e-Navi

Available online at http://www.e-navi.kr/
e-Navigation Journal

Original article

Adaptive Pseudo-Label Calibration for Semi-Supervised PCB Defect Detection

Jingkai Bi[a],Mingzhi Chen[b],Zhuoyi Yang[c*],Kexin Zhang[d]

a School of Naval Architecture and Port Engineering, Shandong Jiaotong University, Weihai 264209, China,13188906372@163.com

b School of Naval Architecture and Port Engineering, Shandong Jiaotong University, Weihai 264209, China,cmz207225@163.com

c*School of Naval Architecture and Port Engineering, Shandong Jiaotong University, Weihai 264209, China,zhuoyi0716@126.com,Corresponding Author

d School of Naval Architecture and Port Engineering, Shandong Jiaotong University, Weihai 264209, China,24226015@stu.sdjtu.edu.cn

Abstract

Printed circuit board (PCB) defect inspection is a key task in industrial automatic optical inspection (AOI). Its deployment in low-label scenarios remains difficult because PCB defects are often small, weakly contrasted, and unevenly distributed across categories and scales. Semi-supervised object detection can reduce annotation demand by using unlabelled images, but its performance depends strongly on the reliability of pseudo labels. In PCB inspection, confidence scores are shaped not only by defect category but also by object scale, which makes fixed or class-only filtering insufficient for small-defect supervision. This paper proposes dual-granularity pseudo-label calibration (DGPC) for semi-supervised PCB surface defect detection. DGPC first fuses weak-view and strong-view teacher predictions into a unified candidate pool. It then estimates pseudo-label thresholds from both class-level and class–scale statistics, and applies a scale-dependent lower-bound relaxation to retain informative small-defect candidates. The method is applied only during training and leaves the detector architecture and inference procedure unchanged. Experiments on Deep PCB at 5%, 10%, and 20%labeled ratios show that DGPC improves pseudo-label-based training under low-label settings. The gains are most consistent when the teacher provides a sufficiently informative candidate pool, especially at 10% and 20% labelled ratios. Ablation, stability, cross-detector, pseudo-label statistics, scale-wise, qualitative, and sensitivity analyses show that class–scale-aware calibration improves pseudo-label selection for annotation-efficient PCB inspection.

Keywords: printed circuit board defect detection, industrial automatic optical inspection, semi-supervised object detection, pseudo-label calibration, annotation-efficient learning, small-scale defect detection

https://doi.org/10.52820/j.enavi.2026.26.038

1. Introduction

Printed circuit board (PCB) inspection is a fundamental stage in electronics manufacturing because subtle surface defects can affect electrical connectivity, assembly reliability, and long-term product stability. In industrial automatic optical inspection (AOI) pipelines, PCB defect detection is therefore a reliability-critical quality inspection task rather than a generic visual recognition problem. Common defects, including open circuits, shorts, mouse bites, pinholes, spurs, and spurious copper, often show weak saliency, irregular morphology, and substantial appearance variation across products and imaging conditions. Recent reviews and benchmark releases have further highlighted the need for robust PCB defect detectors under limited supervision [1,2].

Driven by the advancements in deep learning, using You Only Look Once (YOLO) for object detection has become the mainstream approach in PCB surface inspection. Recent studies have improved detector backbones, feature fusion modules, and lightweight deployment strategies for PCB defect detection. Yuan et al.[3] introduced YOLO-HMC to strengthen subtle-defect feature extraction and multiscale representation, while Lei et al.[4] emphasized reliability and lightweight design for PCB inspection scenarios with constrained computation. More recent methods, such as PCB-YOLO [5], simple and effective PCB defect detection network (SEPD Net)[6], the lightweight high-accuracy frame-work of Hou et al.[7], and HSA-RTDETR [8], have continued to improve the accuracy–speed trade of on PCB benchmarks. Nevertheless, even with these architectural refinements, small-scale defects remain difficult because they occupy only a small fraction of the image and often provide insufficient discriminative evidence for stable localization and classification.

This difficulty is closely related to the broader small-object detection problem. Small objects tend to produce weak feature responses, severe foreground–background imbalance, and less stable confidence estimates[9]. These effects are amplified in PCB images because defects may resemble normal copper textures and appear under dense background patterns. Detector design is therefore only one part of the problem. Low-label PCB inspection also requires reliable use of unlabeled images.

From an application perspective, the annotation bur-den is another major bottleneck. High-quality bounding-box annotation for PCB defects is expensive, tedious, and often dependent on domain expertise, particularly when subtle defects must be distinguished from normal manufacturing variation. This makes semi-supervised object detection (SSOD) an attractive direction for PCB inspection, since it aims to exploit a small labeled sub-set together with a larger collection of unlabeled im-ages. Existing surveys have shown that teacher–student learning and pseudo-label-based self-training have be-come the mainstream SSOD paradigm because they provide a practical balance between implementation simplicity and empirical performance[10,11]. Representative methods such as Unbiased Teacher[12], Soft Teacher[13], and Humble Teacher[14] demonstrated that high confidence pseudo labels generated by a teacher can effectively improve detector learning under scarce an-notations.

However, SSOD depends strongly on pseudo-label reliability. Pseudo-label inconsistency, inaccurate confidence estimates, and noisy box regression can reduce the benefit of unlabeled data. Recent studies have ad-dressed this issue through improved teacher–student training, dense pseudo labels, consistency modelling, and more robust pseudo-target selection [15-18]. These studies suggest that semi-supervised PCB inspection should focus on how pseudo labels are selected, not only on how many pseudo labels are generated.

This selection problem is particularly important for PCB datasets dominated by small defects. Pseudo-label confidence is affected by both defect category and object scale. Even within the same category, small instances may receive lower confidence because their local evidence is weak and their boundaries are ambiguous. A fixed threshold ignores category imbalance, whereas a class-wise threshold still treats different scales within the same category as equivalent. A more suitable filtering rule should therefore account for both class-level variation and intra-class scale variation.

To address this problem, this paper proposes dual-granularity pseudo-label calibration (DGPC) for semi-supervised PCB defect detection. The method first generates candidate pseudo labels from weak and strong teacher prediction views and fuses them into a unified pool. It then estimates filtering thresholds from both class statistics and class–scale statistics. A scale-dependent lower-bound relaxation further preserves small-defect candidates that are informative but assigned lower confidence. DGPC is a training-stage module, so it can be integrated with detector pipelines without changing test-time inference.

Experiments on Deep PCB under 5%, 10%, and 20%labeled ratios evaluate DGPC in representative low-label settings. The results show consistent improvements over a fixed-threshold baseline and clear gains at 10%and 20% labeled ratios. Stability, cross-detector, ablation, pseudo-label statistics, scale-wise, qualitative, and sensitivity analyses further support the roles of class–scale thresholding and lower-bound relaxation.

The main contributions of this paper are summarized as follows:

–To formulate pseudo-label filtering for PCB inspection as a class–scale calibration problem, motivated by the heterogeneous confidence distribution of small surface defects.

–To propose DGPC, which combines class-level and class–scale threshold estimation to select pseudo la-bels for semi-supervised PCB defect detection.

–To introduce scale-dependent lower-bound relaxation and verify its contribution through multi-ratio experiments, cross-detector evaluation, ablation, pseudo-label statistics, scale-wise analysis, qualitative com-parison, and parameter sensitivity analysis.

The remainder of this paper is organized as follows. Section II reviews PCB defect detection, semi-supervised object detection, and pseudo-label calibration. Section III presents the DGPC framework. Section IV reports the experiments. Section V discusses the implications and scope of the method. Section VI concludes the paper.

2.Related Work

2.1 Deep Learning-Based PCB Defect Detection

Deep learning has become the dominant technical route for PCB surface inspection because it can learn task specific discriminative representations directly from image data and is generally more robust than handcrafted pipelines under complex industrial conditions. He et al.[2] recently reviewed PCB defect inspection research over the past decades and noted a clear transition from traditional image-processing methods to deep learning-based detection frameworks. In parallel, the public re-lease and continued use of benchmark datasets such as Deep PCB have supported more standardized evaluation in this area[2].

Recent PCB defect detection studies have mainly focused on improving multiscale representation, light weight deployment, and sensitivity to small-scale defects. Yuan et al.[3] enhanced YOLO-based detection with stronger backbone modeling and feature refinement for PCB defects. Lei et al.[4] designed a reliable and light weight adaptive convolution network to balance accuracy and deployment efficiency. Wei et al. [5] further improved multiscale feature fusion and attention modeling in PCB-YOLO, while Lang et al.[6] proposed SEPD Net as a simple yet effective architecture for PCB surface defect detection. Hou et al. [7] emphasized practical accuracy–efficiency trade-offs, and Wang et al.[8] introduced a hierarchical scale-aware strategy based on RT-DETR for dense and multiscale PCB defects. Although these methods have advanced supervised PCB defect detection, they generally assume sufficient labeled training data and do not directly address the annotation scarcity that is common in industrial inspection. More recently, Hu et al.[19] further emphasized the diversity and practical difficulty of industrial PCB defects by introducing a design-guided detection framework together with the PCB-SCUT dataset.

Another recurring observation in the PCB literature is that small-scale defects remain a central challenge even when modern detectors are employed. This is consistent with the broader small-object detection literature, where weak texture evidence, imprecise localization, and feature dilution across deep layers are widely recognized as major bottlenecks[9]. Therefore, while supervised detector improvements are important, annotation-efficient learning strategies are still needed for practical PCB deployment when dense manual labeling is difficult to obtain.

2.2 Semi-Supervised Object Detection

Semi-supervised object detection seeks to reduce annotation cost by jointly exploiting a small labeled set and a larger unlabeled set. Existing surveys consistently show that pseudo-label-based teacher–student learning has become the mainstream SSOD framework because it integrates naturally with modern object detector sand scales well to realistic data regimes [10,11]. In this paradigm, the teacher predicts pseudo labels for unlabeled images, and the student is trained on both ground-truth annotations and pseudo-labeled samples.

Several representative SSOD methods have shaped this area. Unbiased Teacher[12] addressed pseudo-labeling bias through mutually beneficial teacher–student learning. Soft Teacher[13] used score-weighted classification learning and box jittering to improve regression super-vision. Humble Teacher[14] exploited soft labels from abundant proposals. Later studies extended teacher–student learning to broader detector families, dense pseudo labels, adaptive assignment, feature alignment, and dynamic score revision[15-17,20].

More recent work has also examined SSOD under less idealized unlabeled settings. Liu et al.[18] showed that unlabeled pools in real applications may contain out-of-distribution content or distribution mismatch, making pseudo-label filtering even more critical. Over-all, the SSOD literature suggests that pseudo-label quality, rather than the mere quantity of unlabeled data, is often the dominant factor determining final detector performance. This observation is highly relevant to PCB inspection, where small-scale defects are easy to miss and pseudo-label confidence is intrinsically unstable.

2.3 Pseudo-Label Calibration for Small-Scale Defect Detection

Pseudo-label calibration aims to improve the reliability of unlabelled supervision by filtering or reweighting self-generated labels. In object detection and related semi-supervised settings, a common practice is to use confidence thresholds to retain only reliable pseudo labels. However, threshold design is nontrivial because overconfident filtering reduces useful supervision, whereas loose filtering introduces noise. Wang et al.[21] ad-dressed this issue in sparsely annotated object detection by calibrating teacher confidence so that a shared threshold becomes more transferable across detector sand training stages. In the broader semi-supervised learning literature, class-aware thresholding has also been shown to be more effective than using a single fixed threshold for all categories. For example, Flex Match[22] dynamically adjusts class-wise thresholds according to learning status, and class-aware pseudo-labelling(CAP)[23] explicitly aligns pseudo-label assignment with class distribution in semi-supervised multi-label learning. Xiao et al.[24] further proposed metric-adaptive thresholding to consider both the quality and quantity of pseudo labels during thresh estimation.

These studies provide an important insight: pseudo-label confidence should not be treated as globally homogeneous. Most existing calibration strategies still operate at the image level or class level, whereas PCB defect confidence also varies with object scale. For small-defect inspection, class-only calibration may therefore overlook useful intra-class structure in pseudo-label reliability.

Compared with generic SSOD settings, semi-supervised PCB defect detection requires a more fine-grained pseudo-label selection mechanism. The key challenge is not only to distinguish reliable and unreliable pseudo labels, but also to preserve supervision for small, hard, yet semantically meaningful defect instances. Motivated by this gap, the next section introduces a dual-granularity pseudo-label calibration framework that explicitly links pseudo-label screening to both class variation and scale variation.

3.Methodology

3.1 Overview of the Proposed Framework

The proposed method addresses semi-supervised PCB defect detection from the perspective of pseudo-label reliability. Recent semi-supervised detection studies have shown that the performance of teacher–student frameworks is critically shaped by pseudo-label quality, assignment stability, and confidence calibration. Zhang et al.[25] showed that DETR-based semi-supervised detection is sensitive to inaccurate one-to-one pseudo assignments, while Shehzadi et al.[26] further improved pseudo-label reliability by introducing sparse learnable queries for semi-supervised detection. Hua et al.[27] additionally showed that pseudo-label instability remains a major issue in semi-supervised oriented detection, and Zhang et al.[28] further demonstrated that object-size bias can strongly affect semi-supervised detection performance. These observations are highly relevant to PCB inspection, where defects are visually weak, predominantly small-scale, and expensive to annotate.

Figure 1 illustrates the overall framework. The pipeline follows an offline teacher–student paradigm. First, a teacher detector is trained on the labelled subset. Second, the teacher generates pseudo labels for unlabelled images under weak and strong prediction views. Third, the two pseudo-label sets are fused and filtered by DGPC, which jointly considers defect class and object scale. Finally, a student detector is trained on the union of labelled data and calibrated pseudo-labelled data. The detector architecture is unchanged, so the method improves unlabelled supervision without altering test-time inference.

3.2 Teacher-Based Pseudo-Label Generation

Let $D_l = \{(x_i, y_i)\}_{i=1}^{N_l}$ denote the labeled subset and $D_u = \{u_j\}_{j=1}^{N_u}$ denote the unlabeled subset. A teacher detector $f_t(\cdot)$ is first trained on D_l using standard supervised detection learning. For each unlabeled image $u \in D_u$, two prediction views are constructed, namely a weak prediction view $T_w(u)$ and a strong prediction view $T_s(u)$. In the present implementation, the weak view corresponds to standard teacher inference without test-time augmentation, whereas the strong view is generated using the built-in augmented inference mode of the detector. Therefore, the distinction between the two views is not introduced through a hand-crafted external augmentation policy, but through two complementary inference conditions during pseudo-label generation.

The motivation for this dual-view design is consistent with recent semi-supervised detection literature. Zhang et al.[25] and Shehzadi et al.[26] showed that view inconsistency and pseudo-label instability remain central issues even in modern detector families. Fu et al.[29] similarly emphasized that static pseudo-label screening is insufficient when prediction quality varies across training stages and object characteristics. Therefore, in-stead of using a single prediction branch, the proposed method first constructs a complementary pseudo-label pool from weak and strong teacher predictions.

The teacher predicts two pseudo-label sets:

$$P^w(u) = \{p_i^w\}_{i=1}^{N_w}, \quad P^s(u) = \{p_j^s\}_{j=1}^{N_s} \qquad (1)$$

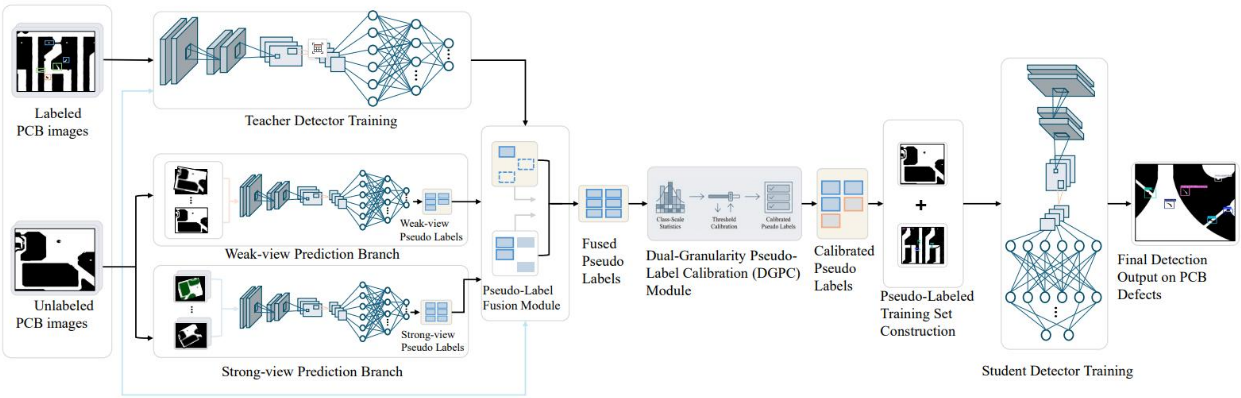


Figure 1: Overview of DGPC Within an Offline Teacher–Student Pipeline for Semi-Supervised PCB Defect Detection.

where each pseudo label $p = (b, c, \gamma)$ consists of a bounding box b, a category label c, and a confidence score $\gamma \in [0,1]$. Because the two views may emphasize difer-ent local evidence, the resulting pseudo-label sets aretreated as complementary candidate predictions rather than identical duplicates.

3.3 Cross-View Pseudo-Label Fusion

One-view pseudo labels may contain view-specific du-plicate boxes or class-inconsistent predictions. The pro-posed framework therefore performs cross-view pseudo-label fusion before threshold calibration.

For each unlabeled image, predictions from $P^w(u)$ and $P^s(u)$ are first regularized by class-wise non-maximum suppression and then merged into a unified candidate pool. Let NMS $(\cdot)$ denote non-maximum suppression and let Fuse $(\cdot)$ denote the cross-view fusion operator. The fused pseudo-label set is written as

$$P^f(u) = \mathrm{Fuse}\left(\mathrm{NMS}\left(P^w(u)\right),\ \mathrm{NMS}\left(P^s(u)\right)\right) \quad (2)$$

where the fusion operator merges weak-view and strong-view predictions and removes class-wise redundant boxes through post-fusion suppression.

This design serves two purposes. First, it reduces evidently redundant pseudo boxes before confidence calibration. Second, it allows the subsequent DGPC module to operate on a more stable candidate pool whose confidence statistics are less contaminated by view-specific outliers. Recent semi-supervised detection studies have similarly highlighted that pseudo-label filtering becomes substantially more effective when it is applied to pseudo labels that have already been regularized by consistency constraints, adaptive selection, or dense prediction refinement[26,29,30]. In the present work, this fused set provides the basis for the subsequent class–scale-aware calibration procedure.

3.4 Dual-Granularity Pseudo-Label Calibration

3.4. 1st Scale-Bin Partition

After cross-view fusion, each fused pseudo label is as-signed to a scale bin according to its normalized box area. For a pseudo label with normalized width w and normalized height h, its normalized area is defined as

$$a = w \times h \quad (3)$$

The pseudo label is then assigned to one of three scale bins:

$$s = \begin{cases} \text{small}, & 0 < a < \tau_s, \\ \text{medium}, & \tau_s \le a \le \tau_m, \\ \text{large}, & a \ge \tau_m, \end{cases} \quad (4)$$

where τ_s and τ_m denote the boundaries between thesmall, medium, and large pseudo-box bins.

This design is motivated by recent evidence that object scale can induce systematic bias in semi-supervised detection. Zhang et al.[28] explicitly identified multiple size-related biases in semi-supervised object detection and showed that small objects are disproportionately harmed by pseudo-label imbalance and biased learning dynamics. Although their study focused on aerial imagery, the underlying observation is highly relevant to PCB inspection, where the majority of defect instances are small-scale.

3.4. 2nd Class-Level and Class-Scale Threshold Estimation

Let C be the set of defect categories. For each class $c \in C$, the confdence values of all fused pseudo labelsbelonging to class c are collected as

$$S_c = \{\gamma_i | p_i \in P^f, cls(p_i) = c\} \quad (5)$$

A class-level threshold is first estimated by percentile statistics:

$$\tau_c = \text{Percentile}(S_c, q) \quad (6)$$

where q is the percentile hyperparameter.

To further account for intra-class scale variation, confidence values are grouped jointly by class and scale:

$$S_{c,s} = \{\gamma_i | p_i \in P^f, \text{cls}(p_i = c, \text{scale}(p_i) = s)\} \quad (7)$$

The class-scale threshold is then defined as

$$\tau_{c,s} = \begin{cases} \text{Perentile}(S_{c,s}, q), & |S_{c,s}| \geq n_{\min} \\ \tau_c, & |S_{c,s}| < n_{\min} \end{cases} \quad (8)$$

where $n_{\min}$ is the minimum sample count required forreliable class-scale threshold estimation.

This fallback mechanism is introduced because some class–scale groups may be too sparse for stable percentile estimation. In that case, reverting to the class-level threshold avoids excessive variance in the filtering rule. The overall rationale is consistent with recent work on detector calibration. Pathiraja et al.[31] showed that modern detectors are mis calibrated not only in class confidence but also in localization behavior, while Kuzucu et al.[32] further argued that detector calibration must be evaluated carefully and that confidence reliability cannot be treated as a trivial post-processing issue. Munir et al.[33] also showed that calibration-aware learning remains important in transformer-based detectors. In the same spirit, the proposed DGPC module does not assume globally homogeneous confidence distributions, but instead explicitly models heterogeneity across both class and scale.

3.4. 3rd Lower-Bound Relaxation

A class-scale threshold alone is still insufficient for PCB defect detection dominated by small-scale instances. In practice, many informative pseudo labels of small-scale defects have relatively low confidence not because they are semantically meaningless, but because they are intrinsically difficult to localize and classify. If the thresh-old is determined only by percentile statistics, pseudo labels from hard small-scale groups may still be over-filtered.

To alleviate this problem, a lower-bound relaxation mechanism is introduced:

$$\hat{\tau}_{c,s} = max(\beta_s \tau_{base}, \alpha_s \tau_{c,s}) \quad (9)$$

Where τ_{base} is a base confidence threshold, and α_s and β_s are scale-dependent factors. Here, α_s controls the relaxation applied to the class-scale threshold, whereas β_s defines the scale-dependent lower bound relative to the base threshold. For small-scale pseudo labels, the relaxation factors are set to prevent the calibrated threshold from becoming excessively strict; for medium and large pseudo labels, stricter filtering is retained because their confidence estimates are generally more reliable.

The final filtering rule is defined as

$$p_i \in P^* \quad \text{if} \quad \text{con}(p_i) \geq \hat{\tau}_{c,s} \quad (10)$$

where P^* denotes the calibrated pseudo-label set used for student training.

The role of lower-bound relaxation should be emphasized. It is not designed to loosen pseudo-label screening globally. Instead, it prevents the class-scale threshold from over-suppressing pseudo labels in the hardest size regime, thereby preserving supervision for the defect instances that are most difficult to detect in PCB inspection. This design is conceptually aligned with recent small-object-oriented SSOD studies. Zhang et al.[28] showed that size-unbiased strategies are necessary when small objects are systematically disadvantaged, and Zhao et al.[30] further demonstrated that dense pseudo-label selection in small-object-dominated scenes requires more careful filtering than conventional sparse pseudo-box screening.

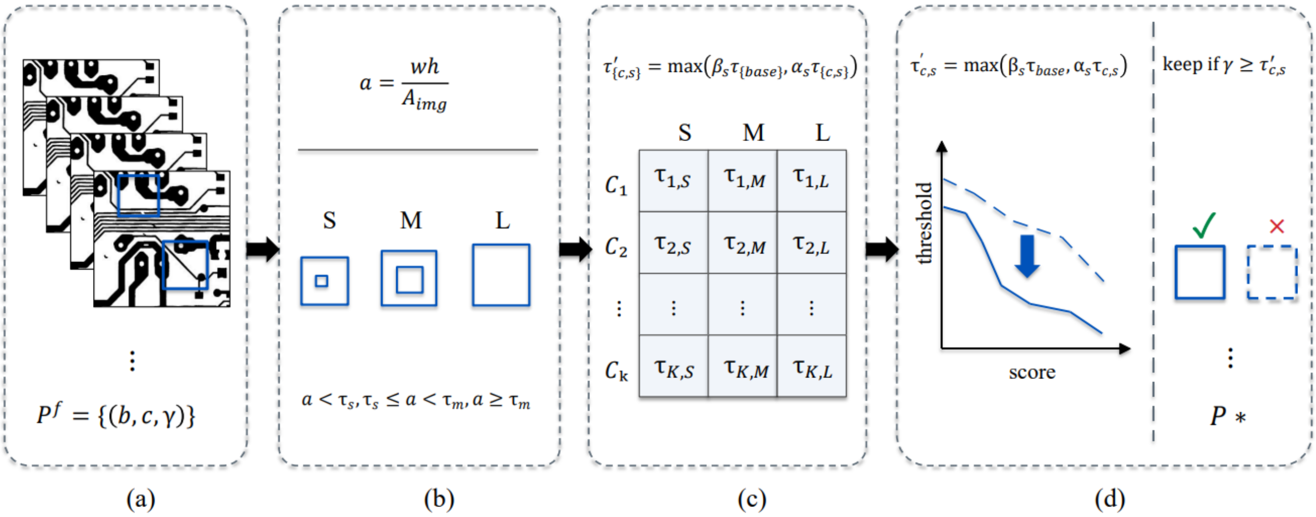


Figure 2: Workflow Of DGPC, Including Scale-Bin Assignment, Class–Scale Threshold Estimation, Lower-Bound Relaxation, and Pseudo-Label Filtering.

3.5 Student Training

After DGPC filtering, a pseudo-labeled training set is constructed by combining the original labeled images with unlabeled images whose annotations are replaced by calibrated pseudo labels. Let $\widetilde{D}_u = \{(u_j, \hat{y}_j)\}$ denote the pseudo-labeled unlabeled subset, where $\hat{y}_j$ is obtained from $P^*(u_j)$. The final student training set is

$$D_{\text{train}} = D_l \cup \widetilde{D}_u \tag{11}$$

The student detector is then trained on D_{train}, where labeled samples provide reliable ground-truth supervision and pseudo-labeled samples provide additional supervisory coverage on unlabeled PCB images.

This design keeps the framework detector-agnostic. Recent real-time detectors increasingly emphasize efficient multi-scale reasoning and end-to-end prediction quality[34,35]. DGPC is complementary to these advances because it improves pseudo-label selection at the training-pipeline level without adding test-time complexity.

4.Experiments

4.1 Lower-Bound Relaxation

4.1. 1st Dataset

Experiments are conducted on the Deep PCB bench-mark, which contains six PCB defect categories, namely *open, short, mouse bite, spur, pinhole,* and *spurious cop-per*.

Following the semi-supervised learning protocol, the training set is divided into labeled and unlabeled subsets under three labeled-data ratios, i.e., 5%, 10%, and 20%. These settings are designed to evaluate the proposed method from low supervision to practically relevant low-label regimes. The validation set is used for model selection and quantitative comparison.

4.1. 2nd Implementation Detail

The proposed framework is implemented using the Ultralytics YOLO pipeline, and both the teacher and student detectors adopt YOLO11n as the base detector unless otherwise specified. For all experiments, the framework follows an offline teacher–student procedure: the teacher is first trained on the labeled subset, pseudo-labels are then generated for the unlabeled subset, and the student is finally trained on the union of labeled and pseudo-labeled data. Both the teacher and the student are trained for 120 epochs with an input resolution of 640× 640 and a batch size of 4. The optimizer is automatically selected by the Ultralytics training engine, and the default optimization configuration is used, including an initial learning rate of 0.01, a final learning-rate factor of 0.01, a weight decay of 0.0005, and a 3epoch warmup stage. Teacher and student training both follow the default Ultralytics augmentation policy, including HSV perturbation, random translation, random scaling, horizontal flipping, mosaic augmentation, and random erasing.

During pseudo-label generation, weak-view and strong-view predictions are produced with the same confidence threshold of 0.05. The weak view corresponds to standard teacher inference, while the strong view is generated using the built-in augmented inference mode. Internal prediction-stage non-maximum suppression is disabled, after which class-wise NMS is applied manually with an intersection over union (IoU) threshold of 0.6. For each image, up to 2000 candidate predictions are retained before NMS, and at most 100 pseudo boxes are finally written. The weak-view and strong-view pseudo labels are then fused using class-wise NMS with a fusion IoU threshold of 0.6.

For pseudo-label calibration, the fixed-threshold baseline uses a unified confidence threshold of 0.25 for all classes and scales. The class-only SCPL variant uses a class-wise percentile threshold with percentile parameter q= 60 and base threshold τ_{base}= 0.25. The proposed DGPC variants further partition pseudo labels according to normalized box area, with τ_s= 0.01 and τ_m= 0.04 defining the boundaries of the small, medium, and large-scale bins. When a class–scale group contains fewer than 20 pseudo labels, the threshold falls back to the corresponding class-level estimate. For full DGPC, the percentile parameter is set to q= 60, the base threshold is set to τ_{base}= 0.25, and the scale-dependentlower-bound relaxation factors are set to (α_s, β_s)=(0.85, 0.80) for small pseudo labels, (0.92, 0.90) for medium pseudo labels, and (1.00, 1.00) for large pseudo labels. The DGPC w/o LB variant uses the same class–scale thresholding strategy but removes the lower-bound relaxation step. All labeled-ratio experiments use the same pipeline and hyperparameter configuration, with only the labeled/unlabeled split changed. Experiments are conducted on a single NVIDIA GeForce RTX 2080 Ti GPU using Python 3.10.20, PyTorch 2.2.2+cu118, and Ultralytics 8.3.235. The random seed is parsed from the experiment tag and applied consistently to Python, Py-Torch CPU/GPU, and Ultralytics training under deter-minimalism mode.

4.1. 3rd Evaluation Metric

Following standard object detection evaluation, Precision (P), Recall (R), mean average precision(mAP) are reported. Higher values indicate better performance. Among them, mAP@0.5 reflects overall detection ac-curacy under a relatively loose IoU criterion, whereas mAP@0.5:0.95 provides a stricter assessment of localization quality across multiple IoU thresholds. Reporting both mAP metrics together with precision and re-call is important for a balanced evaluation of pseudo-label calibration methods.

4.2 Main Results Under Different Labelled Ratio

Table 1 reports the main semi-supervised detection results on Deep PCB under three labelled ratios (5%, 10%, and 20%). Following the multi-seed protocol, each set-ting is evaluated over three random seeds, and the mean± standard deviation is reported. Three pseudo-label filtering strategies are compared under the same teacher–student pipeline, namely a fixed-threshold baseline, class-wise thresholding, and the proposed dual-granularity pseudo-label calibration (DGPC).

The results show that pseudo-label calibration be-comes increasingly valuable as the labeled subset pro-vides a stronger teacher. At the 5% labeled ratio, DGPC improves over the fixed-threshold baseline in mAP@0.5and mAP@0.5:0.95, while the class-wise strategy gives the highest average mAP. This indicates that class-level statistics remain useful when the teacher is trained from very limited supervision.

At the 10% labeled ratio, DGPC achieves the high-est average mAP@0.5 (0.9366) and precision (0.9200). Its mAP@0.5:0.95 is also close to the best class-wise result (0.6870 versus 0.6873). These results show that class–scale calibration can improve pseudo-label selectivity once teacher predictions become more informative.

The benefit is clearest at the 20% labeled ratio. DGPC obtains the highest mAP@0.5 (0.9603), mAP@0.5:0.95(0.7280), precision (0.9419), and recall (0.9106). This setting provides the strongest evidence that class–scale-aware filtering improves the quality of pseudo-label supervision in low-label PCB inspection.

4.3 Stability Analysis Under the 10% Labelled Ratio

To evaluate cross-split stability, this article reports DGPC results at the representative 10% labelled ratio using three random seeds. As shown in Table 2, mAP@0.5 ranges from 0.9290 to 0.9437, and mAP@0.5:0.95 ranges from0.6798 to 0.7008. The limited variance indicates that DGPC remains stable across labeled–unlabeled partitions.

This stability is relevant because SSOD can be sensitive to partition randomness and pseudo-label initialization. The multi-seed results support DGPC as a repeatable calibration procedure for pseudo-label selection.

4.4 Cross-Detector Generalization

To further examine whether the effectiveness of DGPC is specific to a single detector configuration, this article additionally evaluates the same pseudo-label calibration strategies using a larger YOLO variant, namely YOLO11s, under the representative 10% and 20% labeled ratios. This experiment is intended to verify whether the pro-posed method remains effective across detectors with different model capacities while keeping the same semi-supervised training pipeline and pseudo-label generation procedure.

Table 3 reports the multi-seed comparison on YOLO11s. At the 10% labeled ratio, DGPC achieves the highest mAP@0.5 (0.9461), mAP@0.5:0.95 (0.6564), and recall (0.8919). At the 20% labeled ratio, DGPC again achieves the highest mAP@0.5 (0.9662), mAP@0.5:0.95(0.7138), and precision (0.9559).

These results indicate that DGPC is not restricted to YOLO11n. The larger-capacity detector also benefits from class–scale pseudo-label calibration, which sup-ports the use of DGPC as a detector-agnostic training-pipeline module.

4.5 Ablation Study

Table 4 presents the ablation experiment results on the representative 10% labeled dataset under the calculation based on the average of multiple seeds. This experiment is designed to answer two key questions: whether pseudo-label calibration should go beyond a fixed threshold, and whether class–scale calibration alone is sufficient without lower-bound relaxation.

Class-only SCPL improves over the fixed-threshold baseline, especially in mAP@0.5:0.95 and precision. This confirms that pseudo-label filtering benefits from category-aware confidence statistics. Introducing class–scale grouping without lower-bound relaxation does not yield the best result, which indicates that scale-aware filtering also needs a mechanism to protect informative small-defect candidates.

Full DGPC achieves the highest mAP@0.5, mAP@0.5:0.95, and precision. The result supports the proposed design: class–scale statistics improve threshold specificity, while lower-bound relaxation prevents excessive filtering in the small-defect regime.

4.6 Pseudo-Label Statistics Analysis

To examine how DGPC changes pseudo-label selection, this article analyses pseudo-label statistics across scale bins in Figure 3. Most pseudo labels fall into the small-scale bin, confirming that Deep PCB is dominated by small defects. Pseudo-label handling in this bin therefore has a direct effect on student supervision.

Full DGPC retains more small-scale pseudo labels than DGPC w/o LB while preserving filtering in the other bins. Together with the ablation results, this sup-ports the role of lower-bound relaxation: it preserves useful small-defect supervision without globally loosening pseudo-label selection.

Table 1: Main results on Deep PCB under 5%, 10%, and 20% labeled ratios.

Labeled Ratio	Method	mAP@0.5	mAP@0.5:0.95	Precision	Recall
5%	Fixed-threshold	0.8475 ± 0.0506	0.5679 ± 0.0366	0.8114 ± 0.0607	**0.8080 ± 0.0332**
	Class-wise DGPC	**0.8735 ± 0.0311**	**0.5872 ± 0.0267**	**0.8606 ± 0.0423**	0.8002 ± 0.0224
		0.8593 ± 0.0361	0.5793 ± 0.0334	0.8288 ± 0.0516	0.8053 ± 0.0254
10%	Fixed-threshold	0.9294 ± 0.0148	0.6756 ± 0.0173	0.8930 ± 0.0190	**0.8804 ± 0.0136**
	Class-wise DGPC	0.9363 ± 0.0098	**0.6873 ± 0.0115**	0.9191 ± 0.0103	0.8689 ± 0.0044
		0.9366 ± 0.0073	0.6870 ± 0.0119	**0.9200 ± 0.0068**	0.8732 ± 0.0092
20%	Fixed-threshold	0.9538 ± 0.0079	0.7053 ± 0.0164	0.9200 ± 0.0135	0.9059 ± 0.0053
	Class-wise DGPC	0.9510 ± 0.0053	0.7229 ± 0.0158	0.9279 ± 0.0081	0.9023 ± 0.0175
		0.9603 ± 0.0068	**0.7280 ± 0.0168**	**0.9419 ± 0.0116**	**0.9106 ± 0.0177**

Table 2: DGPC stability under the 10% labeled ratio.

Seed	mAP@0.5	mAP@0.5:0.95	Precision	Recall
0	0.9437	0.7008	0.9278	0.8795
1	0.9371	0.6805	0.9155	0.8775

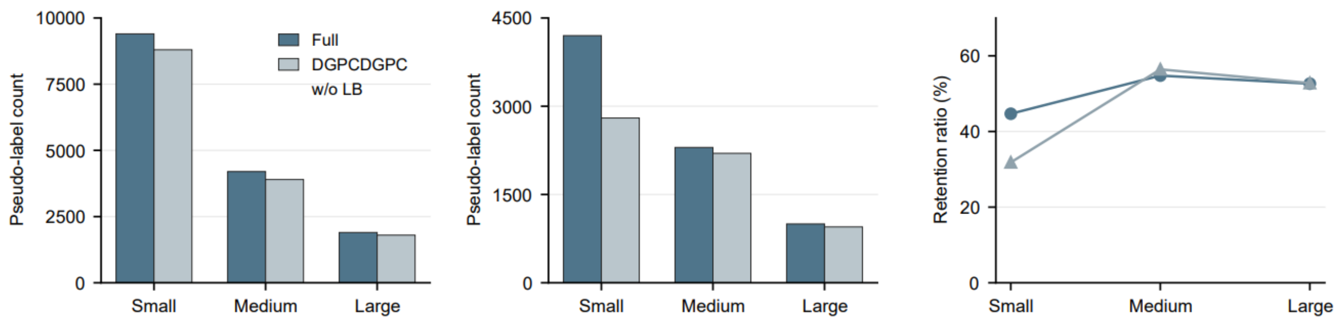
2	0.9290	0.6798	0.9167	0.8626
Mean ± Std	0.9366 ± 0.0073	0.6870 ± 0.0119	0.9200 ± 0.0068	0.8732 ± 0.0092

Table 3: Cross-detector evaluation on YOLO11s.

Labeled Ratio	Method	mAP@0.5	mAP@0.5:0.95	Precision	Recall
10%	Fixed-threshold	0.9327 ± 0.0452	0.6492 ± 0.0405	0.9034 ± 0.0447	0.8899 ± 0.0240
	Class-wise DGPC	0.9313 ± 0.0094	0.6412 ± 0.0189	0.9375 ± 0.0051	0.8558 ± 0.0167
		0.9461 ± 0.0187	**0.6564 ± 0.0210**	0.9306 ± 0.0260	**0.8919 ± 0.0258**
20%	Fixed-threshold	0.9652 ± 0.0094	0.7122 ± 0.0067	0.9383 ± 0.0245	**0.9277 ± 0.0082**
	Class-wise DGPC	0.9471 ± 0.0056	0.6782 ± 0.0233	0.9418 ± 0.0097	0.8846 ± 0.0051
		0.9662 ± 0.0041	**0.7138 ± 0.0317**	0.9559 ± 0.0052	0.9210 ± 0.0060

Table 4: Ablation of pseudo-label calibration strategies at the 10% labeled split.

Setting	mAP@0.5	mAP@0.5:0.95	Precision	Recall
Fixed-threshold baseline	0.957	0.696	0.901	**0.916**
Class-only SCPL	0.960	0.713	0.936	0.891
DGPC w/o LB	0.956	0.705	0.902	0.906
Full DGPC	**0.966**	**0.717**	**0.944**	0.914



（a） Raw Candidates （b） Retained Candidates （c） Retention Ratio

Figure 3: Pseudo-Label Statistics Across Scale Bins Before and After Calibration.

4.7 Scale-Wise Detection Analysis

To further verify whether the gain of DGPC is concentrated in the intended size regime, this article additionally evaluates scale-wise detection performance under the representative 10% labeled split. Following the same normalized box-area partition used in DGPC, validation in-stances are divided into small, medium, and large bins, and AP50 together with recall are computed separately for each scale group.

As shown in Table 5, the validation set contains 666 small instances, 13 medium instances, and 1 large instance. The evaluation is therefore concentrated in the small-defect regime. DGPC improves small-scale AP_{50} from 0.9193 to 0.9342 and recall from 0.9655 to 0.9775. The medium and large bins are included for completeness, but the small-scale results provide the main scale-wise evidence.

Together with Figure 3, this analysis shows that DGPC improves the dominant evaluation regime by retaining useful small-defect pseudo supervision.

4.8 Qualitative Results

Figure 4 shows representative qualitative comparisons on validation images. DGPC recovers several small-scaled effect instances that are missed or weakly detected by the baseline. It also produces stronger responses for some hard defects detected by both methods. These examples are consistent with the quantitative gains and the pseudo-label statistics.

For PCB inspection, recovering subtle defects is more important than changing confidence scores on easy samples. The qualitative results therefore illustrate the practical role of class–scale calibration in small-defect detection.

Table 5: DGPC stability under the 10% labeled ratio.

Method	Scale Bin	AP50	Recall
Baseline	small	0.9193	0.9655
DGPC	small	**0.9342**	**0.9775**
Baseline	medium	0.0293	1.0000
DGPC	medium	**0.0367**	1.0000
Baseline	large	0.0013	1.0000
DGPC	large	**0.0016**	1.0000

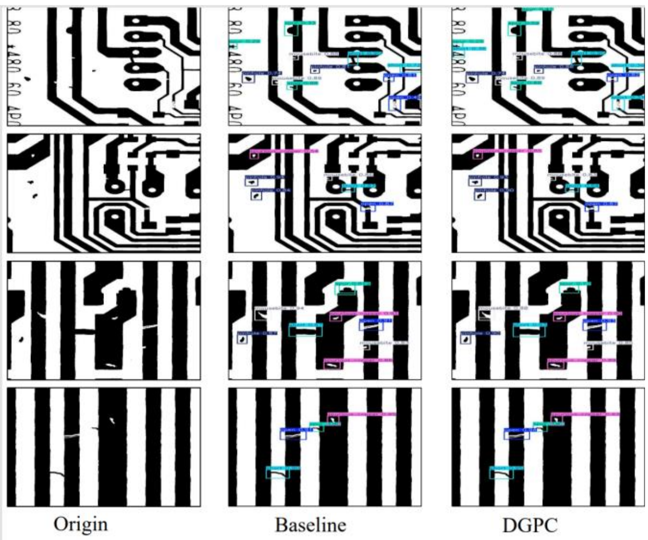


Figure 4: Qualitative Comparison Between the Baseline and DGPC On Representative PCB Defect Images.

4.9 Sensitivity Analysis of the Percentile Parameter

To examine the robustness of the proposed calibration strategy with respect to the percentile parameter, this article further evaluates DGPC under different values of q, including 50, 60, and 70, on the representative 10% labeled split. This experiment is designed to assess whether the performance gain of DGPC depends heavily on a narrowly tuned percentile setting or remains stable un-der moderate parameter variation.

As shown in Table 6, DGPC remains stable under moderate changes in q. The setting $q = 60$ gives the best overall trade-off, with the highest mAP@0.5, mAP@0.5:0.95, and recall. A lower percentile retains more low-confidence pseudo labels, whereas a higher percentile increases precision but reduces recall. The limited variation across $q \in \{50,60,70\}$ indicates that DGPC does not depend on a narrowly tuned percentile. In the current setting, $q = 60$ balances pseudo-label selectivity and retention.

The results show that pseudo-label calibration in semi-supervised PCB inspection should account for heterogeneous confidence distributions. Confidence varies with both defect category and object scale. Small defects often provide weaker local evidence and less stable boundaries, so their pseudo-label scores can be lower even when the predictions remain useful for training. This explains why fixed-threshold and class-only filtering do not fully capture pseudo-label reliability in PCB inspection.

The ablation study further clarifies the role of each component. Class-aware calibration improves over a shared threshold, but class–scale grouping is most effective when coupled with lower-bound relaxation. DGPC therefore balances two requirements: it increases filtering specificity by using class–scale statistics, and it avoids excessive removal of small-defect supervision. This design differs from a simple threshold adjustment because the threshold is conditioned on the structure of the pseudo-label pool.

The method is also practical for deployment-oriented inspection pipelines. DGPC changes only the training procedure and leaves the detector architecture unchanged. It can therefore be combined with compact one-stage detectors without adding inference overhead. This property is useful in AOI systems, where small missed defects can affect downstream quality control and where inference pipelines are often fixed by deployment constraints.

The current study focuses on Deep PCB and uses discrete scale bins with percentile-based threshold estimation. These choices make the method simple and reproducible, but broader industrial datasets and more adaptive calibration functions would further test its generality. Future work will examine DGPC under stronger domain shift, noisier unlabeled pools, and continuous confidence calibration.

5.Conclusion

This paper presented DGPC, a training-stage pseudo-label calibration method for semi-supervised PCB surface defect detection. DGPC estimates thresholds from class-level and class–scale confidence statistics, and uses lower-bound relaxation to retain informative small-defect pseudo labels. Experiments on Deep PCB under three labeled ratios showed that DGPC improves pseudo-label-based training, especially at 10% and 20% labeled ratios. Stability, cross-detector, ablation, pseudo-label statistics, scale-wise, qualitative, and sensitivity analyses support the contribution of class–scale-aware calibration. These findings suggest that annotation-efficient PCB inspection should calibrate pseudo-label reliability according to the scale characteristics of subtle defects. Because DGPC leaves the detector architecture and inference procedure unchanged, it can be integrated into deployment-oriented inspection pipelines with low engineering overhead.

Data availability

Data will be made available on request.

References

[1]He, Z., Lian, Y., Wang, Y., Lu, Z.: A comprehensive reviewof research on surface defect detection of pcbs based onmachine vision. Results in Engineering p. 106437 (2025)

[2]Lv, S., Ouyang, B., Deng, Z., Liang, T., Jiang, S., Zhang,K., Chen, J., Li, Z.: A dataset for deep learning based de-tection of printed circuit board surface defect. ScientifcData 11(1), 811 (2024)

[3]Yuan, M., Zhou, Y., Ren, X., Zhi, H., Zhang, J., Chen, H.:Yolo-hmc: An improved method for pcb surface defect de-tection. IEEE Transactions on Instrumentation and Mea-surement 73, 1–11 (2024)

[4]Lei, L., Li, H.X., Yang, H.D.: Reliable and lightweightadaptive convolution network for pcb surface defect detec-tion. IEEE Transactions on Instrumentation and Measure-ment 73, 1–8 (2024)

[5]Wei, Z., Yang, F., Zhong, K., Yao, L.: Pcb-yolo: Enhanc-ing pcb surface defect detection with coordinate attentionand multi-scale feature fusion. PLoS One 20(6), e0323684(2025)

[6]Lang, D., Lv, Z.: Sepdnet: Simple and efective pcb surfacedefect detection method. Scientifc Reports 15(1), 10919(2025)

[7]Hou, Y., Zhang, X.: A lightweight and high-accuracy frame-work for printed circuit board defect detection. EngineeringApplications of Artifcial Intelligence 148, 110375 (2025)

[8]Liu, C., Zhang, W., Lin, X., Zhang, W., Tan, X., Han,J., Li, X., Ding, E., Wang, J.: Ambiguity-resistant semi-supervised learning for dense object detection. In: Proceed-ings of the IEEE/CVF Conference on Computer Vision andPattern Recognition (CVPR), pp. 15579–15588 (2023)

[9]Nikouei, M., Baroutian, B., Nabavi, S., Taraghi, F., Aghaei,A., Sajedi, A., Moghaddam, M.E.: Small object detection:A comprehensive survey on challenges, techniques and real-world applications. Intelligent Systems with Applicationsp. 200561 (2025)

[10]Wang, Y., Liu, Z., Lian, S.: Semi-supervised object de-tection: A survey on recent research and progress. arXivpreprint arXiv:2306.14106 (2023)

[11]Shehzadi, T., Ifza, I., Liwicki, M., Stricker, D., Afzal, M.Z.:Semi-supervised object detection: A survey on progressfrom cnn to transformer. Sensors 26(1), 310 (2026)

[12]Liu, Y.C., Ma, C.Y., He, Z., Kuo, C.W., Chen, K.,Zhang, P., Wu, B., Kira, Z., Vajda, P.: Unbiased teacherfor semi-supervised object detection. arXiv preprintarXiv:2102.09480 (2021)

[13]Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F.,Bai, X., Liu, Z.: End-to-end semi-supervised object detec-tion with soft teacher. In: Proceedings of the IEEE/CVFInternational Conference on Computer Vision (ICCV), pp.3060–3069 (2021)

[14]Tang, Y., Chen, W., Luo, Y., Zhang, Y.: Humble teachersteach better students for semi-supervised object detection.In: Proceedings of the IEEE/CVF Conference on ComputerVision and Pattern Recognition (CVPR), pp. 3132–3141(2021)

[15]Liu, Y.C., Ma, C.Y., Kira, Z.: Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors. In: Proceedings of the IEEE/CVF Con-ference on Computer Vision and Pattern Recognition(CVPR), pp. 9819–9828 (2022)

[16]Zhou, H., Ge, Z., Liu, S., Mao, W., Li, Z., Yu, H., Sun, J.:Dense teacher: Dense pseudo-labels for semi-supervised ob-ject detection. In: Proceedings of the European Conferenceon Computer Vision (ECCV), pp. 35–50 (2022)

[17]Wang, X., Yang, X., Zhang, S., Li, Y., Feng, L., Fang, S.,Lyu, C., Chen, K., Zhang, W.: Consistent-teacher: Towardsreducing inconsistent pseudo-targets in semi-supervised ob-ject detection. In: Proceedings of the IEEE/CVF Con-ference on Computer Vision and Pattern Recognition(CVPR), pp. 3240–3249 (2023)

[18]Liu, N., Xu, X., Gao, Y., Zhao, Y., Li, H.C.: Semi-supervised object detection with uncurated unlabeled datafor remote sensing images. International Journal of Ap-plied Earth Observation and Geoinformation 129, 103814(2024)

[19]Hu, Z., Luo, J., Hong, Z., Chen, J.: A design-guided indus-trial defect detection method with enhanced camoufageddefect identifcation. Expert Systems with Applications p.128996 (2025)

[20]Liu, C., Zhang, W., Lin, X., Zhang, W., Tan, X., Han,J., Li, X., Ding, E., Wang, J.: Ambiguity-resistant semi-supervised learning for dense object detection. In: Proceed-ings of the IEEE/CVF Conference on Computer Vision andPattern Recognition (CVPR), pp. 15579–15588 (2023)

[21]Wang, H., Liu, L., Zhang, B., Zhang, J., Zhang, W., Gan,Z., Wang, Y., Wang, C., Wang, H.: Calibrated teacher forsparsely annotated object detection. In: Proceedings of theAAAI Conference on Artifcial Intelligence, pp. 2519–2527(2023)

[22]Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J.,Okumura, M., Shinozaki, T.: Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. Ad-vances in Neural Information Processing Systems 34,18408–18419 (2021)

[23]Xie, M.K., Xiao, J., Liu, H.Z., Niu, G., Sugiyama, M.,Huang, S.J.: Class-distribution-aware pseudo-labeling forsemi-supervised multi-label learning. Advances in NeuralInformation Processing Systems 36, 25731–25747 (2023)

[24]Xiao, J.H., Xie, M.K., Fan, H.B., Niu, G., Sugiyama, M.,Huang, S.J.: Dual-decoupling learning and metric-adaptivethresholding for semi-supervised multi-label learning. In:Proceedings of the European Conference on Computer Vi-sion (ECCV), pp. 437–454 (2024)

[25]Zhang, J., Lin, X., Zhang, W., Wang, K., Tan, X., Han,J., Ding, E., Wang, J., Li, G.: Semi-detr: Semi-supervisedobject detection with detection transformers. In: Proceed-ings of the IEEE/CVF Conference on Computer Vision andPattern Recognition (CVPR), pp. 23809–23818 (2023)

[26]Shehzadi, T., Hashmi, K.A., Stricker, D., Afzal, M.Z.:Sparse semi-detr: Sparse learnable queries for semi-supervised object detection. In: Proceedings of theIEEE/CVF Conference on Computer Vision and PatternRecognition (CVPR), pp. 5840–5850 (2024)

[27]Hua, W., Liang, D., Li, J., Liu, X., Zou, Z., Ye, X., Bai, X.:Sood: Towards semi-supervised oriented object detection.In: Proceedings of the IEEE/CVF Conference on ComputerVision and Pattern Recognition (CVPR), pp. 15558–15567(2023)

[28]Zhang, R., Xu, C., Xu, F., Yang, W., He, G., Yu, H., Xia,G.S.: S3od: Size-unbiased semi-supervised object detectionin aerial images. ISPRS Journal of Photogrammetry andRemote Sensing 221, 179–192 (2025)

[29]Fu, R., Chen, C., Yan, S., Wang, X., Chen, H.:Consistency-based semi-supervised learning fororiented object detec-tion. Knowledge-Based Systems 304, 112534 (2024)

[30]Zhao, T., Zeng, Y., Fang, Q., Xu, X., Xie, H.: Semi-supervised object detection for remote sensing images us-ing consistent dense pseudo-labels. Remote Sensing 17(8),1474 (2025)

[31]Pathiraja, B., Gunawardhana, M., Khan, M.H.: Multiclassconfdence and localization calibration for object

detection.In: Proceedings of the IEEE/CVF Conference on ComputerVision and Pattern Recognition (CVPR), pp. 19734–19743(2023)

[32]Kuzucu, S., Oksuz, K., Sadeghi, J., Dokania, P.K.: On cal-ibration of object detectors: Pitfalls, evaluation and base-lines. In: Proceedings of the European Conference on Com-puter Vision (ECCV), pp. 185–204 (2024)

[33]Munir, M.A., Khan, S.H., Khan, M.H., Ali, M., Khan, F.S.:Cal-detr: Calibrated detection transformer. Advances inNeural Information Processing Systems 36, 71619–71631(2023)

[34]Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu,Y., Chen, J.: Detrs beat yolos on real-time object detection.In: Proceedings of the IEEE/CVF Conference on ComputerVision and Pattern Recognition (CVPR), pp. 16965–16974(2024)

[35]Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J.,Ding, G.: Yolov10: Real-time end-to-end object detection.Advances in Neural Information Processing Systems 37,107984–108011 (2024)

Funding: This work was supported in part by the Key Research and Development Program of Shandong Province (Major Science and Technology Innovation Engineering, 2023) under Grant 2023CXGC010203, and in part by the Xin-jiang Talent Development Fund under Grant XJRC-2025-GX-YJ-QY-027.

Received	**05 June 2026**
1st Revised	**29 June 2026**
Accepted	**30 June 2026**