sensors

MDPI

Review

Semi-Supervised Object Detection: A Survey on Progress from CNN to Transformer

Tahira Shehzadi [1,2,3,*,†], Ifza Ifza [1,2,3,†], Marcus Liwicki [4], Didier Stricker [1,2,3] and Muhammad Zeshan Afzal [1,2,3,*]

[1] Department of Computer Science, Technical University of Kaiserslautern, 67663 Kaiserslautern, Germany; didier.stricker@dfki.de (D.S.)
[2] Mindgarage Lab, Technical University of Kaiserslautern, 67663 Kaiserslautern, Germany
[3] German Research Institute for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany
[4] Department of Computer Science, Electrical and Space Engineering University of Technology, 971 87 Luleå, Sweden; marcus.liwicki@ltu.se
* Correspondence: tahira.shehzadi@dfki.de (T.S.); muhammad_zeshan.afzal@dfki.de (M.Z.A.)
† These authors contributed equally to this work.

Abstract

The impressive advancements in semi-supervised learning have driven researchers to explore its potential in object detection tasks within the field of computer vision. Semi-Supervised Object Detection (SSOD) leverages a combination of a small labeled dataset and a larger, unlabeled dataset. This approach effectively reduces the dependence on large labeled datasets, which are often expensive and time-consuming to obtain. Initially, SSOD models encountered challenges in effectively leveraging unlabeled data and managing noise in generated pseudo-labels for unlabeled data. However, numerous recent advancements have addressed these issues, resulting in substantial improvements in SSOD performance. This paper presents a comprehensive review of 28 cutting-edge developments in SSOD methodologies, from Convolutional Neural Networks (CNNs) to Transformers. We delve into the core components of semi-supervised learning and its integration into object detection frameworks, covering data augmentation techniques, pseudo-labeling strategies, consistency regularization, and adversarial training methods. Furthermore, we conduct a comparative analysis of various SSOD models, evaluating their performance and architectural differences. We aim to ignite further research interest in overcoming existing challenges and exploring new directions in semi-supervised learning for object detection.

Keywords: transformer; object detection; DETR; computer vision; deep neural networks

Check for updates

Academic Editor: Bogdan Smolka

Received: 27 September 2025
Revised: 21 December 2025
Accepted: 25 December 2025
Published: 3 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Deep learning [1–3] has become an active area of research with numerous applications in various fields such as pattern recognition [4,5], data mining [6,7], statistical learning [8,9], computer vision [10,11], and natural language processing [12,13]. It has seen significant achievements particularly in supervised learning contexts, by effectively utilizing a substantial amount of high-quality labeled data. However, these supervised learning approaches [14–16], rely on labeled data for training that is costly and time-consuming. Semi-Supervised Object Detection (SSOD) [17,18] bridges this gap by incorporating both labeled and unlabeled data [19]. It shows a significant advancement in the field of computer vision [10,11], particularly for industries where obtaining extensive labeled data [17]

is challenging or costly. SSOD is used in various sectors, including Autonomous vehicles [20,21] as well as medical imaging [22,23]. In industries like agriculture [24–26] and manufacturing [27], where there is lots of data but labeling is time-consuming, SSOD helps make things more efficient.

Semi-supervised methods [28,29] enhance model performance and reduce labeling needs by employing both unlabeled and labeled data [30,31]. Moreover, previous object detection [32,33] approaches primarily involved manual feature engineering [34,35] and the use of simplistic models. These approaches faced difficulties in accurately identifying objects with different shapes and dimensions [36]. Later, the introduction of Convolutional Neural Networks (CNNs) [37,38] revolutionizes object detection by directly extracting hierarchical features [39] from raw data, enabling end-to-end learning [40] and substantially enhancing accuracy and effectiveness. In recent years, Semi-Supervised Object Detection has made significant improvement, driven by advancements in deep learning architectures [41,42], optimization techniques [43], and dataset augmentation strategies [44–47]. Researchers have developed various semi-supervised learning (SSL) approaches tailored for object detection, each with distinct strengths and limitations [48,49]. These approaches are mainly categorized into pseudo-labeling [50–52] and consistency regularization [53], both of which effectively utilize labeled and unlabeled data during training. Moreover, the integration of SSL methods with state-of-the-art object detection architectures such as FCOS [54], Faster R-CNN [55], and YOLO [56] has significantly enhanced the performance and scalability of Semi-Supervised Object Detection systems. This combination not only improves detection accuracy but also helps models work well with new and unseen datasets.

Object detection has seen remarkable progress with the advent of the DEtection TRansformer (DETR) [57–59]. Transformers, originally developed for natural language processing [12,13], excel in capturing long-range dependencies [60] and contextual information [61,62], making them ideal for complex spatial arrangements [63,64] in object detection. Unlike CNNs [38–40], which rely on localized convolutions and require non-maximum suppression (NMS) [65] to filter out redundant detections, DETR uses self-attention mechanisms [66,67] and do not need NMS. It considers the object detection task as a direct set prediction problem, eliminating traditional processes like NMS [65] and anchor generation [68]. Despite its advantages, DETR has limitations, such as slow convergence during training and challenges with small object detection. To address these issues, advancements in DETR enhance performance and efficiency through improved attention mechanisms and optimization techniques [69]. Following DETR's success, researchers are now employing DETR-based networks in Semi-Supervised Object Detection approaches [70–75]. This combines DETR's strengths with semi-supervised learning to use unlabeled data [53], reducing the need for large labeled datasets.

Due to the rapid progress of transformer-based Semi-Supervised Object Detection (SSOD) [19,70] approaches, keeping up with the latest advancements has become increasingly challenging. Therefore, a review of ongoing developments from CNN-based to Transformer-based SSOD methods is essential and would greatly benefit researchers in the field. This paper presents a comprehensive overview of the transition from CNN-based to Transformer-based approaches in Semi-Supervised Object Detection (SSOD). As shown in Figure 1, the survey categorizes SSOD approaches into CNN-based (one-stage and two-stage) [76–99] and Transformer-based approaches [73–75], highlighting techniques like pseudo-labeling and consistency-based labeling. It also provide details about data augmentation strategies [45–47,100–103], including strong, weak, and hybrid techniques.

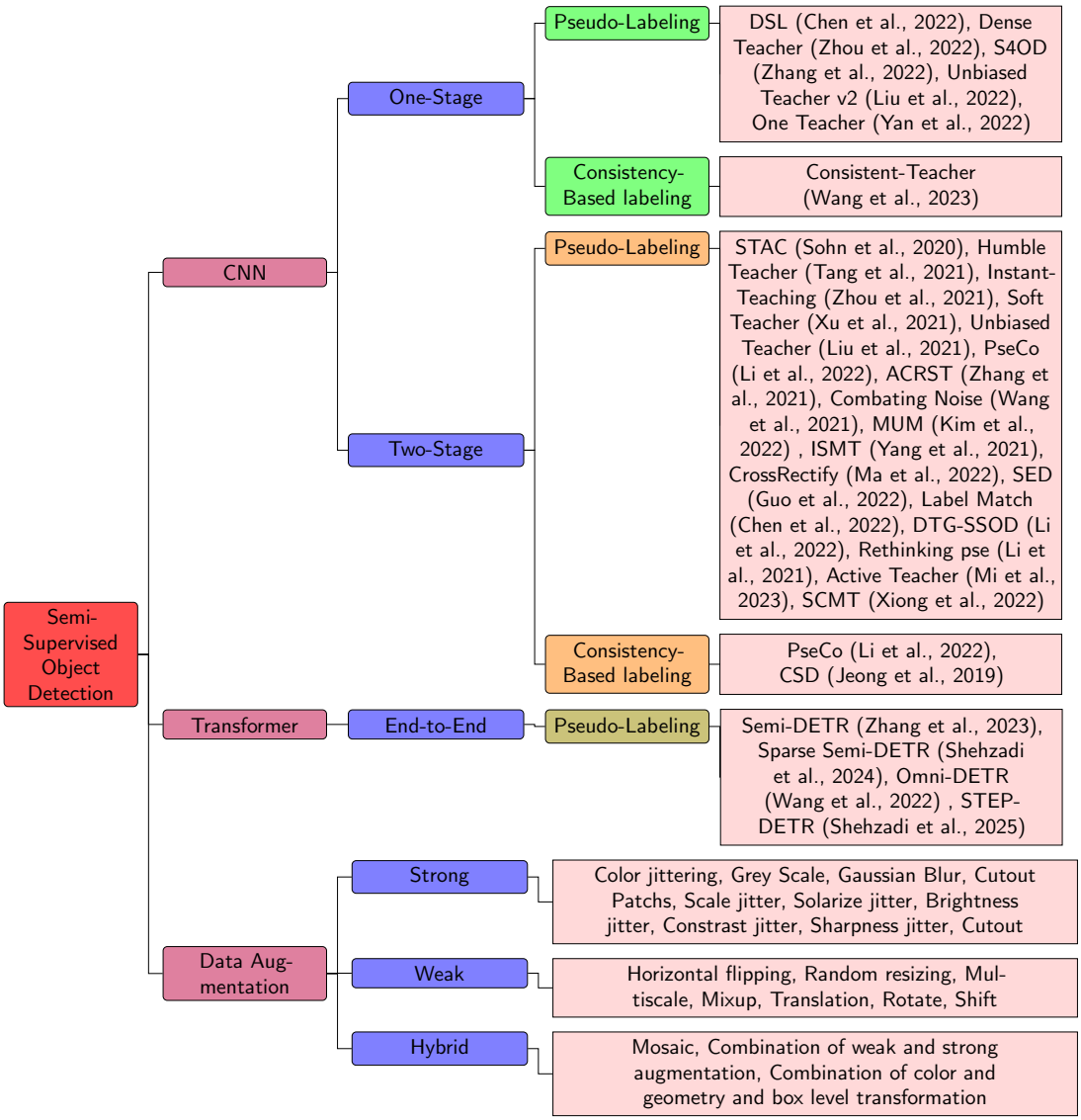

Figure 1. Semi-Supervised Object Detection: A Comprehensive Review and Taxonomy of Techniques.

Figure 2 depicts a teacher–student architecture tailored for semi-supervised object detection. A pretrained teacher model is utilized to generate pseudo-labels for unlabeled data. These pseudo-labels, along with the labeled data, are then utilized to jointly train the student model. By incorporating pseudo-labeled data, the student model learns from a more extensive and diverse dataset, enhancing its ability to detect objects accurately. Additionally, data augmentation methods are applied to both labeled and pseudo-labeled datasets. This collaborative learning approach effectively leverages both labeled and unlabeled data to improve the overall performance of object detection systems.

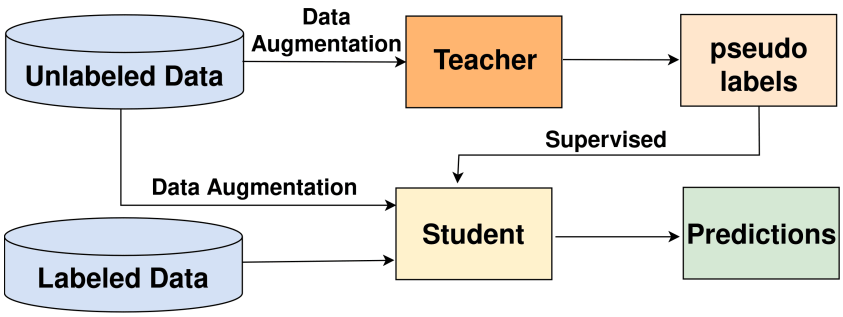

Figure 2. Teacher–Student Architecture for Semi-Supervised Object Detection.

Table 1 provides an overview of previous surveys on object detection, highlighting key research in semi-supervised learning. It covers a range of topics from theoretical advancements [104,105] to practical applications [106] across various domains. These surveys investigate diverse methodologies and their effectiveness, including specific applications in tweet sentiment analysis [107] and medical contexts [108]. Recent works explore improvements within machine learning frameworks [109], addressing challenges posed by small data and industrial applications with noisy or incomplete labels [106]. Notably, some surveys focus on deep visual learning and image classification using semi-supervised [48,110], self-supervised [108,111], and unsupervised methods [112], providing valuable insights into their effectiveness and challenges. Collectively, these surveys offer a detailed understanding of the advancements, challenges, and practical implementations in the field of Semi-Supervised Object Detection. While previous surveys have focused on CNN-based SSOD methods, the rise of Transformer-based Semi-Supervised Object Detection requires thorough evaluation to understand their effectiveness and trends.

Table 1. Overview of previous surveys on object detection. For each paper, the publication details are provided.

Title	Year	Description
Semi-Supervised Learning Literature Survey [104]	2008	This survey examines the landscape of semi-supervised learning literature concentrating on diverse methodologies and applications.
A Survey On Semi-Supervised Learning Techniques [108]	2014	An Analysis investigates various techniques in semi-supervised learning, offering insights into their effectiveness and applications.
A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning [107]	2016	This study provides a thorough comparison and analysis of tweet sentiment methods employing semi-supervised learning techniques.
Semi-supervised learning for medical application: A survey [113]	2018	This paper delves into the integration of semi-supervised learning within medical contexts, offering insights into its applicability and potential advancements.
A survey on semi supervised learning [109]	2019	This comprehensive examination explores the domain of semi-supervised learning, shedding light on its practical implementations and advancements.
Improvability Through Semi-Supervised Learning: A Survey of Theoretical Results [105]	2020	This analysis investigates theoretical advancements facilitated by semi-supervised learning, exploring avenues for improvement within machine learning frameworks.
Small Data Challenges in Big Data Era: A Survey of Recent Progress on Unsupervised and Semi-Supervised Methods [110]	2020	This exploration examines recent progress in unsupervised and semi-supervised methods, addressing challenges posed by small data in the context of the big data era.
A Survey of Un-, Weakly-, and Semi-Supervised Learning Methods for Noisy, Missing and Partial Labels in Industrial Vision Applications [106]	2021	This survey evaluates unsupervised, weakly-supervised, and semi-supervised learning techniques designed to address problems caused by noisy, incomplete, and missing labels in industrial vision applications.
Semi-Supervised and Unsupervised Deep Visual Learning: A Survey [112]	2022	This study explores the field of deep visual learning, with a particular focus on semi-supervised and unsupervised methods. It aims to uncover key insights and advancements in these approaches.
A Survey on Semi-, Self- and Unsupervised Learning for Image Classification [111]	2022	This survey examines image classification, focusing on semi-supervised, self-supervised, and unsupervised learning methods to understand their effectiveness and potential applications.
A survey on semi-supervised learning for delayed partially labelled data streams [114]	2022	This study delves into semi-supervised learning approaches employed for handling delayed data streams with semi labels, focusing on their effectiveness and challenges.
Semi Supervised deep learning for image classification with distribution mismatch: A survey [115]	2022	This study explores Semi-Supervised Deep Learning for image classification with distribution mismatch, providing insights into its strategies and challenges.
A Survey on Deep Semi-supervised Learning [49]	2023	This survey examines the field of deep semi-supervised learning techniques, providing insights into their applications and advancements.
Graph-based semi-supervised learning: A comprehensive review [116]	2023	This comprehensive review examines the effectiveness and applications of graph-based semi-supervised learning methods.

The remainder of this paper is organized as follows: Section 2, the core of this paper, offers a comprehensive overview of SSOD approaches. Section 3 examines different loss functions used in SSOD. Section 4 presents a comparative analysis of SSOD approaches. Section 5 addresses open challenges and future directions.Section 6 explores the role of SSODs in various vision tasks. Finally, Section 7 concludes the paper.

2. Semi Supervised Strategies

Semi-supervised object detection (SSOD) relies on both labeled and unlabeled data to improve detection performance with minimal annotation cost. Existing approaches mainly include one-stage and two-stage methods [71,76,78–80,82,96,99], as well as emerging Transformer-based detectors [73–75]. Together, these approaches highlight the diverse design spaces and methodological innovations within SSOD. To ensure a representative and comprehensive survey, the papers included in this section were selected based on their relevance to SSOD, publication in reputable venues (e.g., CVPR, ICCV, ECCV, NeurIPS), methodological novelty, and demonstrated empirical impact on the field.

2.1. One Stage

2.1.1. One Teacher

One Teacher [99] is a teacher–student framework tailored for one-stage SSOD, particularly optimized for YOLOv5 [117,118]. It addresses key issues in one-stage semi-supervised detection—such as low-quality pseudo labels [50–52] and conflicts among multiple detection tasks [119]. To improve training stability and pseudo-label accuracy, One Teacher introduces Multi-view Pseudo-label Refinement (MPR) [120] and Decoupled Semi-supervised Optimization (DSO) [121]. Together, these components reduce noise in pseudo supervision and enable more effective teacher–student learning. This method leverages a single teacher model for semi-supervised learning, primarily tested on COCO 1/5/10% splits for real-time applications. It employs Focal Loss for classification, GIoU Loss for bounding box regression, consistency loss between weak and strong augmentations, and soft pseudo-labeling loss to guide the student model, as illustrated in Figure 3.

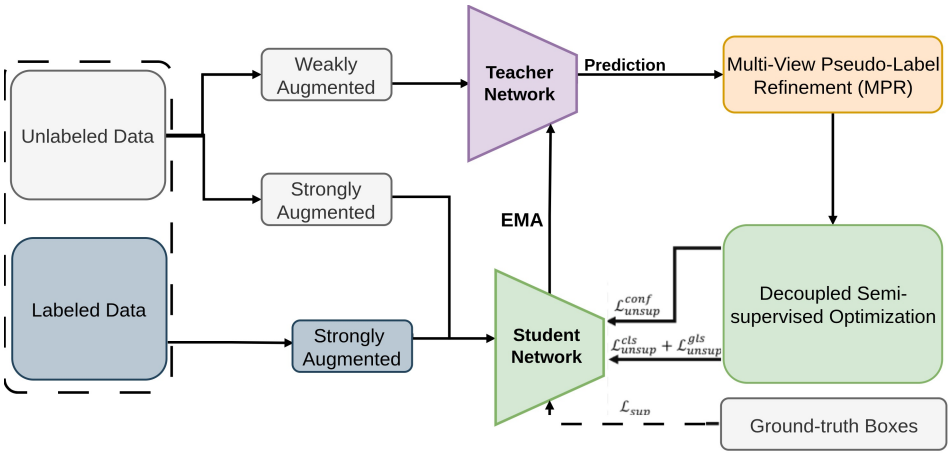

Figure 3. Framework of One Teacher [99]: The teacher generates high-quality pseudo labels for the student via Multi-view Pseudo-label Refinement (MPR) and is updated via EMA. Decoupled Semi-supervised Optimization (DSO) handles multi-task conflicts. YOLOv5 serves as the base model.

2.1.2. DSL

DenSe Learning (DSL) [95] algorithm presents a approach to anchor-free SSOD. As shown in Figure 4, is designed for one stage anchor-free detector like FCOS [54], in contrast to current approaches that mainly concentrate on two stage anchor-based detectors, which are more practical for real-world applications.

DSL addresses key challenges by introducing innovative techniques such as Adaptive Filtering (AF) for precise pseudo-label assignment [91,122], Aggregated Teacher (AT) [123] for enhanced label stability, and uncertainty consistency regularization [124] for improved model generalization. DSL is an anchor-free framework built on FCOS that enhances pseudo-label assignment through adaptive filtering and teacher aggregation. It is evaluated on COCO splits and PASCAL-VOC, employing Focal Loss for classification, IoU/GIoU for regression, uncertainty consistency regularization, and adaptive pseudo-label refinement to improve detection performance.

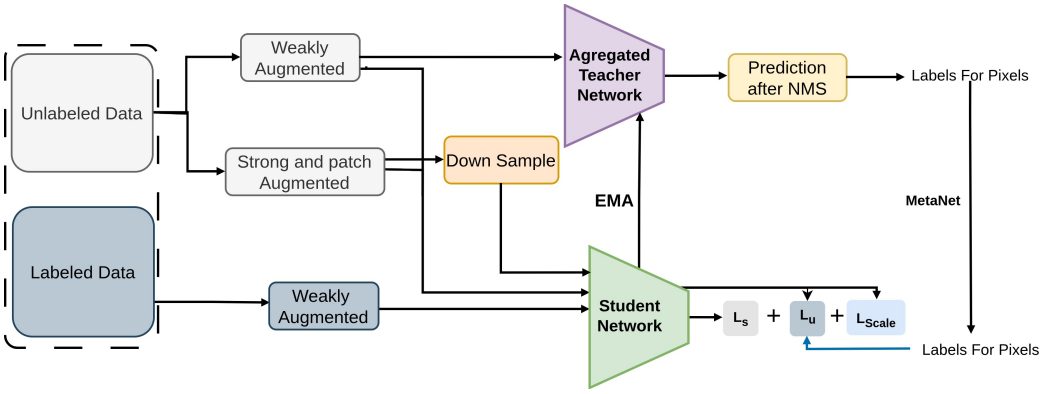

Figure 4. Framework of DSL [95]: The teacher generates pseudo labels for weakly augmented unlabeled images, refined via Adaptive Filtering and MetaNet. Patch-shuffled consistency regularization improves generalization, and the teacher is updated via Aggregated Teacher. The detector is trained with the combined loss.

2.1.3. Dense Teacher

The Dense Teacher [96] framework introduces a innovative approach to Semi-Supervised Object Detection (SSOD) by replacing sparse pseudo-boxes with dense predictions termed Dense Pseudo-Labels (DPL) [125,126], as demonstrated in Figure 5. Post-processing procedures, such as Non-Maximum Suppression [65], are not necessary for this unified pseudo-label [50,51,127] structure. Additionally, a region division strategy is proposed to suppress noise and enhance the focus on key regions, further improving detection accuracy. Overall, Dense Teacher represents a significant advancement in SSOD with its streamlined pipeline and effective utilization of dense pseudo-labels [125,126]. Dense Teacher, compatible with FCOS or RetinaNet, replaces box-level pseudo-labels with full dense predictions, removing the need for NMS. Tested on COCO and VOC, it uses dense pseudo-label supervision, Focal and GIoU losses for the student, and region-division noise suppression to provide richer training signals.

2.1.4. Unbiased Teacher v2

Unbiased Teacher v2 [98] introduces an innovative method that extends the scope of SSOD techniques [78,79,81,87] to anchor-free detectors, alongside the introduction of the Listen2Student mechanism to unsupervised regression loss [78,80] is depicted in Figure 6. Key contributions include expanding the applicability of SSOD to both anchor-based and anchor-free detectors [128], developing a mechanism to address misleading instances in regression pseudo-labels [50,51,127], and reducing performance differences between anchor-free and anchor-based detectors [128] in the semi-supervised domain. An anchor-free and general one-stage method, Unbiased Teacher v2 reduces regression errors using mechanisms like Listen2Student unsupervised regression loss. It also applies Focal Loss for classification, IoU/Smooth-L1 for regression, and weak-strong consistency loss, extending SSOD capabilities to anchor-free detectors on COCO and VOC datasets.

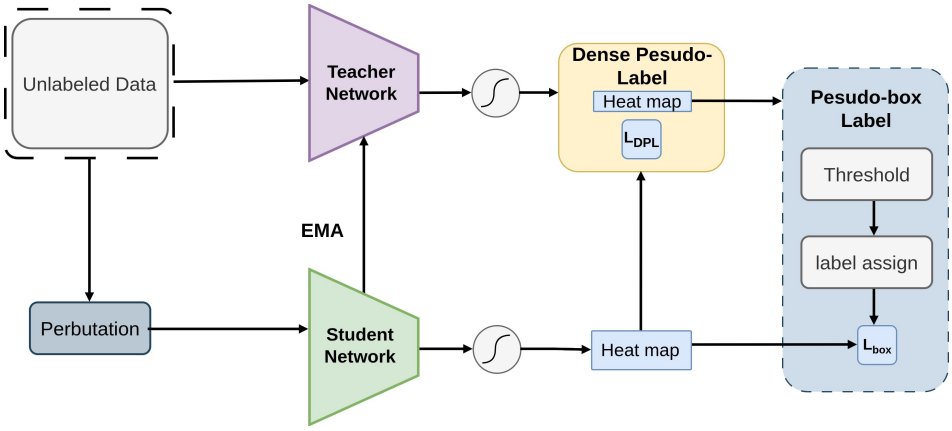

Figure 5. Framework of Dense Teacher [96]: The teacher generates Dense Pseudo-Labels (DPL) for unlabeled images, which guide the student on perturbed inputs. DPL retains rich teacher information, and the total loss combines supervised and unsupervised components.

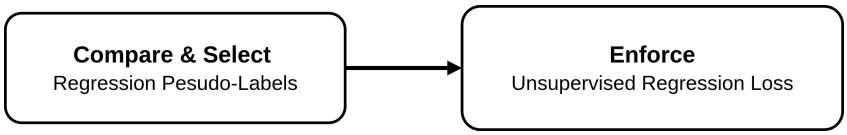

Figure 6. Framework of Unbiased Teacher v2 [96]: Listen2Student improves unsupervised regression by selecting pseudo-labels where the teacher is more confident than the student. Anchor-free detectors show smaller gains from pseudo-labeling compared to anchor-based detectors.

2.1.5. S4OD

S4OD [97], a semi-supervised methodology tailored for one stage detectors, addresses the challenge of extreme class imbalance [129] inherent in these detectors compared to their two stage SSOD [78,79,81]. Shown in Figure 7, S4OD introduces the Dynamic Self-Adaptive Threshold (DSAT) strategy [130]. S4OD dynamically determines pseudo-label selection [50–52], balancing label quality and quantity in the classification branch. Additionally, the NMS-UNC module evaluates regression label quality by computing box uncertainties via Non-Maximum Suppression [65], enhancing regression targets [81,82]. S4OD addresses extreme class imbalance in one-stage detectors by dynamically adjusting pseudo-label quality. It is evaluated on COCO and VOC datasets, combining standard Focal and GIoU losses with NMS-based uncertainty loss and dynamic self-adaptive thresholding for pseudo-label selection.

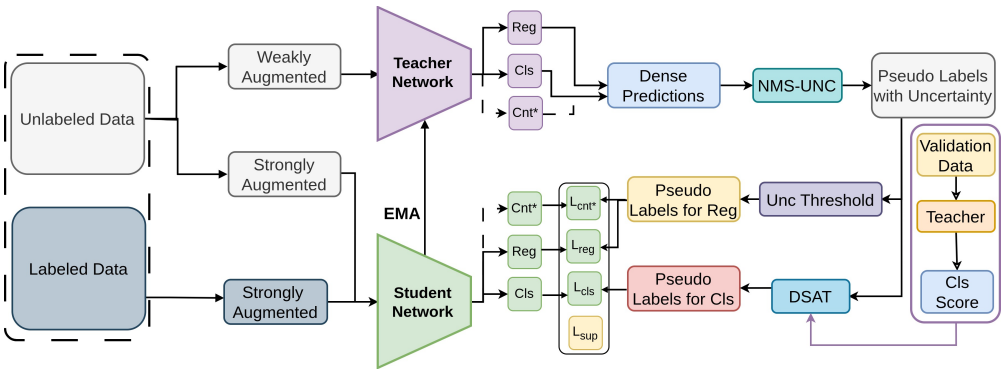

Figure 7. Framework of S4OD [97]: Regression (Reg), Classification (Cls), and Centerness (Cnt*) outputs are used. NMS-UNC generates sparse pseudo labels and retains high-quality regression boxes using an uncertainty threshold. DSAT dynamically selects high-confidence classification boxes based on F1 distribution.

2.1.6. Consistent-Teacher

Inconsistent pseudo labels [50–52] in Semi-Supervised Object Detection (SSOD) pose a challenge that Consistent-Teacher [94] addresses. These pseudo labels introduce noise into the student's training process, which causes serious overfitting [131] problems and compromises the construction of accurate detectors.

As represented in Figure 8, Consistent-Teacher introduces a 3D feature alignment module (FAM- 3D) [132], the Gaussian Mixture Model (GMM), and adaptive anchor assignment (ASA) [118,133] as a strategy to minimize this issue. These components enhance the quality of the pseudo-boxes, dynamically modify the threshold values, and stabilize the pseudo-box matching with anchors. Designed to address inconsistent pseudo-targets, Consistent-Teacher applies 3D feature alignment and adaptive thresholding. On COCO, it uses Gaussian Mixture Model thresholding loss, feature-alignment consistency loss, and adaptive anchor assignment loss to stabilize student learning.

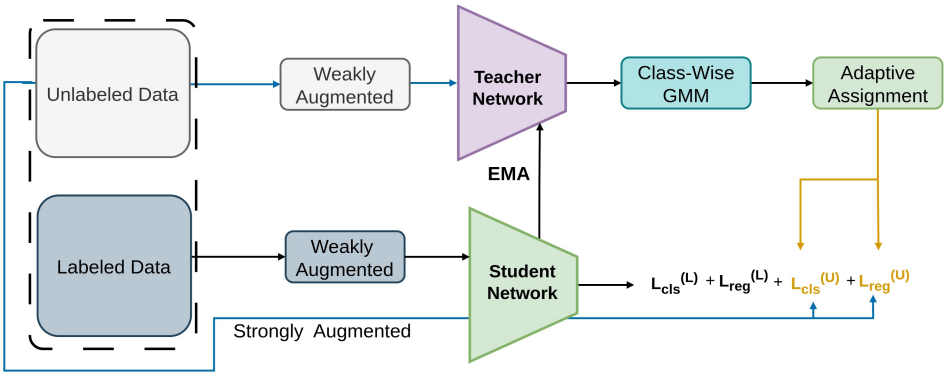

Figure 8. Framework of Consistent Teacher [94]: GMM sets dynamic thresholds, 3D feature alignment improves regression quality, and Adaptive Assignment assigns anchors based on matching cost.

2.2. Two Stage

2.2.1. Rethinking Pse

Rethinking Pse [91], as shown in Figure 9, introduces certainty aware pseudo labels [50–52] that are specifically designed for object detection. These labels accurately assess the quality of both classification and localization [134], providing a more refined method for generating pseudo labels [50–52]. By dynamically adjusting thresholds and reweighting loss functions [135] based on these certainty measurements, this mitigates the challenges posed by class imbalance [129,136–139]. This method introduces certainty-aware pseudo-labels for two-stage detectors. On COCO and VOC, it uses certainty-aware classification loss, certainty-weighted Smooth-L1 regression, and threshold adaptation loss to jointly model localization and classification confidence.

2.2.2. CSD

CSD [77] (Consistency-based Semi-supervised learning method for object Detection), which utilizes consistency constraints [140] to maximize the use of accessible unlabeled data and improve detection performance, as illustrated in Figure 10. This approach extends beyond object classification to include localization [134], ensuring comprehensive model training [134]. Additionally, this introduces Background Elimination(BE) to lessen the adverse effects of background noise on detection accuracy. CSD focuses on feature-level consistency across augmented views to reduce background noise. Evaluated primarily on VOC and COCO, it employs MSE consistency loss, standard CE and Smooth-L1 losses for labeled data, and a background elimination constraint.

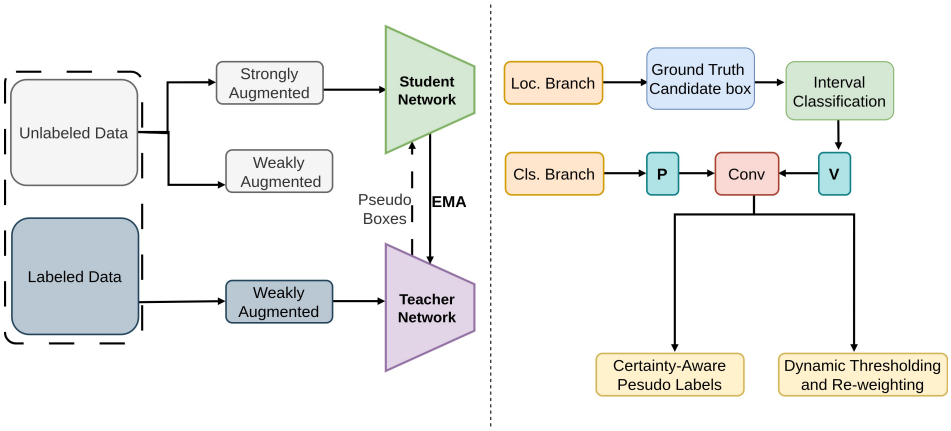

Figure 9. Framework of Rethinking Pse [91]: On the left, the teacher generates pseudo labels from labeled images to train the student. On the right, certainty-aware pseudo labels use classification and localization confidence, with dynamic thresholds and class-wise loss re-weighting to improve localization and mitigate class imbalance.

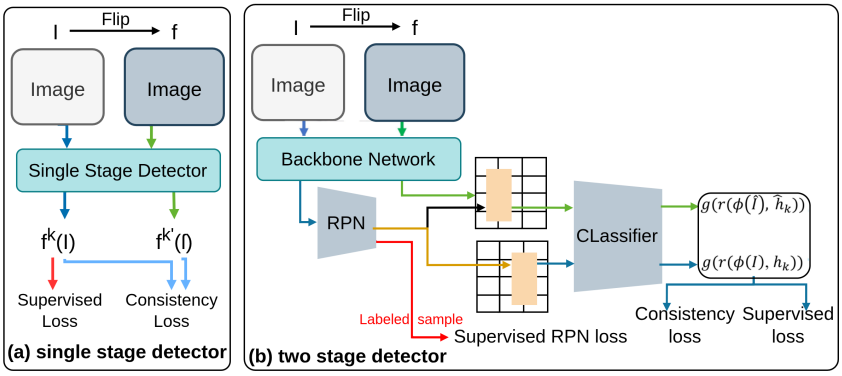

Figure 10. Framework of CSD [77]: (**a**) Single-stage detector applies supervised loss on labeled data and consistency loss between original and flipped features. (**b**) Two-stage detector aligns flipped features with original RoIs, computing supervised and consistency losses similarly. Arrows indicate data or feature flow: blue for supervised loss, green for consistency loss, and red for labeled sample flow.

2.2.3. STAC

STAC [78] is a semi-supervised [19,70] framework designed to enhance detection models for visual object recognition using unlabeled data, as shown in Figure 11. The baseline detector employed in the proposed architecture is Faster R-CNN [55]. It follows a two-step procedure where a trained detector is utilized in the first stage to generate high-confidence pseudo-labels [141] from unlabeled images. To ensure consistency and robustness, the model undergoes further training in the second stage using labeled and pseudo-labeled data along with significant data augmentations [46,100]. STAC combines augmentation-driven consistency regularization [142] and self-training [143,144] to extend the state-of-the-art SSL from image classification [48,110] to object detection. A pioneering method for weak-strong augmentation pipelines, STAC uses hard pseudo-label CE loss and Smooth-L1 regression on COCO and VOC, with augmentation-driven consistency to improve student model performance.

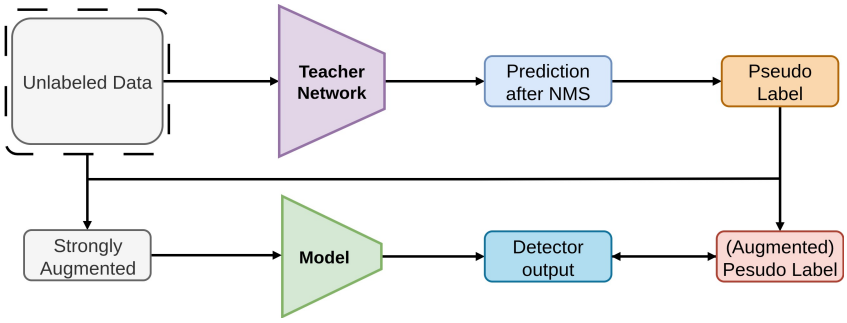

Figure 11. Overview of STAC [78]: Pseudo labels are generated from the teacher model using test-time inference and NMS. Unsupervised loss is applied to high-confidence pseudo labels, with strong augmentations ensuring consistency and target boxes adjusted for global transformations.

2.2.4. Humble Teacher

Humble Teacher [79] proposes semi-supervised approach for contemporary object detectors, utilizing a teacher–student dual model framework, as illustrated in Figure 12. The method incorporates dynamic updates to the teacher model through exponential moving averaging (EMA) [145], employs soft pseudo-labels and multiple region proposals as training targets for the student, and utilizes a detection-specific data ensemble to generate more dependable pseudo-labels. Unlike existing approaches such as STAC [78], which rely on hard labels for sparsely selected pseudo samples, the method leverages soft-labels on multiple proposals, allowing the student to distill richer information from the teacher [78]. Humble Teacher relies on soft pseudo-labels with ensemble-like teacher updates for robust supervision. Tested on COCO and VOC, it uses KL divergence for soft label distillation along with CE and Smooth-L1 for labeled samples.

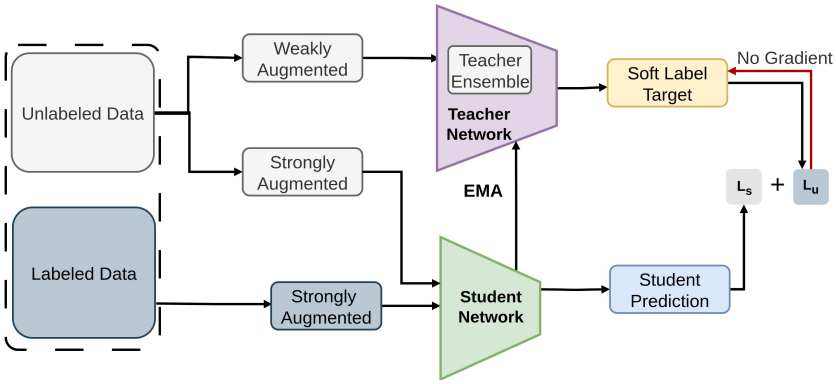

Figure 12. Overview of Humble Teacher [79]: The teacher generates soft pseudo-labels for the student, and its parameters are updated using exponential moving average (EMA).

2.2.5. Combating Noise

The proposal outlined in Combating Noise [84] introduces a method resilient to noise by measuring region uncertainty to mitigate the negative impacts of noisy pseudo-labels [146,147]. With this method, the effects of noisy pseudo-labels are carefully examined, and a metric for measuring region uncertainty is ultimately developed.

By incorporating this metric into the learning framework [148], an uncertainty-aware soft target can be formulated to prevent performance degradation caused by noisy pseudo-labeling [146], as illustrated in Figure 13. Additionally, it mitigates overfitting [131] by allowing multi-peak probability distributions and removing competition among classes. Focused on uncertainty-aware supervision, this method applies region-uncertainty loss and KL divergence for multi-peak distributions on COCO and VOC to handle noisy pseudo-labels.

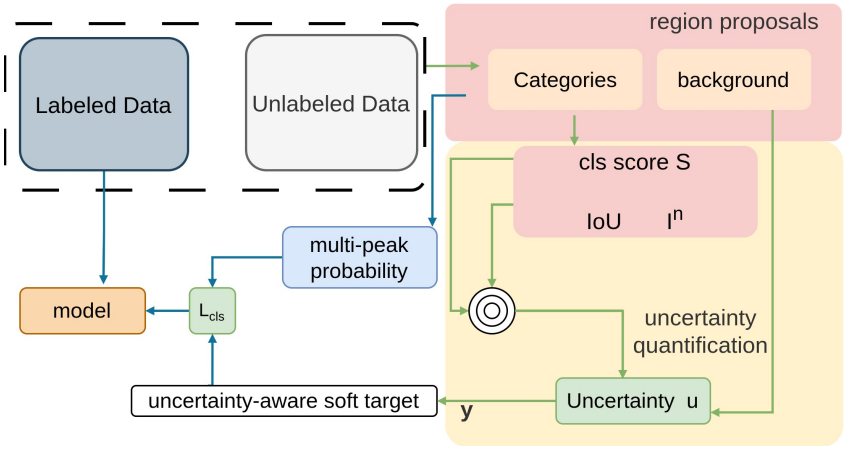

Figure 13. Framework of Combating Noise [84]: Uncertainty is quantified for different regions, and uncertainty-aware soft targets with multi-peak probability distributions are used to incorporate uncertainty into training, enabling noise-resistant learning.

2.2.6. ISMT

A Semi-Supervised Object Detection technique known as Interactive Self-Training with Mean Teachers (ISMT) [86] introduces an approach to rectify the oversight of inconsistencies among detection outcomes in the same image across various training iterations, as shown in Figure 14. By utilizing non maximum suppression [65] to combine detection outcomes from different iterations and employing multiple detection heads to offer complementary information, this approach boosts the stability and quality of pseudo labels. Moreover, the incorporation of the mean teacher model [145] prevents overfitting [131] and aids in the transfer of knowledge between detection heads. ISMT employs multi-head ensemble distillation and iterative NMS-based pseudo-label correction on COCO. It uses Smooth-L1 regression to refine student predictions while improving supervision reliability.

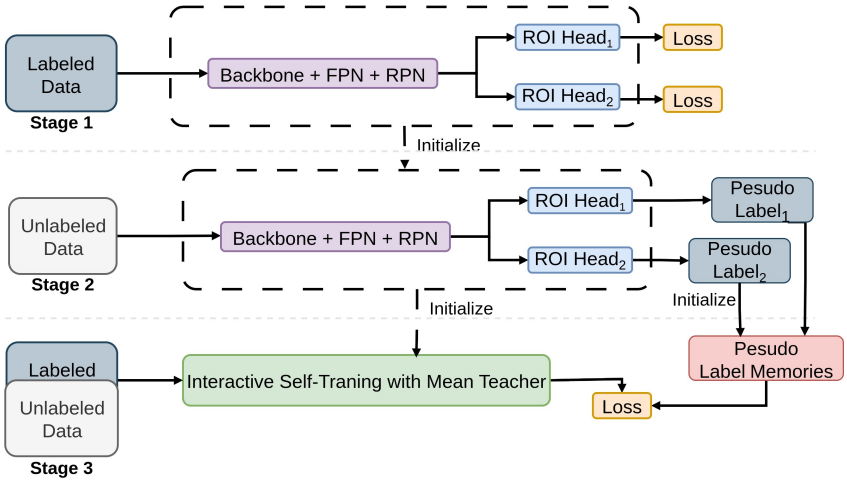

Figure 14. Framework of ISMT [86]: A detection model is first trained on labeled data to generate initial pseudo labels stored in a memory. Semi-supervised training uses the pretrained model with interactive self-training via the mean teacher method. After training, only a single ROI head is needed for inference.

2.2.7. Instant-Teaching

Instant-Teaching [80] leverages instant pseudo labeling [50–52] and extended weak-strong data augmentations [47,103] throughout each training iteration to overcome the limitations of manual annotations in typical supervised object detection frameworks. The system implements Instant-Teaching, a co-rectify approach [87], to improve pseudo annotation

quality and reduce confirmation bias [145], as depicted in Figure 15. Instant-Teaching minimizes confirmation bias by generating instant pseudo-labels, applying CE loss, confidence filtering, and co-rectify loss. Its experiments are conducted on COCO datasets.

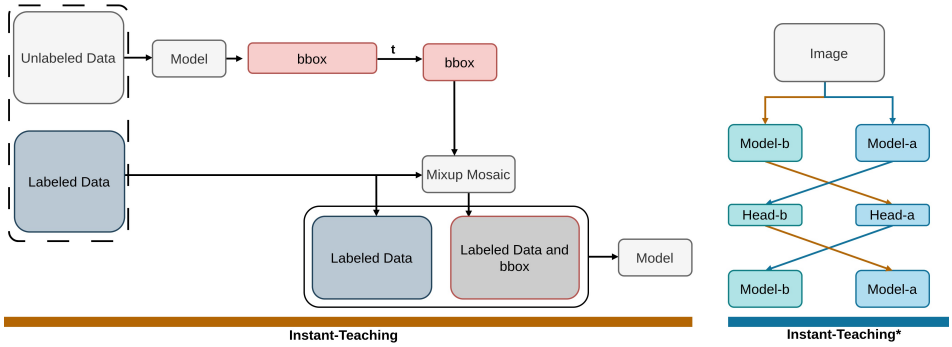

Figure 15. Overview of Instant Teaching [80]: Instant-Teaching applies instant pseudo-labeling along with extended weak and strong augmentations. Instant-Teaching* refers to the approach combined with the co-rectify scheme.

2.2.8. Soft Teacher

In contrast to earlier multi-stage approaches, Soft Teacher [81] introduces an end-to-end solution for Semi-Supervised Object Detection. The object detection training efficiency is increased by this new framework, which progressively enhances pseudo label [50–52] attributes during training [78,149]. As shown in Figure 16, this framework proposes two straightforward yet efficient methods: a box jittering methodology [150] for choosing robust pseudo boxes for box regression learning [151], and a soft teacher mechanism involving classification loss is balanced by the classification score from the teacher network. Soft Teacher emphasizes soft classification guidance and weak-strong consistency with box jittering regression loss on COCO and VOC, providing stable pseudo-supervision for one-stage and two-stage detectors.

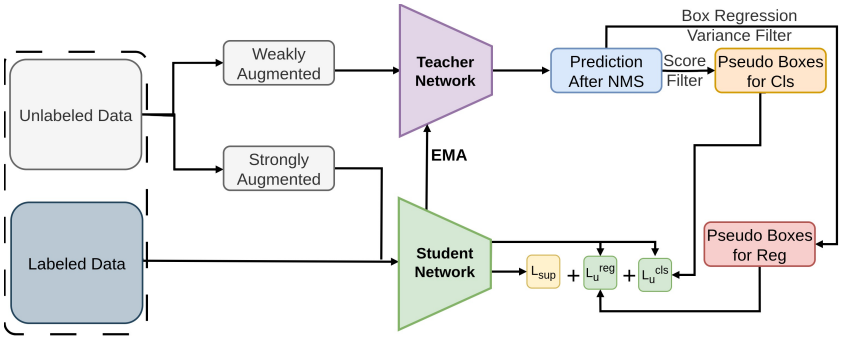

Figure 16. Overview of Soft Teacher [81]: A soft teacher generates pseudo labels on weakly augmented unlabeled images, producing separate sets for classification and regression. The teacher is updated via EMA, and the total loss combines supervised and unsupervised detection losses.

2.2.9. Unbiased Teacher

Unbiased Teacher [82] framework tackles the bias issue in pseudo-labeling [50–52], prevalent in SSOD due to class imbalances [129,136,137], as shown in Figure 17.

By collaborating to train a student and a teacher, who learns slowly, Unbiased Teacher leverages Exponential Moving Average (EMA) [152] and differential data augmentation [101,102,153] to enhance pseudo-label quality and mitigate overfitting [131]. The approach addresses key challenges in SSOD, including class imbalance and overfitting, leading to notable performance enhancements in object detection. A foundational method in SSOD, it uses balanced classification loss, strong-augmentation consistency,

and EMA-based pseudo-supervision on COCO and VOC to improve detection under limited labels.

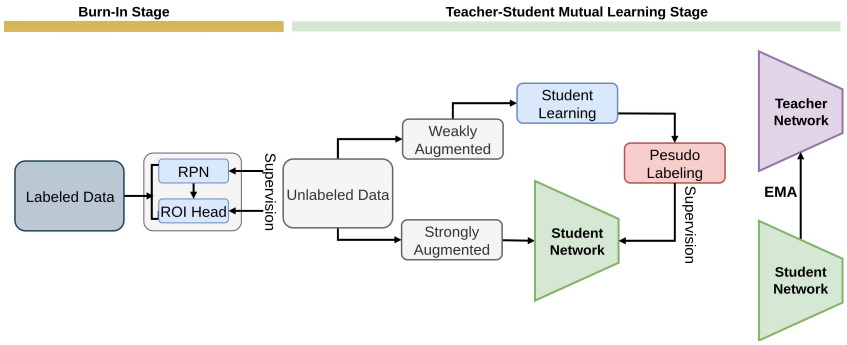

Figure 17. Overview of Unbiased Teacher [82]: Burn-In trains the detector on labeled data. In Teacher–Student Mutual Learning, the fixed teacher generates pseudo-labels for the student, and the student's knowledge is transferred back to the teacher via EMA. The teacher receives weakly augmented inputs, and the student receives strongly augmented inputs.

2.2.10. DTG-SSOD

Using the 'dense-to-dense' methodology, Dense teacher Guidance for Semi-Supervised Object Detection (DTG-SSOD) [90] utilizes dense teacher predictions directly to guide student training. As represented in Figure 18, this method is facilitated through techniques such as Inverse NMS Clustering (INC)and Rank Matching (RM) [90], allows the student model to emulate the teacher's behavior during Non-Maximum Suppression (NMS) [154], thereby receiving dense supervision without relying on sparse pseudo labels. INC clusters candidate boxes similar to the teacher's NMS process, while RM aligns the score rank of clustered candidates between the teacher and student. This dense teacher-guided approach applies dense teacher guidance loss, rank matching loss, and inverse NMS clustering supervision on COCO datasets, improving pseudo-label quality and student learning.

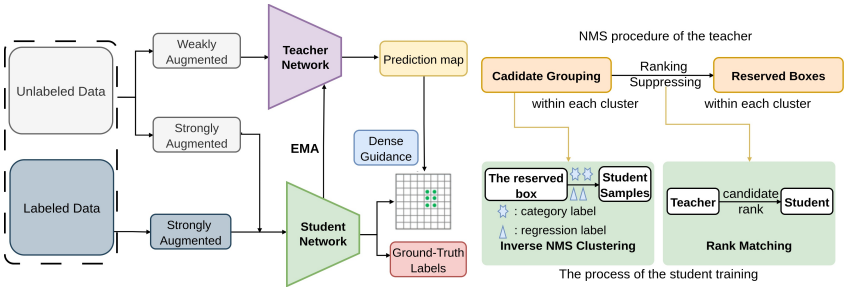

Figure 18. Framework of DTG-SSOD [90]: Training batches contain labeled and unlabeled data. For unlabeled data, the teacher provides dense guidance rather than sparse pseudo labels to supervise the student. The teacher's NMS procedure guides student training: Inverse NMS Clustering enables the student to group candidates like the teacher, and Rank Matching conveys relational information over the clustered candidates.

2.2.11. MUM

MUM [85], a data augmentation approach [101,102,153], is introduced to tackle challenges in effectively utilizing strong data augmentation strategies in SSOD due to potential adverse effects on bounding box localization [103].

As depicted in Figure 19, MUM facilitates mixing and reconstructing feature tiles from mixed image tiles, leveraging interpolation-regularization (IR) [155] for meaningful weak-strong pair generation [156].Unlike traditional SSL methods, MUM allows for the preservation of spatial information crucial for accurate object localization. MUM en-

forces interpolation regularization and consistency across mixed patches, tested on COCO, to increase student robustness to spatial transformations.

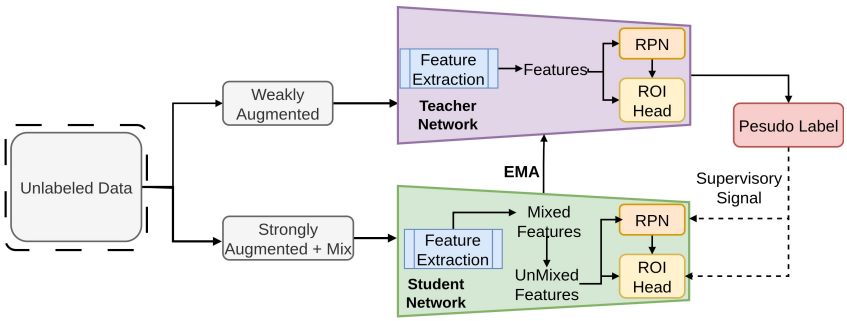

Figure 19. Framework of MUM [85]: The teacher generates pseudo labels to supervise the student. Weakly augmented inputs go to the teacher, and strongly augmented mixed inputs go to the student. Mixed features are unmixed for the student's detection head, and the teacher is updated via EMA of the student's weights

2.2.12. Active Teacher

Iteratively extending the teacher–student structure, the Active Teacher [92] method is used for Semi-Supervised Object Detection (SSOD), as demonstrated in Figure 20. Active Teacher addresses the challenge of data initialization in SSOD by gradually augmenting [45–47] the label set through an active sampling strategy, considering factors such as difficulty, information, and diversity of unlabeled examples. Active Teacher significantly enhances the performance of SSOD by maximizing the utility of limited label information and improving the accuracy of pseudo-labels [50–52]. Active Teacher introduces active sampling to select informative pseudo-labels, applying weighted CE and pseudo-label regression with EMA updates on COCO.

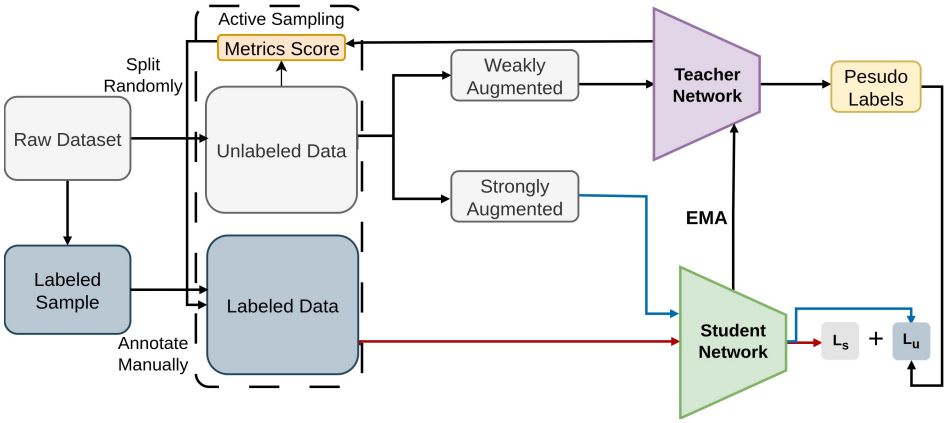

Figure 20. Framework of Active Teacher [92]: Partially initialized labels are gradually expanded. Teacher generates pseudo labels and is updated via EMA, while the Student is trained on both ground-truth and pseudo labels. The Teacher also identifies unlabeled examples for active sampling.

2.2.13. PseCo

Two essential strategies, pseudo-labeling and consistency training (PseCo) [76], in Semi-Supervised Object Detection (SSOD), highlight the shortcomings of these approaches in terms of efficiently using unlabeled data for learning. Specifically, while existing pseudo labeling [50–52] approaches focus solely on classification scores, neglecting the precision of pseudo boxes localization, [134,157] and commonly adopted consistency training methods overlook feature-level consistency crucial for scale invariance. To address these limitations, Noisy Pseudo box Learning (NPL) [146,147] is proposed for robust

pseudo label generation and Multi-view Scale-invariant Learning (MSL) [158] is introduced to ensure both label consistency and feature-level consistency, shown in Figure 21.

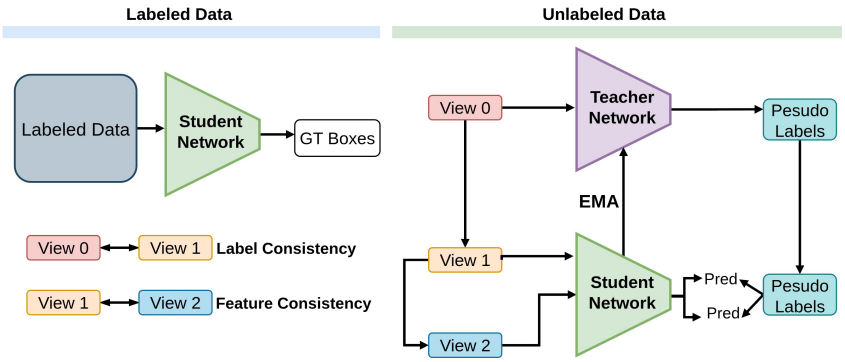

Figure 21. Framework of PseCo [76]: Training batches combine labeled and unlabeled images. The student model is trained on two augmented views (V1 and V2) of the unlabeled data, guided by the same pseudo boxes. The teacher model processes the original input images, denoted as V0.

PseCo combats noisy pseudo-boxes using multi-view scale-invariant consistency loss, alongside CE and Smooth-L1 for supervised data, evaluated on COCO and VOC.

2.2.14. CrossRectify

CrossRectify [87] is a detection framework designed to enhance the accuracy of pseudo labels [50–52], by concurrently training two detectors with different initial parameters, as depicted in Figure 22. By utilizing the disparities between the detectors, CrossRectify implements a cross-rectifying mechanism [87] to identify and improve pseudo labels, thereby addressing the inherent constraints of self-labeling [159] techniques. Extensive experiments conducted across 2D [59] and 3D [160] detection datasets validate the efficacy of CrossRectify in surpassing existing Semi-Supervised Object Detection methods. CrossRectify improves pseudo-label reliability through cross-rectification disagreement loss and consistency KL loss on COCO.

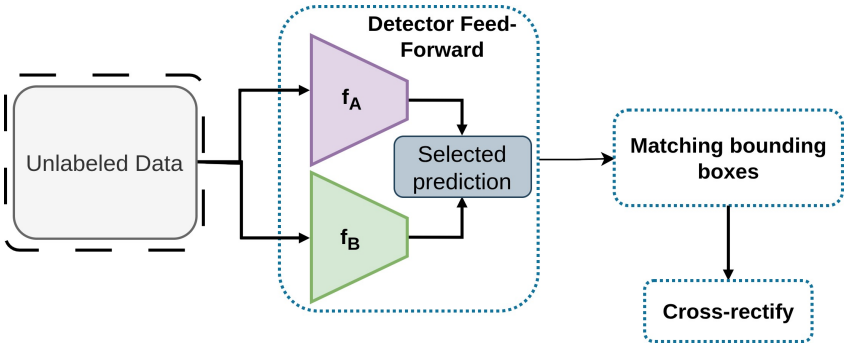

Figure 22. Framework of Cross Rectify [87]: Mechanism of pseudo label generation in the presented framework.

2.2.15. Label Match

Label mismatch is tackled from both distribution-level and instance-level perspectives through the Label Match [89] architecture, shown in Figure 23. A re-distribution mean teacher [145] employs adaptive label-distribution-aware [161] confidence criteria for unbiased pseudo-label [162] creation to address distribution-level incompatibilities [81,82]. By incorporating student suggestions into the teacher's guidance, a proposal self-assignment technique resolves instance-level mismatches stemming [163,164] from label assignment uncertainty.

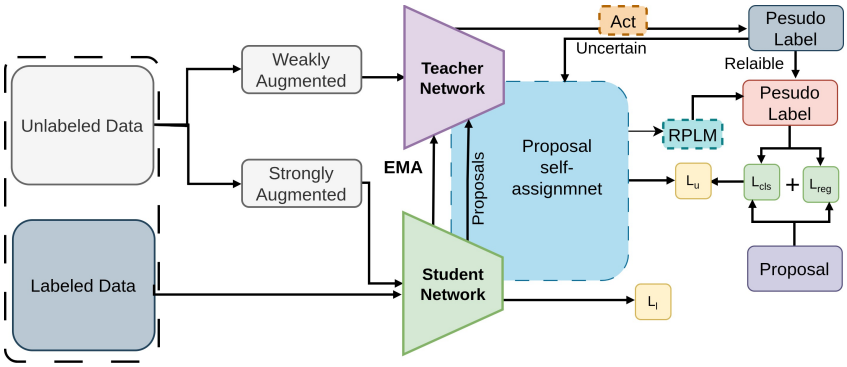

Figure 23. Framework of Label Match [89]: Labeled data train the student with supervised loss. For unlabeled data, the teacher generates pseudo labels using adaptive confidence thresholds (ACT). Reliable labels train the student directly, while uncertain labels are guided by proposal self-assignment. High-quality uncertain labels are gradually upgraded via Reliable Pseudo Label Mining (RPLM).

Furthermore, the utilization of a reliable pseudo label mining technique [165] enhances efficiency by converting ambiguous pseudo-labels into dependable ones. Label Match leverages adaptive label-distribution-aware loss, proposal self-assignment, and reliable pseudo-label mining to improve semi-supervised training on COCO.

2.2.16. ACRST

Adaptive class-rebalancing self-training, or ACRST [83], as illustrated in Figure 24, introduces a new memory module called CropBank to address the major problem of class imbalance [136,137] in SSOD. In SSOD, class imbalance [138,139], especially foreground-background and foreground-foreground imbalances—presents serious difficulties that impact the quality of pseudo-labels [50–52] and the performance of resulting models. By incorporating foreground examples from the CropBank, ACRST dynamically rebalances the training data, thereby reducing the effects of class imbalance. ACRST addresses class imbalance with foreground–background and foreground–foreground rebalancing, along with two-stage pseudo-label filtering loss on COCO and VOC.

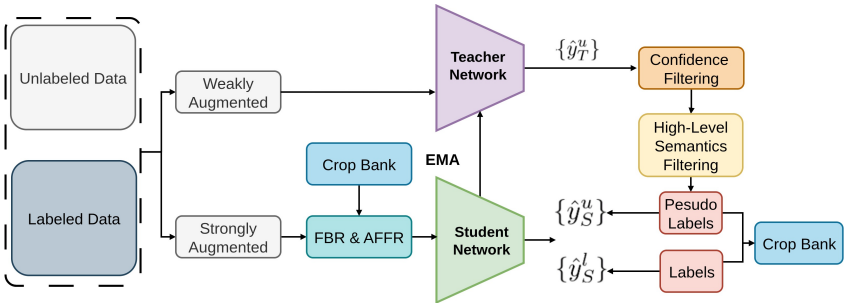

Figure 24. Framework of ACRST [83]: The teacher generates pseudo labels from weakly augmented data, while the student is trained with both ground-truths and pseudo labels. CropBank stores annotations to perform class rebalancing (FBR and AFFR) on strongly augmented data, and a two-stage pseudo-label filtering strategy further improves training.

Additionally, to tackle the issue of noisy pseudo-labels [146,147] in SSOD, a two-stage filtering technique [166] is suggested to produce accurate pseudo-labels.

2.2.17. SED

An innovative method called Scale-Equivalent Distillation (SED) [88] introduces an end-to-end knowledge distillation framework [167] that is both straightforward and efficient. SED diminishes noise from erroneous negative data, enhances localization accuracy,

and deals with high object size variance by enforcing scale consistency regularization [124], as represented in Figure 25.

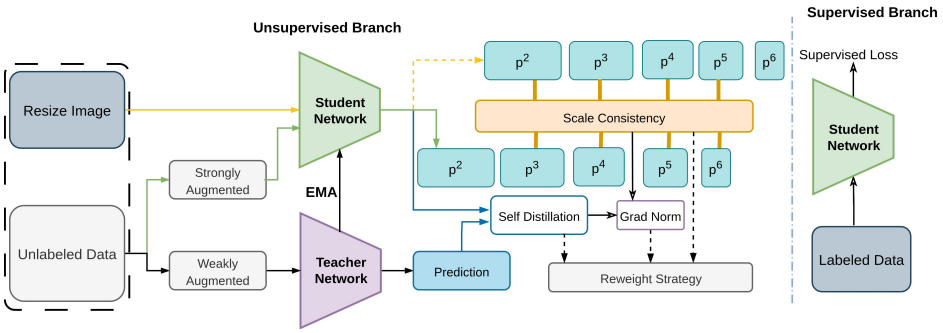

Figure 25. Framework of SED [88]: An FPN-based detector (P2–P6) is used, where the supervised branch shares the student model with the unsupervised branch. 'sg' indicates teacher predictions excluded from gradient updates, and Scale Consistency Regularization enforces consistency across feature levels.

Furthermore, a re-weighting technique [168] effectively minimizes class imbalance [136–139] by implicitly identifying potential foreground areas from unlabeled data. SED applies scale consistency regularization and cross-scale distillation, with re-weighted classification loss, on COCO datasets to improve multi-scale detection.

2.2.18. SCMT

The objective of Self-Correction Mean Teacher(SCMT) [93] is to reduce the negative impact of noise present in pseudo-labels [50–52] by dynamically modifying loss weights for box candidates. Depicted in Figure 26, SCMT effectively prioritizes more reliable box candidates during training by utilizing confidence scores derived from both localization accuracy [134] and classification scores. This novel approach outperforms existing methods [78,79,82], demonstrating its potential to improve the performance of object detection models in real-world contexts. SCMT incorporates self-correction weighting and confidence-based localization weighting alongside standard R-CNN losses, tested on COCO, for robust student training.

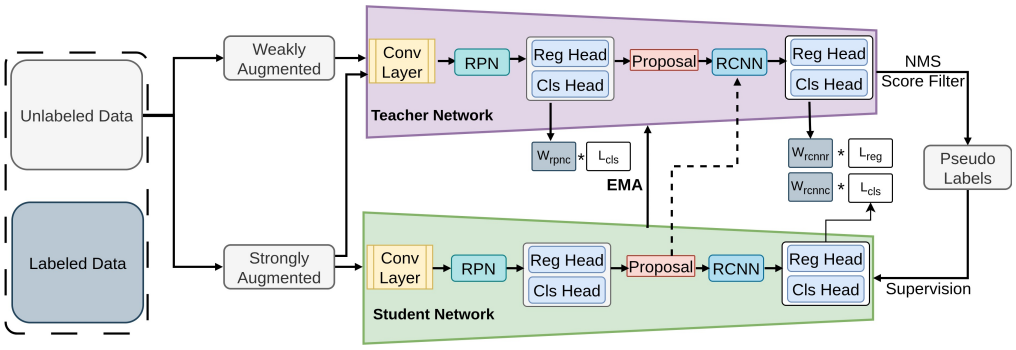

Figure 26. Framework of Self-Correction Mean Teacher [93]: Pseudo-labels are generated from weakly augmented unlabeled data using the teacher model. The unsupervised loss is computed with self-correction weights, indicated by the black dashed line.

2.3. End to End

2.3.1. Omni-DETR

In order to improve detection accuracy while lowering annotation costs, the Omni-DETR [75] framework is shown in Figure 27, incorporates a variety of weak annotations [169], including picture tags, item counts, and points.

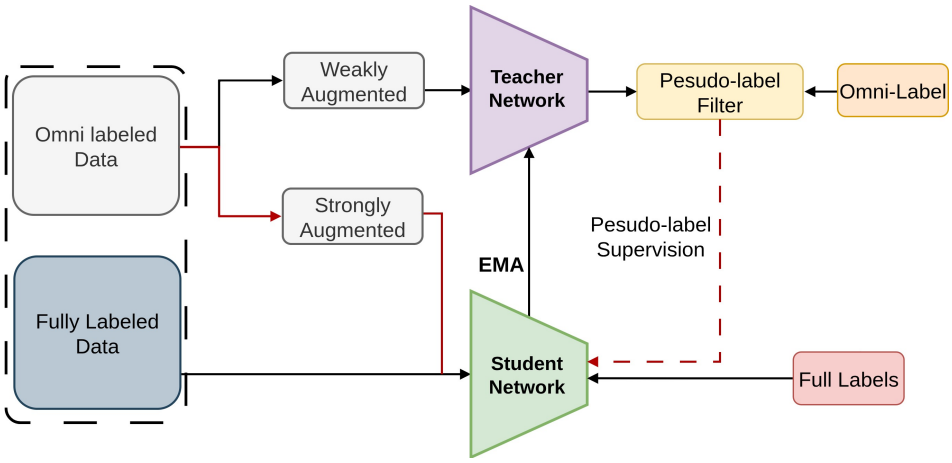

Figure 27. Framework of Omni-DETR [75]: omni-labels filter teacher predictions through a unified pseudo-label filter to generate pseudo-labels for student training.

By integrating recent developments in end-to-end transformer-based detection architecture [170,171] and student-teacher-based Semi-Supervised Object Detection [78,82], Omni-DETR enables the use of unlabeled and poorly labeled data to produce precise pseudo labels [50–52]. Omni-DETR integrates DETR-style bipartite matching with pseudo-label filtering for both COCO and weakly annotated datasets, using CE and GIoU for labeled data.

2.3.2. Semi-DETR

Semi-DETR [73] employs a Stage-wise Hybrid Matching strategy [172] to combine one-to-one [74] and one-to-many [173] assignment strategies, enhancing training efficiency and providing high-quality pseudo-labels. [50–52]. As represented in Figure 28, a Cross-view Query Consistency method [174] eliminates the need for deterministic query correspondence, facilitating the learning of semantic feature invariance. Additionally, the Cost-based Pseudo Label Mining [165] module dynamically identifies reliable pseudo boxes for consistency learning. Semi-DETR employs hybrid matching (one-to-one and one-to-many), cross-view query consistency, and cost-based pseudo-label mining on COCO to improve semi-supervised learning.

2.3.3. Sparse Semi-DETR

Sparse Semi-DETR [74], an end-to-end Semi-Supervised Object Detection system based on transformers. This solution deals with problems regarding the quality of object queries in particular and resolves them. Training efficiency is slowed and model performance is gets worse by inaccurate pseudo-labels [75] and redundant predictions, especially for tiny or obscured objects. As illustrated in Figure 29, to improve object query quality and greatly increase detection capabilities for tiny and partially obscured objects.

Sparse Semi-DETR includes a Query Refinement Module [175]. Robust pseudo-label filtering modules further improve detection accuracy and consistency by filtering only high-quality pseudo-labels [80,81]. Sparse Semi-DETR refines queries and filters unreliable pseudo-labels using Smooth-L1/GIoU losses on COCO, improving transformer-based detection.

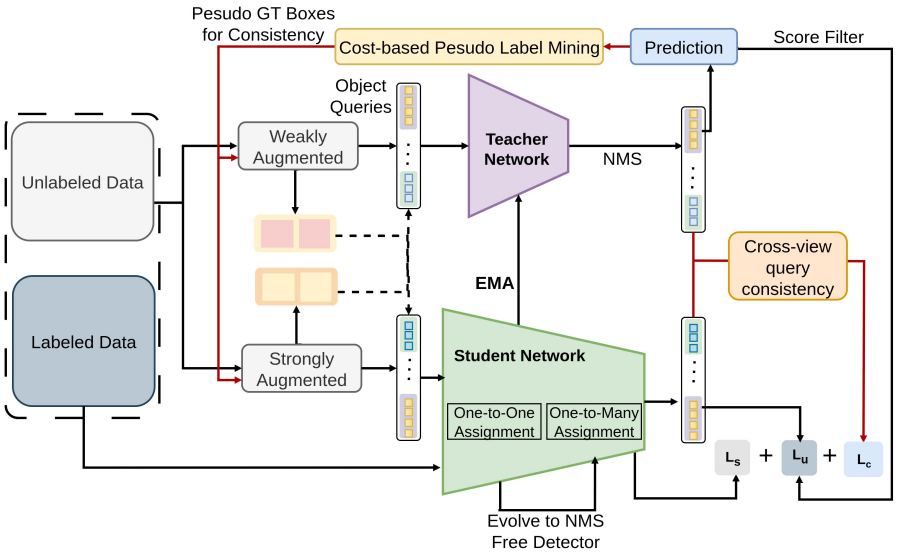

Figure 28. Framework of Semi-Detr [73]: Multi-stage training generates high-quality pseudo labels using Hybrid Matching, followed by one-to-one training. Cross-view query consistency loss with GMM-filtered pseudo boxes enhances overall training consistency.

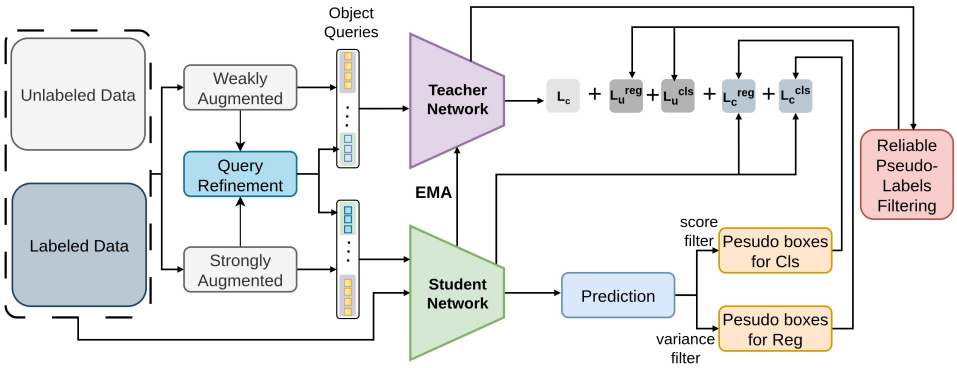

Figure 29. Framework of Sparse Semi-Detr [74]: labeled data trains the student with supervised loss, while the teacher generates pseudo-labels from weakly augmented unlabeled data. Query refinement prevents incorrect matching, and Reliable Pseudo-label Filtering progressively removes low-quality pseudo-labels.

2.3.4. STEP-DETR

STEP-DETR [176] introduces a new paradigm for transformer-based semi-supervised object detection by integrating a Super Teacher model with pseudo-label guided text queries to enhance DETR's reasoning and robustness. As shown in Figure 30, STEP-DETR [176] enriches the detection process by converting high-confidence pseudo-labels [50–52] into textual descriptions that serve as semantic prompts for the detector's query embeddings. This cross-modal guidance enables the model to better align object queries [175] with meaningful semantic cues, significantly reducing ambiguity in query-to-object matching. The Super Teacher provides high-quality pseudo-labels through multi-scale and multi-augmentation fusion, which are then transformed into structured text prompts to refine the student model's query initialization and attention patterns. STEP-DETR introduces text-guided query alignment, supervised DETR losses, and Super Teacher pseudo-label filtering for COCO and weak supervision, enabling cross-modal semi-supervised detection.

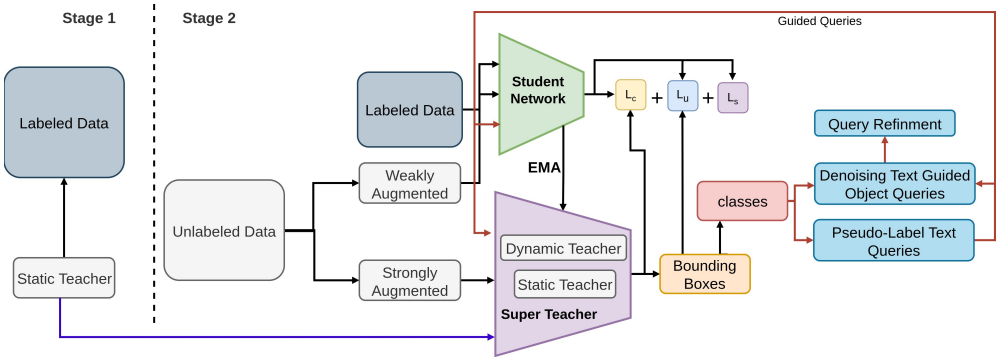

Figure 30. Overview of STEP-DETR [176]. A static teacher is trained on labeled data, while the student learns from both labeled and augmented unlabeled images. The Super Teacher provides high-quality pseudo-labels, aided by text-guided queries and query refinement. Supervised, unsupervised, and consistency losses jointly enhance detection performance.

3. Loss Function

In semi-supervised object detection (SSOD), various loss functions are used to handle labeled and unlabeled data effectively. Below, we provide both the mathematical formulations and the specific roles of each loss.

3.1. Smooth L1 Loss

Smooth L1 loss [89,177,178] is widely used for bounding box regression due to its robustness to outliers and noisy annotations. It applies a quadratic penalty to small errors and a linear penalty to large errors. In SSOD, Smooth L1 is applied to both ground-truth boxes and teacher-generated pseudo-boxes, making it suitable for supervising the student model even when pseudo-labels contain localization noise.

$$\mathcal{L}_{\text{Smooth L1}}(x) = \begin{cases} 0.5\,x^2, & \text{if } |x| < 1, \\ |x| - 0.5, & \text{otherwise.} \end{cases}$$

3.2. Distillation Loss

The transfer of Knowledge [179] from a teacher model based on labeled data to a student model with utilization of unlabeled samples is facilitated by distillation loss [79,180]. Distillation loss can be shown as:

$$\mathcal{L}_{\text{Distill}} = \tau^2 \,\text{KL}\left(\sigma\left(\frac{z_T}{\tau}\right) \,\middle\|\, \sigma\left(\frac{z_S}{\tau}\right)\right)$$

where τ is the temperature, and z_T, z_S are teacher and student logits. In SSOD, the teacher produces predictions for unlabeled images, and the student learns to imitate these predictions. This allows the detector to benefit from unlabeled data without explicit annotations.

3.3. Focal Loss

Focal Loss [181] addresses the severe foreground–background imbalance commonly observed in object detection [129,136–138] by reducing the contribution of easy negative samples during training. It is defined as:

$$\mathcal{L}_{\text{Focal}} = -\alpha_t (1 - p_t)^{\gamma} \log(p_t),$$

where p_t denotes the predicted probability for the ground-truth class, α_t is a weighting factor, and γ is the focusing parameter that controls how strongly easy examples are down-weighted. In the context of SSOD, focal loss is particularly beneficial because pseudo-labels

often include many easy background predictions. By suppressing their influence, focal loss prevents these trivial samples from dominating the optimization process, thereby stabilizing training and improving robustness to pseudo-label noise.

3.4. KL Divergence

Using in semi-supervised scenarios [19,28,70] to align predictions made on labeled and unlabeled data, KL divergence loss [79,84,182,183] minimizes the difference between probability distributions. The KL divergence between distributions P and Q is:

$$\mathrm{KL}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

Consistency-based SSOD methods often minimize KL divergence between predictions obtained under weak and strong augmentations of the same unlabeled image, encouraging stable and coherent estimator behavior.

3.5. Quality Focal Loss

Quality Focal Loss (QFL) [96,184] jointly models the classification probability and the localization quality. Its formulation is:

$$\mathcal{L}_{\mathrm{QFL}} = -|q-p|^{\beta}(q\log(p) + (1-q)\log(1-p))$$

In SSOD, pseudo-boxes may vary widely in quality. QFL naturally reduces the impact of low-quality pseudo-labels while emphasizing reliable teacher predictions.

3.6. Consistency Regularization Loss

The loss of regularization consistency [77,88] ensures consistency in predictions across different views of the same input data, enhancing model robustness and generalization in SSOD. It penalizes inconsistencies, prompting the model to learn invariant features [158], thereby improving performance across varied datasets.A common form is Mean Squared Error (MSE) consistency:

$$\mathcal{L}_{\mathrm{Cons}} = \|f(x) - f(\tilde{x})\|_2^2$$

3.7. Jensen–Shannon Divergence

Jensen–Shannon divergence [185,186] symmetrically measures distribution similarity. It is defined as:

$$\mathrm{JSD}(P\|Q) = \frac{1}{2}\mathrm{KL}(P\|M) + \frac{1}{2}\mathrm{KL}(Q\|M), \quad M = \frac{1}{2}(P+Q)$$

Some SSOD methods regularize predictions by minimizing JSD across predictions from multiple views or across ensemble teachers, helping avoid overconfident or inconsistent pseudo-labels.

3.8. Pseudo-Labeling Loss

Pseudo-Labeling Loss [187] enables semi-supervised learning [19,28,70] by generating labels for unlabeled samples using model predictions. The loss is applied only when the model is sufficiently confident, ensuring reliable supervision from pseudo-labels:

$$\mathcal{L}_{\mathrm{PL}} = -\mathbb{1}(\max(p) > \tau) \sum_c \hat{y}_c \log p_c$$

where $\hat{y}$ is the pseudo-label and τ is the confidence threshold. This strategy is central to teacher–student SSOD frameworks, enabling the student model to learn effectively from high-confidence predictions on unlabeled data.

3.9. Cross-Entropy Loss

The difference between the estimated probability distribution and the actual distribution of labels is measured by the Cross-Entropy Loss [78,82,188]. By encouraging the model to reduce the gap between the ground truth and the predicted probabilities, this loss increases the classification accuracy. Its formulation is:

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{c} y_c \log p_c$$

In SSOD, CE provides strong supervision for labeled data and acts as an anchor that stabilizes training when combined with pseudo-label- or consistency-based objectives.

4. Datasets and Comparison

In the object detection, having challenging datasets is crucial to ensuring fair and accurate evaluations of different algorithms.

4.1. Datasets

Publicly available object detection datasets such as MS-COCO and PASCAL-VOC have become the foundation for most SSOD benchmarks. However, their use in semi-supervised settings differs from conventional supervised tasks. Typically, a small percentage (1%, 5%, or 10%) of the dataset is treated as labeled, while the remaining images are used as unlabeled data. This split simulates real-world conditions where labeled data are scarce and costly to obtain. The Microsoft Common Objects in Context (MS-COCO) dataset contains approximately 118,000 training images and 5000 validation images across 80 object categories. For SSOD experiments, researchers often select a subset of labeled samples and use the remaining unlabeled images to explore model performance under limited supervision. MS-COCO's large scale, object diversity, and scene complexity make it a challenging benchmark for semi-supervised learning, testing both model robustness and generalization. The PASCAL Visual Object Classes (VOC) dataset, consisting of around 20 object categories, offers simpler but well-structured images for evaluating detection performance. SSOD studies often combine VOC 2007 and VOC 2012 for training, splitting the data into small labeled subsets (e.g., 10% or 20%) and using the rest as unlabeled data. This setup allows researchers to test whether semi-supervised techniques can maintain high performance with limited supervision.

In SSOD frameworks, the labeled data provide initial supervision to train a base detector, while unlabeled data are incorporated through pseudo-label generation and consistency regularization. The teacher model generates high-confidence pseudo-labels for the unlabeled set, which the student model uses for joint training. Data augmentation, both weak and strong, is crucial in this process to improve generalization and mitigate label noise. Therefore, these datasets are not only benchmarks but also active participants in the semi-supervised learning pipeline.

4.2. Comparison

The performance of object detection methods has been extensively evaluated on benchmark datasets such as COCO and PASCAL. These evaluations show the progress and effectiveness of both one-stage and two-stage detection approaches, as well as end-to-end methods, in improving detection accuracy over various training epochs. Table 2

offers the performance comparison of various methods on COCO dataset [189]. One stage methods, including One Teacher [99], DSL [95], Dense Teacher [96], demonstrate incremental improvements with increasing training epochs. As illustrated in Figure 31, subfigures (a)–(c) present a comparative visualization of these approaches on the COCO dataset. A clear upward trend is observed, where performance consistently improves from one-stage CNN-based methods to two-stage and finally to transformer-based end-to-end architectures. This confirms that transformer-based SSOD models leverage unlabeled data more effectively and achieve higher mAP with fewer labeled samples.

Two stage methods, such as Rethinking pse [91], STAC [78], and Combating Noise [84], exhibit consistent enhancement in performance metrics over epochs. Notably, DETR-based models like Omni-DETR [75] and Semi-DETR [73] showcase significant performance gains, highlighting the effectiveness of Semi-Supervised Object Detection strategies, as shown in Figure 31. The visual comparison in Figure 31c further demonstrates how transformer-based methods capture long-range dependencies and improve pseudo-label precision, resulting in stronger generalization than CNN-based detectors.

Table 3 shows the performance metrics of various object detection methods across different stages on the PASCAL dataset [190]. In the One stage, methods like S4OD [97], Dense Teacher [96], DSL [95] exhibit competitive performance in terms of AP50, AP50.95, and AP75 scores. Two-stage methods like Soft Teacher [81], Combating Noise [84], and Instant-Teaching [80] display significant variations in performance across different metrics.

Finally, end-to-end methods like Semi-DETR [73] and Sparse Semi-DETR [74] showcase significant performance, indicating the efficacy of SSOD approaches, as illustrated in Figure 32. Figure 32 provides a similar comparison on the PASCAL-VOC dataset, confirming that transformer-based SSOD models consistently outperform CNN-based counterparts even when trained with a small fraction of labeled data. The figure highlights the same progression pattern observed in COCO—demonstrating the scalability and robustness of modern end-to-end SSOD frameworks.

Overall, Figures 31 and 32 emphasize a clear research trend: Semi-Supervised Object Detection has evolved from conventional CNN-based architectures toward more efficient and accurate transformer-based models.

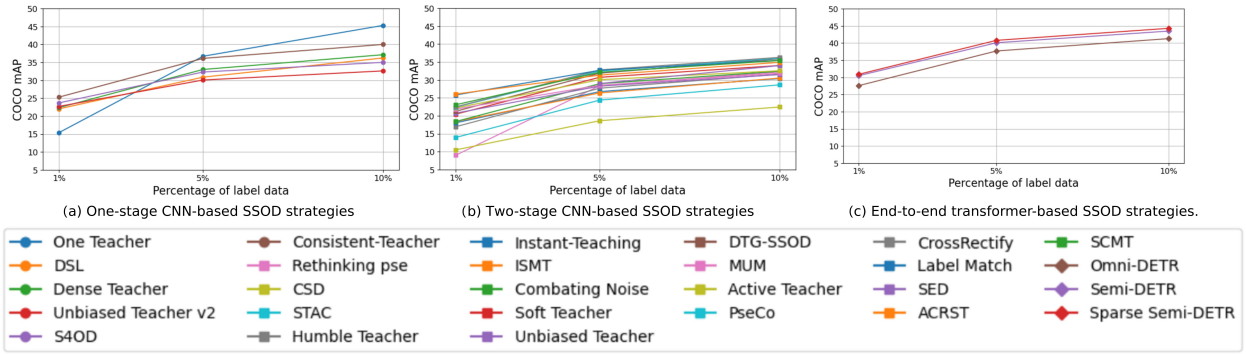

Figure 31. Comparison of CNN-based and transformer-based SSOD strategies on COCO dataset. (**a**) Performance comparison of one-stage CNN-based SSOD Strategies. (**b**) Performance comparison of two-stage CNN-based SSOD Strategies. (**c**) Performance comparison of end to end transformer-based SSOD Strategies.

Table 2. Object Detection Performance on COCO-Partial Dataset. Comparison of object detection methods across different stages on the COCO-Partial dataset.

Methods	Stages	Reference	COCO-Partial		
			1%	5%	10%
One Teacher [99]	One Stage	-	15.4	36.70	45.3
DSL [95]		CVPR22	22.03	30.87	36.22
Dense Teacher [96]		ECCV22	22.38	33.01	37.13
Unbiased Teacher v2 [98]		CVPR22	22.71	30.08	32.61
S4OD [97]		-	23.70	32.30	35.00
Consistent-Teacher [94]		CVPR23	25.30	36.10	40.00
Rethinking pse [91]	Two Stage	AAAI22	9.02	28.40	32.23
CSD [77]		ICML23	10.51	18.63	22.46
STAC [78]		-	13.97	24.38	28.64
Humble Teacher [79]		CVPR22	16.96	27.70	31.61
Instant-Teaching [80]		CVPR21	18.05	26.75	30.40
ISMT [86]		CVPR21	18.41	26.37	30.53
Combating Noise [84]		-	18.41	28.96	32.43
Soft Teacher [81]		ICCV21	20.46	30.74	34.04
Unbiased Teacher [82]		ICLR21	20.75	28.27	31.50
DTG-SSOD [90]		-	21.27	31.90	35.92
MUM [85]		CVPR22	21.88	28.52	31.87
Active Teacher [92]		CVPR22	22.20	30.07	32.58
PseCo [76]		ECCV22	22.43	32.50	36.06
CrossRectify [87]		CVPR22	22.50	32.80	36.30
Label Match [89]		CVPR22	25.81	32.70	35.49
ACRST [83]		-	26.07	31.35	34.92
SED [88]		CVPR22	-	29.01	34.02
SCMT [93]		IJCAI22	23.09	32.14	35.42
Omni-DETR [75]	End to End	CVPR22	27.60	37.70	41.30
Semi-DETR [73]		CVPR23	30.50	40.10	43.5
Sparse Semi-DETR [74]		CVPR24	30.90	40.80	44.30
STEP DETR [176]		ICCV25	31.70	41.1	45.4

Table 3. Object Detection Performance on PASCAL-VOC Dataset. Comparison of object detection methods across different stages on the PASCAL-VOC dataset.

Methods	Stages	Reference	PASCAL-VOC		
			AP_{50}	mAP	AP_{75}
S4OD [97]	One Stage	-	50.1	-	34.0
Dense Teacher [96]		ECCV22	79.89	55.87	-
DSL [95]		CVPR22	80.7	56.8	-
Consistent-Teacher [94]		CVPR23	81.00	59.00	-
Unbiased Teacher v2 [98]		CVPR22	81.29	56.87	-
One Teacher [99]		-	76.1	-	-
Soft Teacher [81]	Two Stage	ICCV21	20.46	30.74	34.04
Combating Noise [84]		-	43.2	62.0	47.5
DTG-SSOD [90]		-	56.4	-	38.8
PseCo [76]		ECCV22	57.2	-	39.2
CSD [77]		ICML23	74.70	-	-
ISMT [86]		CVPR21	77.23	46.23	-
Unbiased Teacher [82]		ICLR21	77.37	48.69	-
STAC [78]		-	77.45	44.64	-
ACRST [83]		-	78.16	50.1	-
MUM [85]		CVPR22	78.94	50.22	-
Rethinking pse [91]		AAAI22	79.0	54.60	59.4
Instant-Teaching [80]		CVPR21	79.20	50.00	54.00
Humble Teacher [79]		CVPR22	80.94	53.04	-
CrossRectify [87]		CVPR22	82.34	-	-
Label Match [89]		CVPR22	85.48	55.11	-
SED [88]		CVPR22	80.60	-	-
Semi-DETR [73]	End to End	CVPR23	86.10	65.2	-
Sparse Semi-DETR [74]		CVPR24	86.30	65.51	-
STEP DETR [176]		ICCV25	86.85	65.87	-

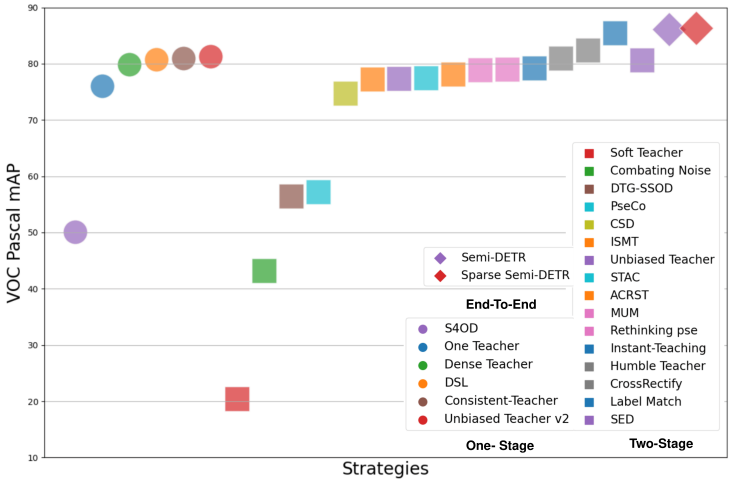

Figure 32. Comparison of CNN-based (one-stage, two-stage) and transformer-based (end-to-end) SSOD strategies on VOC pascal dataset.

5. Open Challenges & Future Directions

Although Semi-Supervised Object Detection (SSOD) has progressed rapidly, from early CNN-based pipelines to sophisticated transformer-driven architectures, this evolution reflects more than just improved accuracy or architectural complexity. A deeper examination reveals an important conceptual shift in how the field interprets uncertainty, supervision, and the role of unlabeled data, as summarized in Table 4, which provides a comparative overview of the strengths and limitations of existing semi-supervised object detection methods.

Table 4. A brief description of Advantages and limitations of Semi Supervised Strategies.

Methods	Advantages	Limitations
Stac [78]	Improves detection performance with minimal complexity.	Low performance with frameworks employing intense hard negative mining, leading to over dependence on noisy pseudo-labels.
Humble Teacher [79]	Improves performance significantly with dynamic teacher model updates and soft pseudo-labels.	More computational resources due to the dynamic updating of the teacher model and the ensemble of numerous teacher models, potentially increasing training time and complexity.
Instant Teaching [80]	Improving model learning with extended weak-strong data augmentation as well as instant pseudo labeling .	Dependency on Extensive weak-strong data augmentation and instant pseudo labeling introduce computational overhead, increase training complexity and time.
Soft Teacher [81]	Enhances detector performance and pseudo label quality simultaneously.	Depending on extensive data augmentation and the soft teacher approach potentially increase training complexity and computational overhead.
Unbiased Teacher [82]	Effectively mitigates pseudo-labeling bias and overfitting in Semi-Supervised Object Detection.	Relies on the balance between the student and teacher models, which require careful tuning and additional computational resources.
ACRST [83]	Improves performance by addressing class imbalance.	Effectiveness relies on the precision of pseudo-labels, which are impacted by noise due to the complexity of detection tasks, requiring robust filtering mechanism.
Combating Noise [84]	Effectively combating noise associated with pseudo labels enhances the robustness of the SSOD Tasks.	Dependence on accurately quantifying region uncertainty is challenging in complex scenes or datasets with diverse object characteristics.
MUM [85]	Effectively augments data for Semi-Supervised Object Detection, enhancing model robustness without significant computational overhead.	Encounter difficulties in accurately locating object boundaries due to the mixing process, potentially affecting localization precision.
ISTM [86]	Effectively leveraging ensemble learning to enhance the usefulness of pseudo labels and stabilize Semi-Supervised Object Detection training.	Introduce additional computational complexity due to the ensemble approach and the use of multiple ROI heads, potentially increasing training time and resource requirements.

Table 4. *Cont.*

Methods	Advantages	Limitations
Cross Rectify [87]	Enhances pseudo label quality and detection performance by rectifying misclassified bounding boxes using detector disagreements.	Simultaneous training of two detectors increase computational overhead, potentially prolonging training time and resource usage.
SED [88]	Improves Semi-Supervised Object Detection by enforcing scale consistency and self-distillation.	Reliance on the IoU threshold criterion, which could not be optimal for all detectors and situations, and its limited benefits from multi-scale testing
Label Match [89]	Improves Semi-Supervised Object Detection by addressing label mismatch through distribution-level and instance-level methods.	Assumes Both unlabeled as well as labeled data have the same distribution, potentially restricting its applicability in diverse scenarios.
DTG-SSOD [90]	Leverages Dense Teacher Guidance for more accurate supervision, enhancing Semi-Supervised Object Detection performance.	Implementation complexity, especially with Inverse NMS Clustering and Rank Matching, increase computational resources during training.
Rethinking Pse [91]	Certainty-aware pseudo labels improve performance by addressing localization precision and class imbalance issues	Implementing certainty-aware pseudo labeling add additional computational complexity during training.
CSD [77]	Leverages consistency constraints for both classification and localization, enhancing object detection performance using unlabeled data.	It shows less performance improvement in two-stage detectors compared to single-stage detectors.
PseCo [76]	Enhances SSOD by integrating object detection attributes into pseudo labeling along with consistency training, leading to superior performance and faster convergence.	Its potential struggle with generalization across diverse datasets due to variability in pseudo-label quality.
Active Teacher [92]	Maximizes limited label information through active sampling, enhancing pseudo-label quality and improving SSOD performance.	Require more training steps compared to other methods, potentially increasing computational overhead.
One Teacher [99]	Improves SSOD on YOLOv5, tackling issues like low-quality pseudo-labeling.	Lowering the threshold for pseudo-labeling due to noisy pseudo-labeling in one-stage detection makes it difficult to maximize the effectiveness of one-stage teacher–student learning.
Dense Teacher [96]	Simplifies the SSOD pipeline by using Dense Pseudo-Labels, improving efficiency and performance.	Contain high-level noise, potentially impacting detection performance if not properly addressed.
Unbiased Teacher v2 [82]	Expands the applicability of SSOD to anchor-free detectors, improving performance across various benchmarks.	Challenges remain in scaling the method to large datasets, integrating localization uncertainty estimation for boundary prediction with the relative thresholding mechanism, and addressing domain shift issues.
S4OD [97]	Dynamically adjusts pseudo-label selection to balance quality and quantity, enhancing single-stage detector performance	DSAT strategy's increased time cost is due to F1-score computation, and using the CPU version of NMS for uncertainty computation slows down training.
Consistent-Teacher [94]	Improves SSOD performance by addressing inconsistent pseudo-targets with feature alignment, adaptive anchor assignment, and dynamic threshold adjustment.	Performance is validated mainly on single-stage detectors, with effectiveness on stage-two detectors and DETR-based models yet to be confirmed.
Omni-DETR [75]	utilize diverse weak annotations to enhance performance and annotation efficiency.	Effectiveness on larger datasets is uncertain, and its simplified annotation process could raise concerns about potential misuse.
Semi-DETR [73]	Combines Cross-view query consistency and stage-wise hybrid matching to improve training efficiency.	Encounter challenges due to the absence of deterministic connection between the predictions and the input queries.
Sparse Semi-DETR [74]	Introduces a Query Refinement Module to improve object query functionality, enhancing detection performance for small and obscured objects.	Require additional computational resources due to the integration of novel modules, potentially increasing training time and complexity.

CNN-based SSOD methods traditionally treated pseudo-labels as unreliable approximations of ground truth. As a result, much of the research effort centered on preventing the student from overfitting to noisy teacher predictions through threshold tuning, uncertainty modeling, and ensemble-based refinement. These strategies expose a fundamental limitation: CNN detectors rely heavily on local heuristics such as anchor assignment, IoU thresholds, and NMS rules—elements that amplify pseudo-label errors in semi-supervised settings.

Transformer-based SSOD approaches offer a different paradigm. Their global attention mechanisms reduce reliance on these brittle heuristics, shifting the focus from correcting pseudo-labels to interpreting and leveraging them. Methods such as Semi-DETR, Sparse Semi-DETR, and STEP-DETR reimagine pseudo-labels as semantic cues that guide query refinement, cross-view consistency, and even text-based reasoning. This marks a conceptual turning point: SSOD is evolving from pseudo-label cleaning toward representation-level alignment and high-level semantic guidance. Yet this progress introduces new challenges that must be addressed for SSOD to mature into a practical, deployable technology. The following subsections outline the major open issues and potential research directions.

System Complexity and Deployment: Many state-of-the-art SSOD frameworks rely on multi-stage pipelines with teacher–student models, pseudo-labeling, and consistency regularization. These components improve accuracy but increase computational cost and memory usage, making real-time deployment difficult. Transformer-based SSOD models, in particular, struggle with dense attention mechanisms and iterative optimization. Lightweight architectures, model pruning, quantization, and distributed training are promising directions for reducing system complexity. Future work must explore model compression, distillation, quantization, and efficient transformer variants to balance performance with practicality. Designing lightweight SSOD systems without sacrificing robustness remains an open research front.

Maintaining Accuracy and Robustness: Noisy pseudo-labels, domain shifts, and real-world conditions such as occlusion or class imbalance can degrade performance. Techniques like adaptive pseudo-label reweighting, uncertainty-aware learning, and online teacher updates can help maintain accuracy. Continuous and domain-adaptive learning strategies further support robust generalization to unseen data. A key challenge is developing models that not only detect noise in pseudo-labels but actively correct and learn from it, closing the loop between representation learning and uncertainty estimation.

Evaluation and Benchmarking: Existing benchmarks such as COCO and PASCAL-VOC often fail to capture realistic deployment challenges. Future research should establish datasets with domain diversity, incremental annotation, and real-world constraints. Standardized evaluation protocols and reproducible pipelines are essential for fair comparison across SSOD methods. Future benchmarks should introduce mixed-quality annotations, temporal or streaming data and cross-domain unlabeled pools. Standardizing evaluation pipelines and reporting practices will help ensure fair comparisons and improve reproducibility.

Scalability and Reproducibility: The growing complexity of SSOD architectures makes training resource-intensive and hyperparameter tuning challenging. Open-source frameworks, transparent reporting, and standardized evaluation settings are crucial to ensure reproducibility and wider accessibility. To improve scalability and reproducibility, the field needs more transparent reporting of experiment settings, robust open-source implementations, standardized SSOD training recipes and automated hyperparameter tuning or adaptive schedules. Without addressing reproducibility, many SSOD innovations risk becoming difficult to validate or build upon.

Domain Adaptation and Hybrid Methods: Improving generalization through domain adaptation and transfer learning is key for real-world deployment. Hybrid approaches combining semi-supervised, self-supervised, and transfer learning, alongside model compression, can enhance both efficiency and detection performance. Hybrid approaches that blend semi-supervised learning, self-supervised learning, and domain adaptation appear promising. Cross-modal supervision as seen in STEP-DETR's text-guided queries may further enhance generalization.

SSOD has moved beyond simple pseudo-label refinement toward a more nuanced understanding of representation learning, semantic guidance, and uncertainty modeling. However, limitations in computational efficiency, robustness, benchmark realism, and training stability continue to hinder widespread adoption. Addressing these open challenges will determine whether SSOD advances from a promising research direction into a reliable, scalable technology suitable for industry-level applications.

6. Applications

6.1. Image Classification

Semi-supervised learning has significantly advanced image classification [22,23], especially in domains with limited labeled data [191]. In medical imaging [192–194], SSOD enables accurate disease diagnosis from X-rays and MRIs even with few labeled examples [195]. Similarly, remote sensing [196,197] benefits from improved classification of land cover and environmental changes, aiding urban planning and disaster management. In autonomous vehicles [198,199], it enhances object and pedestrian classification, promoting safer navigation. The primary challenge SSOD addresses in image classification is the scarcity and imbalance of labeled data across diverse visual categories. By leveraging unlabeled images, it reduces annotation costs while maintaining robustness and generalization under domain shifts. Architecturally, effective SSOD models incorporate adaptive feature extractors, teacher–student frameworks, and noise-tolerant pseudo-labeling pipelines capable of handling heterogeneous image sources and visual variations. Techniques like consistency regularization [124] and pseudo-labeling [50–52] are critical in stabilizing training and improving accuracy.

6.2. Document Analysis

SSOD is increasingly applied to document analysis [200–203], efficiently detecting and classifying text blocks, tables, and images [204–207]. In legal, financial, and academic contexts, where large volumes of documents must be processed [208–211], SSOD reduces reliance on extensive labeled datasets. The key challenges here include variability in document layouts, fonts, and noise from scanning or handwriting. SSOD addresses these by learning structure-invariant representations through multi-scale attention and contextual feature aggregation. Architecturally, systems often combine text-visual fusion modules with region-proposal refinement layers, ensuring semantic consistency across labeled and unlabeled samples. Methods like self-training [212] and consistency-based regularization [124] improve detection robustness under diverse document formats.

6.3. Three-Dimensional Object Detection

In 3D object detection [213,214], SSOD improves accuracy and robustness by leveraging both labeled and unlabeled point cloud or multi-modal data. For autonomous driving [20,21], it enhances detection of pedestrians, vehicles, and obstacles using LIDAR and camera inputs [215–218]. In robotics, it supports precise manipulation and obstacle avoidance, while in AR/VR, it ensures accurate spatial integration of virtual elements with real-world environments. Challenges include sparse or incomplete 3D data, dynamic environments, and cross-modal inconsistencies. Architecturally, effective solutions use cross-modal consistency modules, voxel-based pseudo-label refinement, and memory-efficient 3D feature backbones to enable real-time performance while maintaining robustness under partial observations.

6.4. Network Traffic Classification

SSOD is effective for network traffic classification [219,220], identifying and categorizing traffic patterns even with limited labeled data [221]. This is crucial for detecting anomalies and security threats while maintaining network performance. The main challenge is the high volume, heterogeneity, and evolving nature of network traffic. SSOD addresses these issues by exploiting unlabeled traffic data to improve detection of malicious activities [222,223]. Architecturally, models integrate temporal pattern learning, feature embedding regularization, and adaptive pseudo-labeling to maintain robustness across changing network conditions.

6.5. Speech Recognition

In speech recognition [224–227], SSOD improves transcription and phoneme classification even with limited labeled audio. It enhances the separation of speech from background noise and adapts to diverse linguistic and acoustic conditions [228–231]. Challenges include speaker variability, noisy environments, and limited labeled corpora. SSOD addresses these via cross-modal consistency modules, pseudo-label refinement, and memory-efficient feature extraction, enabling real-time transcription and scalable deployment in voice-controlled systems.

6.6. Drug Discovery and Bioinformatics

SSOD accelerates drug discovery [232,233] and bioinformatics tasks [193,234] by improving identification and classification of molecular structures [235,236] and biological entities [237,238]. It reduces reliance on scarce labeled molecular data while handling high-dimensional biological features. The challenge is the complexity and heterogeneity of biochemical data. SSOD addresses this by learning latent molecular representations that generalize across molecular variations. Architecturally, effective models use graph-based detection networks, uncertainty calibration layers, and transfer learning mechanisms [179] to integrate heterogeneous datasets, enabling scalable and interpretable molecular detection for precision medicine.

7. Conclusions

This survey presented a comprehensive overview of Semi-Supervised Object Detection (SSOD) by examining both CNN-based and Transformer-based approaches within a unified framework. The primary contribution of this work is its explicit effort to bridge these two methodological lines, which are often treated separately in existing literature. By analyzing their architectural characteristics, learning strategies, and pseudo-labeling mechanisms in parallel, the survey highlights the continuity and differences between CNN-driven and Transformer-driven designs. A second major contribution is the introduction of a structured comparative perspective through a consistent taxonomy and benchmark-oriented organization. This structure enables clearer and more systematic comparison across SSOD methods, offering a coherent view of how architectural developments shape the handling of unlabeled data. Through this, the survey identifies common patterns, gaps, and limitations in current approaches, outlining the areas where further research is needed. Overall, the shift from CNN-based to Transformer-based designs marks an important transition in SSOD. The analysis in this survey provides a consolidated reference for ongoing developments and future research directions in semi-supervised object detection.

Author Contributions: Conceptualization was carried out by T.S. and I.I.; methodology was developed by T.S. and I.I.; investigation was performed by T.S. and I.I.; writing—original draft preparation was done by T.S. and I.I.; writing—review and editing was carried out by T.S., I.I.; supervision was

provided by M.Z.A.; and project administration was handled by M.Z.A., D.S., and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: The work has been partially funded by the European project AIRISE under Grant Agreement ID 101092312.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Deng, L.; Yu, D. Deep learning: Methods and applications. In *Foundations and Trends® in Signal Processing*; Now Publishers: Boston, MA, USA; Delft, The Netherlands, 2014. [CrossRef]
2. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
3. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [CrossRef]
4. V, H. An overview of pattern recognition. *Int. J. Res. Publ. Rev.* **2022**, *3*, 1883–1889. [CrossRef]
5. Singh, C. Machine Learning in Pattern Recognition. *Eur. J. Eng. Technol. Res.* **2023**, *8*, 63–68. [CrossRef]
6. Liao, S.H.; Chu, P.H.; Hsiao, P.Y. Data mining techniques and applications—A decade review from 2000 to 2011. *Expert Syst. Appl.* **2012**, *39*, 11303–11311. [CrossRef]
7. Fang, F. A Study on the Application of Data Mining Techniques in the Management of Sustainable Education for Employment. *Data Sci. J.* **2023**, *22*, 23. [CrossRef]
8. von Luxburg, U.; Schoelkopf, B. Statistical Learning Theory: Models, Concepts, and Results. *arXiv* **2008**, arXiv:0810.4752. [CrossRef]
9. Tsai, S.C.; Chen, C.H.; Shiao, Y.T.; Ciou, J.S.; Wu, T.N. Precision education with statistical learning and deep learning: A case study in Taiwan. *Int. J. Educ. Technol. High. Educ.* **2020**, *17*, 12. [CrossRef]
10. Bebis, G.; Egbert, D.; Shah, M. Review of computer vision education. *IEEE Trans. Educ.* **2003**, *46*, 2–21. [CrossRef]
11. Canedo, D.; Neves, A.J.R. Facial Expression Recognition Using Computer Vision: A Systematic Review. *Appl. Sci.* **2019**, *9*, 4678. [CrossRef]
12. Khurana, D.; Koli, A.; Khatter, K.; Singh, S. Natural Language Processing: State of The Art, Current Trends and Challenges. *Multimed. Tools Appl.* **2022**, *82*, 3713–3744. [CrossRef]
13. Chang, K.H. Natural Language Processing: Recent Development and Applications. *Appl. Sci.* **2023**, *13*, 11395. [CrossRef]
14. Praveena, M.; Jaiganesh, V. A Literature Review on Supervised Machine Learning Algorithms and Boosting Process. *Int. J. Comput. Appl.* **2017**, *169*, 32–35. [CrossRef]
15. Nasteski, V. An overview of the supervised machine learning methods. *Horizons B* **2017**, *4*, 51–62. [CrossRef]
16. El Mrabet, M.A.; El Makkaoui, K.; Faize, A. Supervised Machine Learning: A Survey. In Proceedings of the 2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet), Rabat, Morocco, 3–5 December 2021; pp. 1–10. [CrossRef]
17. Ouali, Y.; Hudelot, C.; Tami, M. An Overview of Deep Semi-Supervised Learning. *arXiv* **2020**, arXiv:2006.05278. [CrossRef]
18. Shehzadi, T.; Azeem Hashmi, K.; Stricker, D.; Liwicki, M.; Zeshan Afzal, M. Towards End-to-End Semi-Supervised Table Detection with Deformable Transformer. In Proceedings of the Document Analysis and Recognition—ICDAR 2023: 17th International Conference, San José, CA, USA, 21–26 August 2023; Proceedings, Part II; Springer: Cham, Switzerland , 2023; pp. 51–76. [CrossRef]
19. Allabadi, G.; Lucic, A.; Pao-Huang, P.; Wang, Y.X.; Adve, V. Semi-Supervised Object Detection in the Open World. *arXiv* **2023**. [CrossRef]
20. Hwang, S.; Kim, Y.; Kim, S.; Bahk, S.; Kim, H.S. UpCycling: Semi-supervised 3D Object Detection without Sharing Raw-level Unlabeled Scenes. *arXiv* **2023**, arXiv:2211.11950.
21. Mao, J.; Shi, S.; Wang, X.; Li, H. 3D Object Detection for Autonomous Driving: A Comprehensive Survey. *arXiv* **2023**, arXiv:2206.09474. [CrossRef]
22. Ye, Z.; Li, H.; Song, Y.; Wang, J.; Benediktsson, J.A. A novel semi-supervised learning framework for hyperspectral image classification. *Int. J. Wavelets Multiresolut. Inf. Process.* **2016**, *14*, 1640005 . [CrossRef]
23. Li, S.; Kou, P.; Ma, M.; Yang, H.; Huang, S.; Yang, Z. Application of Semi-Supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data. *IEEE Access* **2024**, *12*, 27331–27343 . [CrossRef]

24. Tseng, G.; Sinkovics, K.; Watsham, T.; Rolnick, D.; Walters, T.C. Semi-Supervised Object Detection for Agriculture. In Proceedings of the 2nd AAAI Workshop on AI for Agriculture and Food Systems, Virtual, 14 February 2023. Available online: https://openreview.net/forum?id=AR4SAOzcuz (accessed on 5 May 2024).
25. Yousaf, A.; Sazonov, E. Food Intake Detection in the Face of Limited Sensor Signal Annotations. In Proceedings of the 2024 Tenth International Conference on Communications and Electronics (ICCE), Danang, Vietnam, 31 July–2 August 2024; pp. 351–356. [CrossRef]
26. Patrício, D.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [CrossRef]
27. Ngobeni, R.; Sadare, O.; Daramola, M.O. Synthesis and Evaluation of HSOD/PSF and SSOD/PSF Membranes for Removal of Phenol from Industrial Wastewater. *Polymers* **2021**, *13*, 1253. [CrossRef] [PubMed]
28. Sheikhpour, R.; Sarram, M.A.; Gharaghani, S.; Chahooki, M.A.Z. A Survey on semi-supervised feature selection methods. *Pattern Recognit.* **2017**, *64*, 141–158. [CrossRef]
29. Bhowmick, K.; Narvekar, M. A Comprehensive Study and Analysis of Semi Supervised Learning Techniques. *Int. J. Eng. Res. Technol.* **2019**, *8*, 810–816. Available online: https://www.ijert.org/a-comprehensive-study-and-analysis-of-semi-supervised-learning-techniques (accessed on 5 May 2024).
30. Chi, S.; Li, X.; Tian, Y.; Li, J.; Kong, X.; Ding, K.; Weng, C.; Li, J. Semi-supervised learning to improve generalizability of risk prediction models. *J. Biomed. Inform.* **2019**, *92*, 103117. [CrossRef]
31. Mey, A.; Loog, M. Improved Generalization in Semi-Supervised Learning: A Survey of Theoretical Results. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 4747–4767. [CrossRef]
32. Wu, X.; Sahoo, D.; Hoi, S.C.H. Recent Advances in Deep Learning for Object Detection. *arXiv* **2019**, arXiv:1908.03673. [CrossRef]
33. Vaishnavi, K.; Reddy, G.; Reddy, T.; Iyengar, N.; Shaik, S. Real-time Object Detection Using Deep Learning. *J. Adv. Math. Comput. Sci.* **2023**, *38*, 24–32. [CrossRef]
34. Rawat, T.; Khemchandani, V. Feature Engineering (FE) Tools and Techniques for Better Classification Performance. *Int. J. Innov. Eng. Technol.* **2019**, *8*, 169–179. [CrossRef]
35. Devi, B.; Aruldoss, C.K.; Murugan, R. Feature Extraction and Object Detection Using Fast-Convolutional Neural Network for Remote Sensing Satellite Image. *J. Indian Soc. Remote Sens.* **2022**, *50*, 961–973. [CrossRef]
36. Gambo, F.L.; Haruna, A.S.; Muhammad, U.S.; Abdullahi, A.A.; Ahmed, B.A.; Dabai, U.S. Advances, Challenges and Opportunities in Deep Learning Approach for Object Detection: A Review. In Proceedings of the 2023 2nd International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), Abuja, Nigeria, 1–3 November 2023; Volume 1, pp. 1–6. [CrossRef]
37. Arkin, E.; Yadikar, N.; Muhtar, Y.; Ubul, K. A Survey of Object Detection Based on CNN and Transformer. In Proceedings of the 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 16–18 July 2021; pp. 99–108. [CrossRef]
38. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6. [CrossRef]
39. Jogin, M.; Madhulika, M.S.; Divya, G.; Meghana, R.; Apoorva, S. Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning. In Proceedings of the 2021 8th Swiss Conference on Data Science (SDS), Bangalore, India, 18–19 May 2018; pp. 2319–2323. [CrossRef]
40. Nguyen, N.M.; Ray, N. End-to-end Learning of Convolutional Neural Net and Dynamic Programming for Left Ventricle Segmentation. *arXiv* **2019**, arXiv:1812.00328.
41. Synnaeve, G.; Xu, Q.; Kahn, J.; Likhomanenko, T.; Grave, E.; Pratap, V.; Sriram, A.; Liptchinsky, V.; Collobert, R. End-to-end ASR: From Supervised to Semi-Supervised Learning with Modern Architectures. *arXiv* **2020**, arXiv:1911.08460.
42. Abdel-Basset, M.; Chang, V.; Hawash, H.; Chakrabortty, R.K.; Ryan, M. FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. *Knowl.-Based Syst.* **2021**, *212*, 106647. [CrossRef] [PubMed]
43. Chapelle, O.; Sindhwani, V.; Keerthi, S.S. Optimization Techniques for Semi-Supervised Support Vector Machines. *J. Mach. Learn. Res.* **2008**, *9*, 203–233.
44. Frommknecht, T.; Zipf, P.A.; Fan, Q.; Shvetsova, N.; Kuehne, H. Augmentation Learning for Semi-Supervised Classification. *arXiv* **2022**, arXiv:2208.01956.
45. Mumuni, A.; Mumuni, F. Data augmentation: A comprehensive survey of modern approaches. *Array* **2022**, *16*, 100258. [CrossRef]
46. Berthelot, D.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Sohn, K.; Zhang, H.; Raffel, C. ReMixMatch: Semi-Supervised Learning with Distribution Alignment and Augmentation Anchoring. *arXiv* **2020**, arXiv:1911.09785.
47. Xie, Q.; Dai, Z.; Hovy, E.; Luong, M.T.; Le, Q.V. Unsupervised Data Augmentation for Consistency Training. *arXiv* **2020**, arXiv:1904.12848. [CrossRef]
48. Pise, N.N.; Kulkarni, P. A Survey of Semi-Supervised Learning Methods. In Proceedings of the 2008 International Conference on Computational Intelligence and Security, Suzhou, China, 13–17 December 2008; Volume 2, pp. 30–34. [CrossRef]

49. Yang, X.; Song, Z.; King, I.; Xu, Z. A Survey on Deep Semi-Supervised Learning. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 8934–8954. [CrossRef]
50. Arazo, E.; Ortego, D.; Albert, P.; O'Connor, N.E.; McGuinness, K. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8. [CrossRef]
51. Xu, Q.; Likhomanenko, T.; Kahn, J.; Hannun, A.; Synnaeve, G.; Collobert, R. Iterative Pseudo-Labeling for Speech Recognition. *arXiv* **2020**, arXiv:2005.09267. [CrossRef]
52. Zhu, H.; Gao, D.; Cheng, G.; Povey, D.; Zhang, P.; Yan, Y. Alternative Pseudo-Labeling for Semi-Supervised Automatic Speech Recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2023**, *31*, 3320–3330. [CrossRef]
53. Lin, H.; Lou, J.; Xiong, L.; Shahabi, C. SemiFed: Semi-supervised Federated Learning with Consistency and Pseudo-Labeling. *arXiv* **2021**, arXiv:2108.09412.
54. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. *arXiv* **2019**, arXiv:1904.01355. [CrossRef]
55. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2016**, arXiv:1506.01497. [CrossRef]
56. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
57. Jia, D.; Yuan, Y.; He, H.; Wu, X.; Yu, H.; Lin, W.; Sun, L.; Zhang, C.; Hu, H. DETRs with Hybrid Matching. *arXiv* **2023**, arXiv:2207.13080. [CrossRef]
58. Li, F.; Zhang, H.; Liu, S.; Guo, J.; Ni, L.M.; Zhang, L. DN-DETR: Accelerate DETR Training by Introducing Query DeNoising. *arXiv* **2022**, arXiv:2203.01305. [CrossRef]
59. Shehzadi, T.; Hashmi, K.A.; Stricker, D.; Afzal, M.Z. Object Detection with Transformers: A Review. *arXiv* **2023**, arXiv:2306.04670.
60. Hai Son, L.; Allauzen, A.; Yvon, F. Measuring the influence of long range dependencies with neural network language models. In Proceedings of the WLM@NAACL-HLT, Montréal, QC, Canada, 8 June 2012; pp. 1–10. [CrossRef]
61. Vierlboeck, M.; Dunbar, D.; Nilchiani, R. Natural Language Processing to Extract Contextual Structure from Requirements. In Proceedings of the 2022 IEEE International Systems Conference (SysCon), Montreal, QC, Canada, 25–28 April 2022; pp. 1–8. [CrossRef]
62. Pérez, J.M.; Luque, F.; Zayat, D.; Kondratzky, M.; Moro, A.; Serrati, P.; Zajac, J.; Miguel, P.; Debandi, N.; Gravano, A.; et al. Assessing the impact of contextual information in hate speech detection. *arXiv* **2023**, arXiv:2210.00465. [CrossRef]
63. Chen, X.; Gupta, A. Spatial Memory for Context Reasoning in Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4106–4116. [CrossRef]
64. Deléarde, R.; Kurtz, C.; Wendling, L. Description and recognition of complex spatial configurations of object pairs with Force Banner 2D features. *Pattern Recognit.* **2022**, *123*, 108410. [CrossRef]
65. Hosang, J.; Benenson, R.; Schiele, B. Learning non-maximum suppression. *arXiv* **2017**, arXiv:1705.02950. [CrossRef]
66. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762. [PubMed]
67. Huang, Z.; Liang, M.; Qin, J.; Zhong, S.; Lin, L. Understanding Self-attention Mechanism via Dynamical System Perspective. *arXiv* **2023**, arXiv:2308.09939. [CrossRef]
68. Huang, Z.; Tang, F.; Zhang, Y.; Cun, X.; Cao, J.; Li, J.; Lee, T.Y. Make-Your-Anchor: A Diffusion-based 2D Avatar Generation Framework. *arXiv* **2024**, arXiv:2403.16510.
69. Hoanh, N.; Pham, T.V. Focus-Attention Approach in Optimizing DETR for Object Detection from High-Resolution Images. *Knowl.-Based Syst.* **2024**, *296*, 111939. [CrossRef]
70. Tang, P.; Ramaiah, C.; Xu, R.; Xiong, C. Proposal Learning for Semi-Supervised Object Detection. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Virtual, 5–9 January 2021; pp. 2290–2300. Available online: https://api.semanticscholar.org/CorpusID:210699986 (accessed on 5 May 2024).
71. Shehzadi, T.; Ifza, I.; Stricker, D.; Afzal, M.Z. FD-SSD: Semi-supervised Detection of Bone Fenestration and Dehiscence in Intraoral Images. In Proceedings of the Medical Image Understanding and Analysis (MIUA) 2025, Leeds, UK, 15–17 July 2025; Springer: Cham, Switzerland, 2025; Volume 15917.
72. Shehzadi, T.; Hashmi, K.A.; Pagani, A.; Liwicki, M.; Stricker, D.; Afzal, M.Z. Mask-Aware Semi-Supervised Object Detection in Floor Plans. *Appl. Sci.* **2022**, *12*, 9398. [CrossRef]
73. Zhang, J.; Lin, X.; Zhang, W.; Wang, K.; Tan, X.; Han, J.; Ding, E.; Wang, J.; Li, G. Semi-DETR: Semi-Supervised Object Detection with Detection Transformers. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 23809–23818. [CrossRef]

74. Shehzadi, T.; Hashmi, K.A.; Stricker, D.; Afzal, M.Z. Sparse Semi-DETR: Sparse Learnable Queries for Semi-Supervised Object Detection. *arXiv* **2024**, arXiv:2404.01819.
75. Wang, P.; Cai, Z.; Yang, H.; Swaminathan, G.; Vasconcelos, N.; Schiele, B.; Soatto, S. Omni-DETR: Omni-Supervised Object Detection with Transformers. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 9357–9366. Available online: https://api.semanticscholar.org/CorpusID:247792844 (accessed on 5 May 2024).
76. Li, G.; Li, X.; Wang, Y.; Wu, Y.; Liang, D.; Zhang, S. PseCo: Pseudo Labeling and Consistency Training for Semi-Supervised Object Detection. *arXiv* **2022**, arXiv:2203.16317.
77. Jeong, J.; Lee, S.; Kim, J.; Kwak, N. Consistency-based Semi-supervised Learning for Object detection. In Proceedings of the Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019. Available online: https://api.semanticscholar.org/CorpusID:202782547 (accessed on 5 May 2024).
78. Sohn, K.; Zhang, Z.; Li, C.L.; Zhang, H.; Lee, C.Y.; Pfister, T. A Simple Semi-Supervised Learning Framework for Object Detection. *arXiv* **2020**, arXiv:2005.04757. [CrossRef]
79. Tang, Y.; Chen, W.; Luo, Y.; Zhang, Y. Humble Teachers Teach Better Students for Semi-Supervised Object Detection. *arXiv* **2021**, arXiv:2106.10456. [CrossRef]
80. Zhou, Q.; Yu, C.; Wang, Z.; Qian, Q.; Li, H. Instant-Teaching: An End-to-End Semi-Supervised Object Detection Framework. *arXiv* **2021**, arXiv:2103.11402.
81. Xu, M.; Zhang, Z.; Hu, H.; Wang, J.; Wang, L.; Wei, F.; Bai, X.; Liu, Z. End-to-End Semi-Supervised Object Detection with Soft Teacher. *arXiv* **2021**, arXiv:2106.09018.
82. Liu, Y.C.; Ma, C.Y.; He, Z.; Kuo, C.W.; Chen, K.; Zhang, P.; Wu, B.; Kira, Z.; Vajda, P. Unbiased Teacher for Semi-Supervised Object Detection. *arXiv* **2021**, arXiv:2102.09480. Available online: https://api.semanticscholar.org/CorpusID:231951546 (accessed on 5 May 2024).
83. Zhang, F.; Pan, T.; Wang, B. Semi-Supervised Object Detection with Adaptive Class-Rebalancing Self-Training. *arXiv* **2021**, arXiv:2107.05031. [CrossRef]
84. Wang, Z.; Li, Y.; Guo, Y.; Wang, S. Combating noise: Semi-supervised learning by region uncertainty quantification. In Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS '21, Virtual, 6–14 December 2021. Available online: https://dl.acm.org/doi/10.5555/3540261.3540991 (accessed on 5 May 2024).
85. Kim, J.; Jang, J.; Seo, S.; Jeong, J.; Na, J.; Kwak, N. MUM: Mix Image Tiles and UnMix Feature Tiles for Semi-Supervised Object Detection. *arXiv* **2022**, arXiv:2111.10958.
86. Yang, Q.; Wei, X.; Wang, B.; Hua, X.S.; Zhang, L. Interactive Self-Training with Mean Teachers for Semi-supervised Object Detection. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 5937–5946. [CrossRef]
87. Ma, C.; Pan, X.; Ye, Q.; Tang, F.; Dong, W.; Xu, C. CrossRectify: Leveraging Disagreement for Semi-supervised Object Detection. *arXiv* **2022**, arXiv:2201.10734. [CrossRef]
88. Guo, Q.; Mu, Y.; Chen, J.; Wang, T.; Yu, Y.; Luo, P. Scale-Equivalent Distillation for Semi-Supervised Object Detection. *arXiv* **2022**, arXiv:2203.12244.
89. Chen, B.; Chen, W.; Yang, S.; Xuan, Y.; Song, J.; Xie, D.; Pu, S.; Song, M.; Zhuang, Y. Label Matching Semi-Supervised Object Detection. *arXiv* **2022**, arXiv:2206.06608. [CrossRef]
90. Li, G.; Li, X.; Wang, Y.; Wu, Y.; Liang, D.; Zhang, S. DTG-SSOD: Dense Teacher Guidance for Semi-Supervised Object Detection. *arXiv* **2022**, arXiv:2207.05536.
91. Li, H.; Wu, Z.; Shrivastava, A.; Davis, L.S. Rethinking Pseudo Labels for Semi-Supervised Object Detection. *arXiv* **2021**, arXiv:2106.00168. [CrossRef]
92. Mi, P.; Lin, J.; Zhou, Y.; Shen, Y.; Luo, G.; Sun, X.; Cao, L.; Fu, R.; Xu, Q.; Ji, R. Active Teacher for Semi-Supervised Object Detection. *arXiv* **2023**, arXiv:2303.08348. [CrossRef]
93. Xiong, F.; Tian, J.; Hao, Z.; He, Y.; Ren, X. SCMT: Self-Correction Mean Teacher for Semi-supervised Object Detection. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, Vienna, Austria, 23–29 July 2022; Raedt, L.D., Ed.; International Joint Conferences on Artificial Intelligence Organization; pp. 1488–1494. [CrossRef]
94. Wang, X.; Yang, X.; Zhang, S.; Li, Y.; Feng, L.; Fang, S.; Lyu, C.; Chen, K.; Zhang, W. Consistent-Teacher: Towards Reducing Inconsistent Pseudo-targets in Semi-supervised Object Detection. *arXiv* **2023**, arXiv:2209.01589.
95. Chen, B.; Li, P.; Chen, X.; Wang, B.; Zhang, L.; Hua, X.S. Dense Learning based Semi-Supervised Object Detection. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 4805–4814. [CrossRef]
96. Zhou, H.; Ge, Z.; Liu, S.; Mao, W.; Li, Z.; Yu, H.; Sun, J. Dense Teacher: Dense Pseudo-Labels for Semi-supervised Object Detection. *arXiv* **2022**, arXiv:2207.02541.

97. Zhang, Y.; Yao, X.; Liu, C.; Chen, F.; Song, X.; Xing, T.; Hu, R.; Chai, H.; Xu, P.; Zhang, G. S4OD: Semi-Supervised learning for Single-Stage Object Detection. *arXiv* **2022**, arXiv:2204.04492.
98. Liu, Y.C.; Ma, C.Y.; Kira, Z. Unbiased Teacher v2: Semi-supervised Object Detection for Anchor-free and Anchor-based Detectors. *arXiv* **2022**, arXiv:2206.09500.
99. Luo, G.; Zhou, Y.; Jin, L.; Sun, X.; Ji, R. Towards End-to-end Semi-supervised Learning for One-stage Object Detection. *arXiv* **2023**, arXiv:2302.11299.
100. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. RandAugment: Practical automated data augmentation with a reduced search space. *arXiv* **2019**, arXiv:1909.13719. [CrossRef]
101. Sajjadi, M.; Javanmardi, M.; Tasdizen, T. Regularization With Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning. *arXiv* **2016**, arXiv:1606.04586. [CrossRef]
102. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. *arXiv* **2017**, arXiv:1708.04896. [CrossRef]
103. Sohn, K.; Berthelot, D.; Li, C.L.; Zhang, Z.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Zhang, H.; Raffel, C. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *arXiv* **2020**, arXiv:2001.07685.
104. Zhu, X. *Semi-Supervised Learning Literature Survey*; Technical Report 1530; Computer Sciences, University of Wisconsin-Madison: Madison, WI, USA, 2005. Available online: https://minds.wisconsin.edu/handle/1793/60444 (accessed on 5 May 2024).
105. Mey, A.; Loog, M. Improvability Through Semi-Supervised Learning: A Survey of Theoretical Results. *arXiv* **2020**, arXiv:1908.09574. [CrossRef]
106. Simmler, N.; Sager, P.; Andermatt, P.; Chavarriaga, R.; Schilling, F.P.; Rosenthal, M.; Stadelmann, T. A Survey of Un-, Weakly-, and Semi-Supervised Learning Methods for Noisy, Missing and Partial Labels in Industrial Vision Applications. In Proceedings of the 2021 8th Swiss Conference on Data Science (SDS), Lucerne, Switzerland, 9 June 2021; pp. 26–31. [CrossRef]
107. Silva, N.F.F.D.; Coletta, L.F.S.; Hruschka, E.R. A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning. *ACM Comput. Surv.* **2016**, *49*, 15. [CrossRef]
108. Prakash, V.J.; Nithya, D.L. A Survey On Semi-Supervised Learning Techniques. *Int. J. Comput. Trends Technol.* **2014**, *8*, 25–29. [CrossRef]
109. van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2019**, *109*, 373–440. Available online: https://api.semanticscholar.org/CorpusID:254738406 (accessed on 5 May 2024).
110. Qi, G.J.; Luo, J. Small Data Challenges in Big Data Era: A Survey of Recent Progress on Unsupervised and Semi-Supervised Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2168–2187 . [CrossRef] [PubMed]
111. Schmarje, L.; Santarossa, M.; Schroder, S.M.; Koch, R. A Survey on Semi-, Self- and Unsupervised Learning for Image Classification. *IEEE Access* **2021**, *9*, 82146–82168. [CrossRef]
112. Chen, Y.; Mancini, M.; Zhu, X.; Akata, Z. Semi-Supervised and Unsupervised Deep Visual Learning: A Survey. *arXiv* **2022**, arXiv:2208.11296. [CrossRef]
113. Chebli, A.; Djebbar, A.; Marouani, H.F. Semi-Supervised Learning for Medical Application: A Survey. In Proceedings of the 2018 International Conference on Applied Smart Systems (ICASS), Medea, Algeria, 24–25 November 2018; pp. 1–9. Available online: https://api.semanticscholar.org/CorpusID:67876194 (accessed on 5 May 2024).
114. Gomes, H.M.; Grzenda, M.; Mello, R.; Read, J.; Le Nguyen, M.H.; Bifet, A. A Survey on Semi-supervised Learning for Delayed Partially Labelled Data Streams. *ACM Comput. Surv.* **2022**, *55*, 75. [CrossRef]
115. Calderon-Ramirez, S.; Yang, S.; Elizondo, D. Semisupervised Deep Learning for Image Classification With Distribution Mismatch: A Survey. *IEEE Trans. Artif. Intell.* **2022**, *3*, 1015–1029. [CrossRef]
116. Song, Z.; Yang, X.; Xu, Z.; King, I. Graph-Based Semi-Supervised Learning: A Comprehensive Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 8174–8194. [CrossRef]
117. Keles, M.C.; Salmanoglu, B.; Guzel, M.S.; Gursoy, B.; Bostanci, G.E. Evaluation of YOLO Models with Sliced Inference for Small Object Detection. *arXiv* **2022**, arXiv:2203.04799. [CrossRef]
118. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430. [CrossRef]
119. Bai, L.; Gupta, A.; Ong, Y.S. Multi-Task Learning with Multi-Task Optimization. *arXiv* **2024**, arXiv:2403.16162.
120. Xiong, B.; Fan, H.; Grauman, K.; Feichtenhofer, C. Multiview Pseudo-Labeling for Semi-supervised Learning from Video. *arXiv* **2021**, arXiv:2104.00682.
121. Kong, D.; Huang, Y.; Xie, J.; Honig, E.; Xu, M.; Xue, S.; Lin, P.; Zhou, S.; Zhong, S.; Zheng, N.; et al. Dual-Space Optimization: Improved Molecule Sequence Design by Latent Prompt Transformer. *arXiv* **2024**, arXiv:2402.17179. [CrossRef]
122. Zhu, L.; Ke, Z.; Lau, R. Towards Self-Adaptive Pseudo-Label Filtering for Semi-Supervised Learning. *arXiv* **2023**, arXiv:2309.09774.
123. Kim, J.; Ryoo, K.; Lee, G.; Cho, S.; Seo, J.; Kim, D.; Cho, H.; Kim, S. AggMatch: Aggregating Pseudo Labels for Semi-Supervised Learning. *arXiv* **2022**, arXiv:2201.10444.
124. Fan, Y.; Kukleva, A.; Schiele, B. Revisiting Consistency Regularization for Semi-Supervised Learning. *arXiv* **2021**, arXiv:2112.05825. [CrossRef]

125. Shang, C.; Ma, T.; Ren, W.; Li, Y.; Yang, J. Sparse Generation: Making Pseudo Labels Sparse for Point Weakly Supervised Object Detection on Low Data Volume. In Proceedings of the ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5. [CrossRef]
126. Liu, W.; Wang, H.; Luo, H.; Zhang, K.; Lu, J.; Xiong, Z. Pseudo-label growth dictionary pair learning for crowd counting. *Appl. Intell.* **2021**, *51*, 8913–8927. [CrossRef]
127. Jo, Y.; Kahng, H.; Kim, S.B. Deep Semi-Supervised Regression via Pseudo-Label Filtering and Calibration. *Appl. Soft Comput.* **2024**, *161*, 111670 . [CrossRef]
128. Liu, S.; Zhou, H.; Li, C.; Wang, S. Analysis of Anchor-Based and Anchor-Free Object Detection Methods Based on Deep Learning. In Proceedings of the 2020 IEEE International Conference on Mechatronics and Automation (ICMA), Beijing, China, 13–16 October 2020; pp. 1058–1065. [CrossRef]
129. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2018**, arXiv:1708.02002. [CrossRef]
130. Wang, S.; Zhang, A.; Wang, H. A Feature Extraction Algorithm for Enhancing Graphical Local Adaptive Threshold. In Proceedings of the Intelligent Computing Theories and Application: 18th International Conference, ICIC 2022, Xi'an, China, 7–11 August 2022; Proceedings, Part I; Springer: Cham, Switzerland, 2022; pp. 277–291. [CrossRef]
131. Salman, S.; Liu, X. Overfitting Mechanism and Avoidance in Deep Neural Networks. *arXiv* **2019**, arXiv:1901.06566. [CrossRef]
132. Feng, C.; Zhong, Y.; Gao, Y.; Scott, M.R.; Huang, W. TOOD: Task-aligned One-stage Object Detection. *arXiv* **2021**, arXiv:2108.07755.
133. Ge, Z.; Liu, S.; Li, Z.; Yoshie, O.; Sun, J. OTA: Optimal Transport Assignment for Object Detection. *arXiv* **2021**, arXiv:2103.14259. [CrossRef]
134. Jiang, B.; Luo, R.; Mao, J.; Xiao, T.; Jiang, Y. Acquisition of Localization Confidence for Accurate Object Detection. *arXiv* **2018**, arXiv:1807.11590. [CrossRef]
135. Ciampiconi, L.; Elwood, A.; Leonardi, M.; Mohamed, A.; Rozza, A. A survey and taxonomy of loss functions in machine learning. *arXiv* **2023**, arXiv:2301.05579. [CrossRef]
136. Oksuz, K.; Cam, B.C.; Kalkan, S.; Akbas, E. Imbalance Problems in Object Detection: A Review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3388–3415. [CrossRef]
137. Peng, J.; Bu, X.; Sun, M.; Zhang, Z.; Tan, T.; Yan, J. Large-Scale Object Detection in the Wild From Imbalanced Multi-Labels. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 9706–9715. [CrossRef]
138. Cao, Y.; Chen, K.; Loy, C.C.; Lin, D. Prime Sample Attention in Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11580–11588. [CrossRef]
139. Chen, K.; Li, J.; Lin, W.; See, J.; Wang, J.; Duan, L.; Chen, Z.; He, C.; Zou, J. Towards Accurate One-Stage Object Detection With AP-Loss. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019, pp. 5114–5122. [CrossRef]
140. Wisdom, S.; Hershey, J.R.; Wilson, K.W.; Thorpe, J.; Chinen, M.; Patton, B.; Saurous, R.A. Differentiable Consistency Constraints for Improved Deep Speech Enhancement. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2018; pp. 900–904. Available online: https://api.semanticscholar.org/CorpusID:53739933 (accessed on 5 May 2024).
141. Shen, Z.; Cao, P.; Yang, H.; Liu, X.; Yang, J.; Zaiane, O.R. Co-training with High-Confidence Pseudo Labels for Semi-supervised Medical Image Segmentation. *arXiv* **2023**, arXiv:2301.04465.
142. Zoph, B.; Cubuk, E.D.; Ghiasi, G.; Lin, T.Y.; Shlens, J.; Le, Q.V. Learning Data Augmentation Strategies for Object Detection. *arXiv* **2019**, arXiv:1906.11172. [CrossRef]
143. Scudder, H. Probability of error of some adaptive pattern-recognition machines. *IEEE Trans. Inf. Theory* **1965**, *11*, 363–371. [CrossRef]
144. McLachlan, G.J. Iterative Reclassification Procedure for Constructing an Asymptotically Optimal Rule of Allocation in Discriminant Analysis. *J. Am. Stat. Assoc.* **1975**, *70*, 365–369. Available online: http://www.jstor.org/stable/2285824 (accessed on 5 May 2024). [CrossRef]
145. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv* **2018**, arXiv:1703.01780.
146. Xia, K.; Wang, L.; Zhou, S.; Hua, G.; Tang, W. Learning from Noisy Pseudo Labels for Semi-Supervised Temporal Action Localization. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 10126–10135. [CrossRef]
147. Wang, B.; Li, J.; Liu, Y.; Cheng, J.; Rong, Y.; Wang, W.; Tsung, F. Deep Insights into Noisy Pseudo Labeling on Graph Data. *arXiv* **2023**, arXiv:2310.01634. [CrossRef]
148. Zhao, W.; Mou, L.; Chen, J.; Bo, Y.; Emery, W.J. Incorporating Metric Learning and Adversarial Network for Seasonal Invariant Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2720–2731. [CrossRef]

149. Zoph, B.; Ghiasi, G.; Lin, T.Y.; Cui, Y.; Liu, H.; Cubuk, E.D.; Le, Q.V. Rethinking pre-training and self-training. In Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, Virtual, 6–12 December 2020. Available online: https://dl.acm.org/doi/proceedings/10.5555/3495724 (accessed on 5 May 2024).
150. He, Y.; Chen, W.; Liang, K.; Tan, Y.; Liang, Z.; Guo, Y. Pseudo-label Correction and Learning For Semi-Supervised Object Detection. *arXiv* **2023**, arXiv:2303.02998.
151. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740. [CrossRef]
152. Klinker, F. Exponential moving average versus moving exponential average. *Math. Semesterber.* **2011**, *58*, 97–107. [CrossRef]
153. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C. MixMatch: A Holistic Approach to Semi-Supervised Learning. *arXiv* **2019**, arXiv:1905.02249.
154. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—Improving Object Detection With One Line of Code. *arXiv* **2017**, arXiv:1704.04503.
155. Verma, V.; Kawaguchi, K.; Lamb, A.; Kannala, J.; Solin, A.; Bengio, Y.; Lopez-Paz, D. Interpolation consistency training for semi-supervised learning. *Neural Netw.* **2022**, *145*, 90–106. [CrossRef]
156. French, G.; Laine, S.; Aila, T.; Mackiewicz, M.; Finlayson, G. Semi-supervised semantic segmentation needs strong, varied perturbations. *arXiv* **2020**, arXiv:1906.01916. [CrossRef]
157. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *arXiv* **2020**, arXiv:2006.04388. [CrossRef]
158. van Noord, N.; Postma, E. Learning scale-variant and scale-invariant features for deep image classification. *arXiv* **2016**, arXiv:1602.01255. [CrossRef]
159. Asano, Y.M.; Rupprecht, C.; Vedaldi, A. Self-labelling via simultaneous clustering and representation learning. *arXiv* **2020**, arXiv:1911.05371. [CrossRef]
160. Qi, C.R.; Litany, O.; He, K.; Guibas, L.J. Deep Hough Voting for 3D Object Detection in Point Clouds. *arXiv* **2019**, arXiv:1904.09664. [CrossRef]
161. Cao, K.; Wei, C.; Gaidon, A.; Arechiga, N.; Ma, T. Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss. *arXiv* **2019**, arXiv:1906.07413.
162. Higashimoto, R.; Yoshida, S.; Horihata, T.; Muneyasu, M. Unbiased Pseudo-Labeling for Learning with Noisy Labels. *IEICE Trans. Inf. Syst.* **2024**, *107*, 44–48. [CrossRef]
163. Wang, S.; Zhuang, S.; Zuccon, G. Large Language Models for Stemming: Promises, Pitfalls and Failures. *arXiv* **2024**, arXiv:2402.11757. [CrossRef]
164. Ismailov, A.; Jalil, M.A.; Abdullah, Z.; Rahim, N.A. A comparative study of stemming algorithms for use with the Uzbek language. In Proceedings of the 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 15–17 August 2016; pp. 7–12. [CrossRef]
165. Qian, X.; Li, C.; Wang, W.; Yao, X.; Cheng, G. Semantic segmentation guided pseudo label mining and instance re-detection for weakly supervised object detection in remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *119*, 103301. [CrossRef]
166. Desai, B.; Paliwal, M.; Nagwanshi, K.K. Study on Image Filtering—Techniques, Algorithm and Applications. *arXiv* **2022**, arXiv:2207.06481.
167. Guo, Q.; Wang, X.; Wu, Y.; Yu, Z.; Liang, D.; Hu, X.; Luo, P. Online Knowledge Distillation via Collaborative Learning. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11017–11026. [CrossRef]
168. Seiffert, C.; Khoshgoftaar, T.M.; Hulse, J.V.; Napolitano, A. Resampling or Reweighting: A Comparison of Boosting Implementations. In Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence, Dayton, OH, USA, 3–5 November 2008; Volume 1, pp. 445–451. [CrossRef]
169. Zhang, Y.; Zhao, S.; Gu, H.; Mazurowski, M.A. How to Efficiently Annotate Images for Best-Performing Deep Learning Based Segmentation Models: An Empirical Study with Weak and Noisy Annotations and Segment Anything Model. *arXiv* **2023**, arXiv:2312.10600. [CrossRef]
170. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. *arXiv* **2020**, arXiv:2005.12872.
171. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *arXiv* **2020**, arXiv:2010.04159 .
172. M., A.; Govindharaju, K.; John, A.; Mohan, S.; Ahmadian, A.; Ciano, T. A hybrid learning approach for the stage-wise classification and prediction of COVID-19 X-ray images. *Expert Syst.* **2022**, *39*, e12884, [CrossRef]
173. Chen, Q.; Chen, X.; Wang, J.; Zhang, S.; Yao, K.; Feng, H.; Han, J.; Ding, E.; Zeng, G.; Wang, J. Group DETR: Fast DETR Training with Group-Wise One-to-Many Assignment. *arXiv* **2023**, arXiv:2207.13085.
174. Wang, Z.; Zhao, Z.; Xing, X.; Xu, D.; Kong, X.; Zhou, L. Conflict-Based Cross-View Consistency for Semi-Supervised Semantic Segmentation. *arXiv* **2023**, arXiv:2303.01276.

175. Mosbah, M. Query Refinement into Information Retrieval Systems: An Overview. *J. Inf. Organ. Sci.* **2023**, *47*, 133–151. [CrossRef]
176. Shehzadi, T.; Hashmi, K.A.; Sarode, S.; Stricker, D.; Afzal, M.Z. STEP-DETR: Advancing DETR-based Semi-Supervised Object Detection with Super Teacher and Pseudo-Label Guided Text Queries. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–23 October 2025; pp. 3069–3079. Available online: https://openaccess.thecvf.com/content/ICCV2025/html/Shehzadi_STEP-DETR_Advancing_DETR-based_Semi-Supervised_Object_Detection_with_Super_Teacher_and_ICCV_2025_paper.html (accessed on 5 May 2024).
177. Sutanto, A.R.; Kang, D.K. A Novel Diminish Smooth L1 Loss Model with Generative Adversarial Network. In Proceedings of the Intelligent Human Computer Interaction: 12th International Conference, IHCI 2020, Daegu, Republic of Korea, 24–26 November 2020; Proceedings, Part I; Springer: Cham, Switzerland, 2020; pp. 361–368. [CrossRef]
178. Chen, L.; Yang, T.; Zhang, X.; Zhang, W.; Sun, J. Points as Queries: Weakly Semi-supervised Object Detection by Points. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 8819–8828. [CrossRef]
179. Gong, C.; Zhang, X. Knowledge Transfer for Object Detection. *J. Phys. Conf. Ser.* **2020**, *1549*, 052119. [CrossRef]
180. Banitalebi-Dehkordi, A. Knowledge Distillation for Low-Power Object Detection: A Simple Technique and Its Extensions for Training Compact Models Using Unlabeled Data. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 769–778. [CrossRef]
181. Li, B.; Yao, Y.; Tan, J.; Zhang, G.; Yu, F.; Lu, J.; Luo, Y. Equalized Focal Loss for Dense Long-Tailed Object Detection. *arXiv* **2022**, arXiv:2201.02593. [CrossRef]
182. Yang, X.; Yang, X.; Yang, J.; Ming, Q.; Wang, W.; Tian, Q.; Yan, J. Learning High-Precision Bounding Box for Rotated Object Detection via Kullback-Leibler Divergence. *arXiv* **2022**, arXiv:2106.01883.
183. Seo, G.; Yoo, J.; Choi, J.; Kwak, N. KL-Divergence-Based Region Proposal Network for Object Detection. *arXiv* **2020**, arXiv:2005.11220.
184. Li, X.; Wang, W.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized Focal Loss V2: Learning Reliable Localization Quality Estimation for Dense Object Detection. *arXiv* **2020**, arXiv:2011.12885. [CrossRef]
185. Hoyos-Osorio, J.K.; Posso-Murillo, S.; Sanchez-Giraldo, L.G. The Representation Jensen-Shannon Divergence. *arXiv* **2023**, arXiv:2305.16446. [CrossRef]
186. Nielsen, F. On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means. *Entropy* **2019**, *21*, 485. [CrossRef] [PubMed]
187. Lee, S.; Kim, T.; Heo, J.p. Cross-Loss Pseudo Labeling for Semi-Supervised Segmentation. *IEEE Access* **2023**, *11*, 96761–96772. [CrossRef]
188. Mao, A.; Mohri, M.; Zhong, Y. Cross-Entropy Loss Functions: Theoretical Analysis and Applications. *arXiv* **2023**, arXiv:2304.07288. [CrossRef]
189. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312. [CrossRef]
190. Everingham, M.; Van Gool, L.; Williams, C.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
191. Hu, X.; Niu, Y.; Miao, C.; Hua, X.S.; Zhang, H. On Non-Random Missing Labels in Semi-Supervised Learning. *arXiv* **2022**, arXiv:2206.14923.
192. Solatidehkordi, Z.; Zualkernan, I. Survey on Recent Trends in Medical Image Classification Using Semi-Supervised Learning. *Appl. Sci.* **2022**, *12*, 12094. [CrossRef]
193. Inés, A.; Domínguez, C.; Heras, J.; Mata, E.; Pascual, V. Biomedical image classification made easier thanks to transfer and semi-supervised learning. *Comput. Methods Programs Biomed.* **2021**, *198*, 105782. [CrossRef]
194. Yousaf, A.; Shehzadi, T.; Farooq, A.; Ilyas, K. Protein active site prediction for early drug discovery and designing. *Int. Rev. Appl. Sci. Eng.* **2021**, *13*, 98–105. [CrossRef]
195. Shehzadi, T.; Majid, A.; Hameed, M.; Farooq, A.; Yousaf, A. Intelligent predictor using cancer-related biologically information extraction from cancer transcriptomes. In Proceedings of the 2020 International Symposium on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS), Islamabad, Pakistan, 20–22 October 2020; Volume 5, pp. 1–5. [CrossRef]
196. Yan, P.; He, F.; Yang, Y.; Hu, F. Semi-Supervised Representation Learning for Remote Sensing Image Classification Based on Generative Adversarial Networks. *IEEE Access* **2020**, *8*, 54135–54144. [CrossRef]
197. Wan, L.; Tang, K.; Li, M.; Zhong, Y.; Qin, A.K. Collaborative Active and Semisupervised Learning for Hyperspectral Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2384–2396. [CrossRef]
198. Li, J.; Gang, H.; Ma, H.; Tomizuka, M.; Choi, C. Important Object Identification with Semi-Supervised Learning for Autonomous Driving. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 2913–2919. [CrossRef]

199. Chen, W.; Yan, J.; Huang, W.; Ge, W.; Liu, H.; Yin, H. Robust Object Detection for Autonomous Driving Based on Semi-supervised Learning. *Secur. Saf.* **2024**, *3*, 2024002. [CrossRef]
200. Shehzadi, T.; Hashmi, K.A.; Stricker, D.; Liwicki, M.; Afzal, M.Z. Bridging the Performance Gap between DETR and R-CNN for Graphical Object Detection in Document Images. *arXiv* **2023**, arXiv:2306.13526. [CrossRef]
201. Shehzadi, T.; Sarode, S.; Stricker, D.; Afzal, M.Z. Towards End-to-End Semi-supervised Table Detection with Semantic Aligned Matching Transformer. In Proceedings of the Document Analysis and Recognition—ICDAR 2024, Athens, Greece, 30 August–4 September 2024; Barney Smith, E.H., Liwicki, M., Peng, L., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2024; Volume 14808. [CrossRef]
202. Shehzadi, T.; Stricker, D.; Afzal, M.Z. A Hybrid Approach for Document Layout Analysis in Document images. *arXiv* **2024**, arXiv:2404.17888. [CrossRef]
203. Shehzadi, T.; Ifza, I.; Stricker, D.; Afzal, M.Z. Efficient Additive Attention for Transformer-based Semi-supervised Document Layout Analysis. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–23 October 2025; pp. 3495–3503.
204. Kallempudi, G.; Hashmi, K.A.; Pagani, A.; Liwicki, M.; Stricker, D.; Afzal, M.Z. Toward Semi-Supervised Graphical Object Detection in Document Images. *Future Internet* **2022**, *14*, 176. [CrossRef]
205. Hashmi, K.A.; Pagani, A.; Liwicki, M.; Stricker, D.; Afzal, M.Z. CasTabDetectoRS: Cascade Network for Table Detection in Document Images with Recursive Feature Pyramid and Switchable Atrous Convolution. *J. Imaging* **2021**, *7*, 214. [CrossRef]
206. Schreiber, S.; Agne, S.; Wolf, I.; Dengel, A.; Ahmed, S. DeepDeSRT: Deep Learning for Detection and Structure Recognition of Tables in Document Images. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 1, pp. 1162–1167. [CrossRef]
207. Hao, L.; Gao, L.; Yi, X.; Tang, Z. A Table Detection Method for PDF Documents Based on Convolutional Neural Networks. In Proceedings of the 2016 12th IAPR Workshop on Document Analysis Systems (DAS), Santorini, Greece, 11–14 April 2016; pp. 287–292. [CrossRef]
208. Fang, J.; Gao, L.; Bai, K.; Qiu, R.; Tao, X.; Tang, Z. A Table Detection Method for Multipage PDF Documents via Visual Seperators and Tabular Structures. In Proceedings of the 2011 International Conference on Document Analysis and Recognition, Beijing, China, 18–21 September 2011; pp. 779–783. [CrossRef]
209. Kasar, T.; Barlas, P.; Adam, S.; Chatelain, C.; Paquet, T. Learning to Detect Tables in Scanned Document Images Using Line Information. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 25–28 August 2013; pp. 1185–1189. [CrossRef]
210. Shehzadi, T.; Ifza, I.; Stricker, D.; Afzal, M.Z. DocSemi: Efficient Document Layout Analysis with Guided Queries. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–23 October 2025; pp. 7536–7546.
211. Ehsan, I.; Shehzadi, T.; Stricker, D.; Afzal, M.Z. End-to-End Semi-Supervised approach with Modulated Object Queries for Table Detection in Documents. *Int. J. Doc. Anal. Recognit.* **2024**, *27*, 363–378. Available online: https://api.semanticscholar.org/CorpusID:269626070 (accessed on 5 May 2024). [CrossRef]
212. Xie, Q.; Luong, M.T.; Hovy, E.; Le, Q.V. Self-Training With Noisy Student Improves ImageNet Classification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10684–10695. [CrossRef]
213. Wang, H.; Cong, Y.; Litany, O.; Gao, Y.; Guibas, L.J. 3DIoUMatch: Leveraging IoU Prediction for Semi-Supervised 3D Object Detection. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 14610–14619. [CrossRef]
214. Yin, J.; Fang, J.; Zhou, D.; Zhang, L.; Xu, C.Z.; Shen, J.; Wang, W. Semi-supervised 3D Object Detection with Proficient Teachers. *arXiv* **2022**, arXiv:2207.12655. [CrossRef]
215. Barrera, A.; Guindel, C.; Beltrán, J.; García, F. BirdNet+: End-to-End 3D Object Detection in LiDAR Bird's Eye View. *arXiv* **2020**, arXiv:2003.04188.
216. Bai, X.; Hu, Z.; Zhu, X.; Huang, Q.; Chen, Y.; Fu, H.; Tai, C.L. TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers. *arXiv* **2022**, arXiv:2203.11496.
217. Hazarika, A.; Vyas, A.; Rahmati, M.; Wang, Y. Multi-Camera 3D Object Detection for Autonomous Driving Using Deep Learning and Self-Attention Mechanism. *IEEE Access* **2023**, *11*, 64608–64620. [CrossRef]
218. Ma, X.; Ouyang, W.; Simonelli, A.; Ricci, E. 3D Object Detection from Images for Autonomous Driving: A Survey. *arXiv* **2023**, arXiv:2202.02980. [CrossRef] [PubMed]
219. Jin, Z.; Liang, Z.; He, M.; Peng, Y.; Xue, H.; Wang, Y. A federated semi-supervised learning approach for network traffic classification. *Int. J. Netw. Manag.* **2023**, *33*, e2222 . [CrossRef]
220. Erman, J.; Mahanti, A.; Arlitt, M.; Cohen, I.; Williamson, C. Offline/realtime traffic classification using semi-supervised learning. *Perform. Eval.* **2007**, *64*, 1194–1213. [CrossRef]

221. Saeed, W.; Saleh, M.S.; Gull, M.N.; Raza, H.; Saeed, R.; Shehzadi, T. Geometric features and traffic dynamic analysis on 4-leg intersections. *Int. Rev. Appl. Sci. Eng.* **2024**, *15*, 171–188. [CrossRef]
222. Pei, X.; Deng, X.; Tian, S.; Zhang, L.; Xue, K. A Knowledge Transfer-Based Semi-Supervised Federated Learning for IoT Malware Detection. *IEEE Trans. Dependable Secur. Comput.* **2023**, *20*, 2127–2143. [CrossRef]
223. Zhang, S.; Du, C. Semi-Supervised Deep Learning based Network Intrusion Detection. In Proceedings of the 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Chongqing, China, 29–30 October 2020; pp. 35–40. [CrossRef]
224. Nitin Washani, S.S. Speech Recognition System: A Review. *Int. J. Comput. Appl.* **2015**, *115*, 7–10. [CrossRef]
225. Jain, N.; Rastogi, S. Speech recognition systems—A comprehensive study of concepts and mechanism . *Acta Inform. Malays.* **2019**, *3*, 1–3. [CrossRef]
226. Malik, M.; Malik, M.K.; Mehmood, K.; Makhdoom, I. Automatic speech recognition: A survey. *Multimed. Tools Appl.* **2021**, *80*, 9411–9457. [CrossRef]
227. Gaikwad, S.K.; Gawali, B.W.; Yannawar, P.Y. Article:A Review on Speech Recognition Technique. *Int. J. Comput. Appl.* **2010**, *10*, 16–24. [CrossRef]
228. Shrawankar, U.; Thakare, V.M. Techniques for Feature Extraction In Speech Recognition System: A Comparative Study. *arXiv* **2013**, arXiv:1305.1145. [CrossRef]
229. Safeel, M.; Sukumar, T.; Shashank, K.S.; Arman, M.D.; Shashidhar, R.; Puneeth, S.B. Sign Language Recognition Techniques—A Review. In Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 6–8 November 2020; pp. 1–9. Available online: https://api.semanticscholar.org/CorpusID:230513563 (accessed on 5 May 2024).
230. Rabiner, L. Applications of speech recognition in the area of telecommunications. In Proceedings of the 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings, Santa Barbara, CA, USA, 17 December 1997; pp. 501–510. [CrossRef]
231. Cohen, J. Embedded speech recognition applications in mobile phones: Status, trends, and challenges. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 5352–5355. [CrossRef]
232. Zhang, X.; Wang, S.; Zhu, F.; Xu, Z.; Wang, Y.; Huang, J. Seq3seq Fingerprint: Towards End-to-end Semi-supervised Deep Drug Discovery. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '18, Washington, DC, USA, 29 August–1 September 2018; pp. 404–413. [CrossRef]
233. Yan, C.; Duan, G.; Zhang, Y.; Wu, F.X.; Pan, Y.; Wang, J. Predicting Drug-Drug Interactions Based on Integrated Similarity and Semi-Supervised Learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 168–179. [CrossRef]
234. Nguyen, T.P.; Ho, T.B. Detecting disease genes based on semi-supervised learning and protein-protein interaction networks. *Artif. Intell. Med.* **2012**, *54*, 63–71. [CrossRef]
235. Hao, Z.; Lu, C.; Huang, Z.; Wang, H.; Hu, Z.; Liu, Q.; Chen, E.; Lee, C. ASGN: An Active Semi-supervised Graph Neural Network for Molecular Property Prediction. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20, Virtual, 23–27 August 2020; pp. 731–752. [CrossRef]
236. Luo, X.; Zheng, S.; Jiang, Z.; Chen, Z.; Huang, Y.; Zeng, S.; Zeng, X. Semi-supervised deep learning for molecular clump verification. *Astron. Astrophys.* **2024**, *683*, A104. [CrossRef]
237. Kuksa, P.P.; Qi, Y. Semi-Supervised Bio-Named Entity Recognition with Word-Codebook Learning. In Proceedings of the 2010 SIAM International Conference on Data Mining (SDM), Columbus, OH, USA, 29 April–1 May 2010; pp. 25–36. [CrossRef]
238. Pourreza Shahri, M.; Kahanda, I. Deep semi-supervised learning ensemble framework for classifying co-mentions of human proteins and phenotypes. *BMC Bioinform.* **2021**, *22*, 500. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.