1	85.0

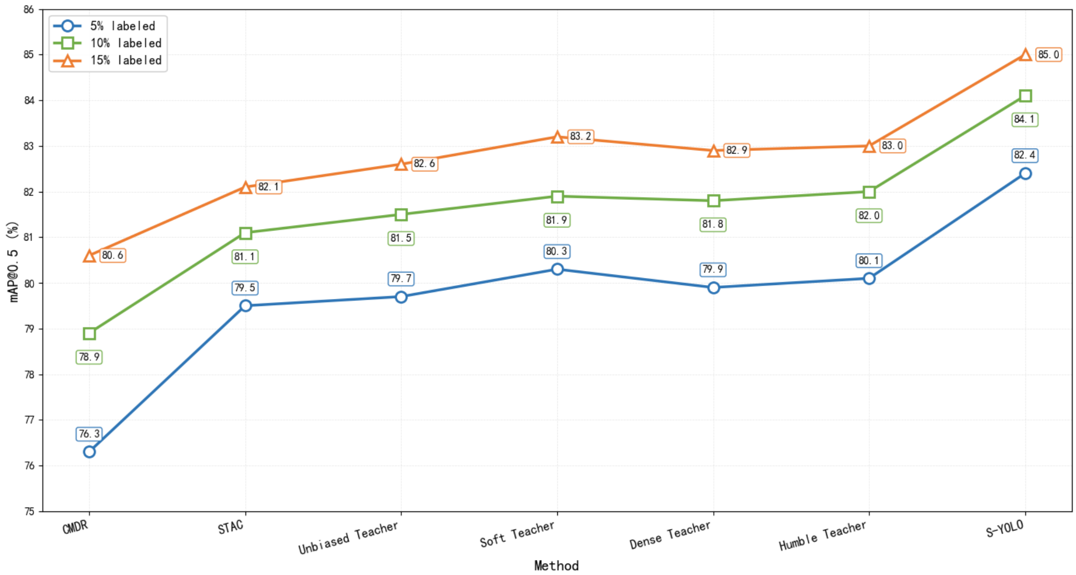

Fig.6 Performance Comparison under Different Labeling Ratios

As shown in Table 5, detection performance consistently improves with increasing labeled data across all methods. This is expected, as more labeled samples provide stronger supervision.

Under the extremely low-label setting (5%), the fully supervised YOLO achieves only 76.3% mAP. With semi-supervised learning, all methods improve performance by leveraging unlabeled data. Among them, STAC, Unbiased Teacher, and Dense Teacher reach around 79% mAP, demonstrating the effectiveness of semi-supervised learning.

In contrast, the proposed S-YOLO achieves 82.4%, outperforming the fully supervised baseline by 6.1 percentage points and surpassing the best competing method (Soft Teacher) by 2.1 percentage points. This indicates superior utilization of unlabeled data.

At the 10% labeling ratio, all methods show further improvements, with several exceeding 81% mAP. The proposed method achieves 84.1%, maintaining a clear advantage. This demonstrates that the combination of DTCV, PLOM, and UWL effectively enhances pseudo-label quality and improves training efficiency.

At the 15% labeling ratio, performance gaps between methods narrow as more labeled data become available. However, S-YOLO still achieves the highest mAP of 85.0%, outperforming Soft Teacher and Humble Teacher by 1.8 and 2.0 percentage points, respectively.

Overall, S-YOLO consistently achieves the best performance across all labeling ratios, with particularly significant gains under low-label conditions. This confirms that the proposed pseudo-label optimization strategies effectively improve unlabeled data utilization and enhance detection performance in complex underground environments.

5. Conclusion

To address the challenges of insufficient labeled data and unstable pseudo-label quality in underground drill rod detection under complex coal mine environments, this paper proposes a semi-supervised object detection method based on a teacher–student framework, incorporating a multi-strategy pseudo-label optimization mechanism. The proposed method focuses on three key stages: pseudo-label generation, filtering, and utilization. By introducing a dual-teacher consistency verification mechanism, an adaptive thresholding strategy, and a scene noise estimation method, the approach effectively reduces pseudo-label noise caused by factors such as dust interference, illumination variation, and occlusion in complex environments.

Furthermore, from the perspective of bounding box regression stability, a pseudo-label localization quality optimization method is developed. By analyzing the consistency of multiple regression predictions, the spatial localization accuracy of pseudo labels is significantly improved. In addition, considering the varying reliability of pseudo labels under complex conditions, an uncertainty-aware weighting strategy is proposed. This strategy integrates confidence scores, scene noise factors, and localization quality into a unified framework, enabling dynamic weighting of pseudo-label training samples and thereby enhancing training stability and model robustness.

Extensive experiments are conducted on a self-constructed underground drill rod detection dataset under different labeling ratios. The results demonstrate that the proposed method significantly improves detection performance, especially under low-label conditions. Compared with conventional fully supervised methods, the semi-supervised framework effectively leverages the latent information in unlabeled data. Moreover, with the integration of the proposed multi-strategy pseudo-label optimization mechanism, further improvements are achieved in both detection accuracy and robustness.

Ablation studies verify the effectiveness of each module in enhancing pseudo-label quality and optimizing the training process, while comparative experiments show that the proposed method consistently outperforms existing state-of-the-art semi-supervised detection approaches across different labeling ratios, with more pronounced advantages under low-label settings. Meanwhile, the model maintains relatively high inference efficiency while achieving improved accuracy, indicating its potential for practical deployment in real-world underground drilling scenarios.

Overall, this study demonstrates that, under the practical conditions of "large-scale data with limited annotations and highly complex, dynamic environments" in underground coal mines, multi-dimensional modeling and systematic optimization of pseudo-label quality can effectively enhance the utilization of unlabeled data, thereby improving detection performance in challenging scenarios. The proposed method not only provides an efficient and cost-effective solution for underground drill rod detection but also offers valuable insights for the application of semi-supervised object detection in complex industrial environments.

Despite the promising results, some limitations remain. The current approach primarily focuses on image-level information for pseudo-label optimization and does not fully exploit temporal correlations in video sequences. In addition, pseudo-label quality may still be further improved under extreme occlusion or high-noise conditions. Future work will explore incorporating temporal modeling and cross-frame consistency constraints to enable more refined dynamic pseudo-label optimization. Moreover, the development of more efficient lightweight model architectures will be investigated to further enhance deployment performance in real-world underground environments.

Conflicts of interest

The authors declare no conflicts of interest to report regarding the present study.

Acknowledgements

The authors express their gratitude to Henan Polytechnic University, for administrative and technical support.

References

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.

[2] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[3] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[4] SUN C, SHRIVASTAVA A, SINGH S, et al. Revisiting the unreasonable effectiveness of data in deep learning era[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017.

[5] Prasad R, Hussain M. A comprehensive review of YOLOv5 towards real-time detection[J]. Materials Today: Proceedings, 2024, 85: 102-110.

[6] ZHU X J, GHAHRAMANI Z. Learning from labeled and unlabeled data[R]. University of Wisconsin-Madison, 2002.

[7] SOHN K, BERTHELOT D, LIU C L, et al. A simple semi-supervised learning framework for object detection[C]. International Conference on Learning Representations, 2021.

[8] LIU Y C, MAO Z, ZHANG Z, et al. Unbiased Teacher for semi-supervised object detection[C]. International Conference on Learning Representations, 2021.

[9] XU M, ZHANG Z, et al. End-to-end semi-supervised object detection with Soft Teacher[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[10] ARAZO E, ORTEGO D, ALONSO A, et al. Pseudo-labeling and confirmation bias in deep semi-supervised learning[C]. International Conference on Learning Representations, 2020.

[11] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning[C]. Advances in Neural Information Processing Systems, 2017.

[12] Chen L, Zhang Y, Wang H, et al. Collaboration of Teachers for Semi-supervised Object Detection[EB/OL]. arXiv:2405.13374, 2024.

[13] ZHOU H, GE Z, LIU S, et al. Dense Teacher: Dense pseudo-labels for semi-supervised object detection[C]. Computer Vision – ECCV 2022. Lecture Notes in Computer Science. Springer, 2022.

[14] TANG Y, CHEN W, LUO Y, et al. Humble Teacher: Towards more reliable pseudo-labels for semi-supervised object detection[J]. arXiv preprint, 2022.