DECOUPLING DIRECTION AND MAGNITUDE IN FEDERATED REMOTE SENSING OBJECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Federated learning provides a practical paradigm for collaborative remote sensing object detection without centralizing raw data, yet it is severely hindered by client heterogeneity. This heterogeneity manifests as a dual bottleneck: local detectors struggle to isolate discriminative object cues from complex aerial backgrounds, and conflicting classifier updates across clients cause destructive vector cancellation during aggregation, which shrinks classifier norms and severely attenuates prediction confidence. To address this, we propose CoRD-Fed, a dual-level federated framework that synergistically resolves both issues. On the client side, an object-centric region distillation module restricts knowledge transfer to informative target regions via ground-truth-constrained matching and CAM-guided spatial–channel alignment, effectively suppressing background interference. On the server side, a True Energy-Restored Consensus Aggregation (TERCA) strategy explicitly decouples update direction from magnitude. It estimates category-wise directional consensus to suppress conflicting gradients, then restores the aggregated classifier norm using the uncompromised sum of individual update energies, thereby preventing confidence collapse. Extensive experiments on DIOR and DOTA under severe non-IID partitioning demonstrate that CoRD-Fed consistently outperforms state-of-the-art baselines, yielding notable improvements in small-object detection and average recall.

1 INTRODUCTION

Remote sensing object detection underpins a broad range of applications, from traffic monitoring and disaster assessment to urban planning (Cheng et al., 2023; Guo et al., 2023). In practice, remote sensing imagery is typically scattered across multiple institutions and geographic regions, and cannot be directly centralized owing to privacy, security, or policy constraints (Zhu et al., 2017). Federated learning (FL) (McMahan et al., 2017) offers a natural solution by enabling collaborative model training without raw-data sharing, yet applying FL to remote sensing object detection introduces a distinctive difficulty: extreme client heterogeneity. Local datasets exhibit highly non-independent and identically distributed (non-IID) category compositions, scene layouts, and object scale distributions, all further amplified by geographic variation (Kairouz & McMahan, 2021). As visualized in Appendix A, object category distributions across clients become severely fragmented under realistic federated partitioning. Such heterogeneity gives rise to a *dual bottleneck* that simultaneously degrades client-side representation learning and server-side model aggregation (Li et al., 2020b; Qi et al., 2023; Tan et al., 2022).

On the client side, local detectors struggle to acquire clean, object-centric representations. Although knowledge distillation (KD) has been widely adopted to transfer knowledge from a stronger teacher (Zheng et al., 2022; Wang et al., 2024; Yang et al., 2022; 2024; Wang et al., 2025; Cao et al., 2022), most existing methods perform indiscriminate feature imitation over the entire image. In cluttered aerial scenes, this causes overwhelming background noise to dominate the distillation signal, obscuring the informative target regions that matter most for detection (Yang et al., 2024; Wang et al., 2025).

On the server side, heterogeneity introduces a subtler but equally critical failure mode. Under severe label skew, classifier updates for certain categories may drift in conflicting directions across clients (Shi et al., 2025; Li et al., 2023). When such divergent updates are fused through direct parameter

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

averaging, vector cancellation inevitably occurs, shrinking the effective weight norm of the affected categories. Because classification logit scale is positively correlated with classifier weight norm, this shrinkage attenuates prediction confidence. In dense remote sensing scenes, where detection relies heavily on confidence thresholds and Non-Maximum Suppression (NMS), such attenuation suppresses valid detections and degrades recall, particularly for small objects. In extreme cases, this category-wise degradation can escalate into a confidence collapse phenomenon (Salimans & Kingma, 2016; Kang et al., 2019; Lin et al., 2017; Ren et al., 2015).

To address both bottlenecks in a unified manner, we propose **CoRD-Fed**, a dual-level federated framework for remote sensing object detection. On the client side, we design an *object-centric region distillation* module that confines knowledge transfer to informative target regions. By combining ground-truth-constrained teacher–student matching, discrepancy-driven hard instance mining, and CAM-guided spatial–channel feature alignment, this module enables local detectors to acquire discriminative representations while filtering out complex background clutter. On the server side, we propose a *True Energy-Restored Consensus Aggregation* (TERCA) strategy that explicitly decouples update direction from magnitude. TERCA first estimates a consensus-guided direction to suppress destructive interference among conflicting client updates, then restores the aggregated update magnitude using the uncompromised sum of individual client update norms, thereby preserving classifier scale and stabilizing prediction confidence across heterogeneous categories.

Our main contributions are as follows:

- We identify and analyze the dual bottleneck of client-side background interference and server-side conflict-driven confidence attenuation in federated remote sensing object detection, and propose CoRD-Fed to address both issues within a single unified framework.
- We design an object-centric region distillation module that leverages ground-truth-aware matching, discrepancy-driven hard instance selection, and CAM-guided spatial–channel alignment to transfer pure target knowledge while suppressing background clutter.
- We introduce TERCA, a consensus-aware aggregation mechanism that decouples update direction and magnitude, restores the uncompromised classifier norm, and mitigates conflict-induced confidence attenuation.
- Extensive experiments on DIOR and DOTA under severe non-IID settings demonstrate that CoRD-Fed consistently outperforms both federated optimization and knowledge distillation baselines, with notable gains in small-object recall.

2 Related Work

2.1 Federated Learning under Heterogeneous Data

Statistical heterogeneity across clients remains the central obstacle to reliable federated training (McMahan et al., 2017; Kairouz & McMahan, 2021). A large body of work targets optimization stability: proximal regularization constrains local drift toward the global model (Li et al., 2020b; Acar et al., 2021), adaptive server-side solvers replace naive averaging with momentum-based updates (Reddi et al., 2021; Lee et al., 2023), and robust client selection strategies down-weight statistically divergent participants (Ning et al., 2024). While effective at improving convergence, these methods treat the model as a monolithic parameter block and do not distinguish the unique vulnerability of individual layers.

Beyond optimization instability, non-IID data intrinsically induces category-wise inconsistency in local classifiers. Recent studies show that classifier layers are particularly susceptible to label skew, producing scattered decision boundaries (Li et al., 2023) and destructive interference when heterogeneous updates are fused in parameter space (Fanì et al., 2024; Lee et al., 2023). Such interference is especially consequential for object detection, where prediction reliability depends not only on boundary alignment but also on the preservation of classifier scale (Gao et al., 2022). In remote sensing scenarios characterized by extreme scene variation and category imbalance, naive parameter averaging can shrink the effective update magnitude for certain categories, weakening confidence and degrading recall.

Despite these observations, most heterogeneous FL methods are developed for image classification (Zhang et al., 2023) and do not explicitly address category-wise classifier inconsistency under structured detection heads, nor the sensitivity of detection pipelines to confidence calibration. Our work fills this gap by introducing a consensus-guided aggregation mechanism that mitigates conflict-driven update shrinkage and preserves classifier scale in federated object detection.

2.2 Knowledge Distillation for Object Detection in Remote Sensing

Knowledge distillation transfers representational capacity from a strong teacher to a compact student (Hinton et al., 2015; Chen et al., 2024). In object detection, effective distillation requires structured, region-selective transfer rather than holistic feature regression (Chen et al., 2021; Yang et al., 2022; Wang et al., 2024; Deng et al., 2025). This selectivity is especially critical for remote sensing imagery, where densely packed small objects, extreme scale variation, and cluttered backgrounds make indiscriminate imitation counterproductive (Xia et al., 2018; Sun et al., 2022; Li et al., 2020a; Liu et al., 2019a;b).

Recent efforts have tailored distillation strategies to aerial scenes, including object-level supervision (Feng et al., 2024; Shi et al., 2026), bidirectional teacher–student interaction (Yang et al., 2024), relation-aware transfer (Wang et al., 2025), and generative object distillation (Zhou et al., 2026). These methods improve representation quality in centralized settings.

However, existing KD frameworks implicitly assume IID data and a stationary teacher, overlooking the aggregation-induced inconsistency inherent to federated training. In heterogeneous federated remote sensing, even if local distillation successfully refines object-centric features, uncoordinated server-side aggregation can distort classifier updates and erode the locally distilled knowledge. To bridge this gap, our framework tightly couples client-side object-centric region distillation with server-side consensus-aware classifier aggregation, ensuring that refined representations are faithfully preserved through global model fusion.

3 Method

3.1 Framework Overview

Fig. 1 illustrates the overall architecture of CoRD-Fed, which operates under a standard server–client federated paradigm. The design is motivated by the dual bottleneck identified in Sec. **??**: client-side background interference and server-side conflict-driven confidence attenuation. To address the former, we introduce an *object-centric region distillation* module (Sec. 3.2) that restricts knowledge transfer to informative target regions, enabling each local detector to learn discriminative representations free from background clutter. To address the latter, we propose the *True Energy-Restored Consensus Aggregation* (TERCA) strategy (Sec. 3.3) that decouples update direction from magnitude during global fusion, suppressing destructive interference while preserving classifier scale. The two components are complementary: the client-side module ensures that locally distilled knowledge is of high purity, while the server-side strategy ensures that such knowledge survives aggregation without distortion.

3.2 Client-Side Object-Centric Region Distillation

The region distillation module enhances local representation learning by confining knowledge transfer to informative target regions. It proceeds through three sequential stages: (i) ground-truth-constrained matching and hard instance selection, (ii) object-centric mask generation, and (iii) spatial–channel joint distillation.

3.2.1 GT-Constrained Matching and Hard Instance Selection

To ensure that knowledge transfer is anchored to object-relevant evidence, we first establish reliable teacher–student correspondences. For each ground-truth instance g_i with category label y_i, we select the teacher prediction b_i^t and student prediction b_i^s that maximize the Intersection-over-Union (IoU) under class consistency.

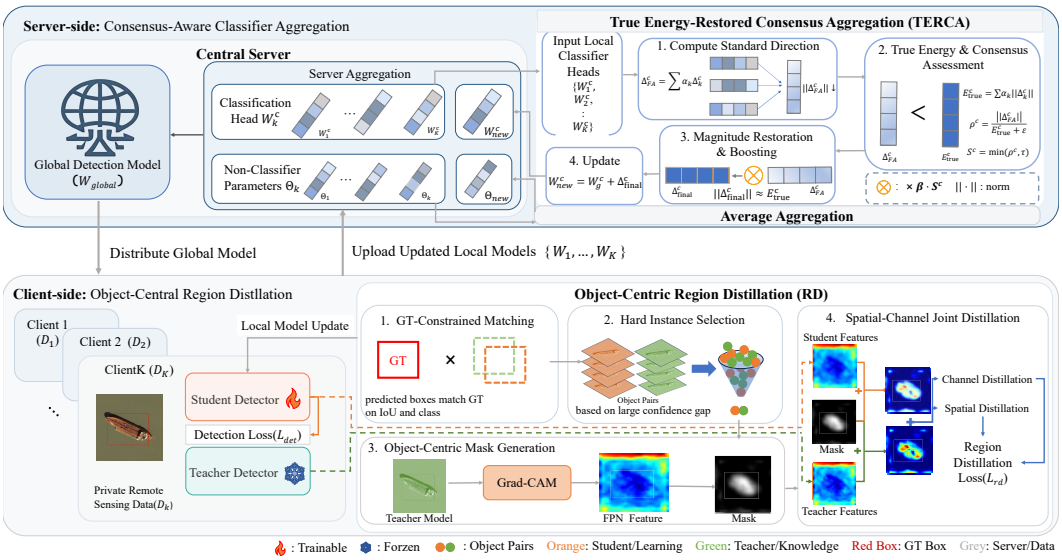


Figure 1: Overview of CoRD-Fed: client-side object-centric region distillation coupled with server-side consensus-aware classifier aggregation.

Not all matched instances contribute equally to learning. To identify the most informative cases, we compute the confidence discrepancy for each matched pair:

$$\Delta_i = |p_i^t - p_i^s|, \tag{1}$$

where p_i^t and p_i^s denote the teacher and student confidence scores, respectively. For each category, the top-K instances with the largest Δ_i are retained. A large discrepancy signals that the student has failed to capture discriminative cues already preserved by the teacher, making such instances hard yet learnable targets for distillation.

3.2.2 Object-Centric Mask Generation

For each selected teacher prediction, we construct a soft mask that localizes the most informative regions. A class-aware Grad-CAM response is generated from the teacher branch (Selvaraju et al., 2017), projected onto the feature map P_2, and constrained by the predicted bounding box:

$$M_i = \mathrm{Norm}\big(\mathrm{Pool}\big(\mathrm{CAM}(x, b_i^t)\big)\big) \odot R(b_i^t), \tag{2}$$

where $\mathrm{Pool}(\cdot)$ denotes adaptive resizing and $R(b_i^t)$ is the bounding-box mask that suppresses off-target activations. We specifically extract masks at the higher-resolution level P_2: as visualized in Appendix B, object responses at this level exhibit greater spatial compactness and activation concentration than those at coarser levels, providing a cleaner spatial prior for mask generation.

3.2.3 Spatial–Channel Joint Distillation

Given teacher and student feature maps $F \in \{F_t, F_s\} \in \mathbb{R}^{C\times H\times W}$, knowledge transfer is restricted to the masked regions defined by M_i. We perform targeted alignment along both spatial and channel dimensions rather than indiscriminate global imitation.

Channel-wise distillation. The masked features are spatially pooled into L_2-normalized channel descriptors. The c-th element of the descriptor $z \in \{z_t, z_s\}$ is computed as:

$$z^c = \mathrm{Norm}\left(\sum_{u,v} M_i(u,v)\, F^c(u,v)\right). \tag{3}$$

The channel loss $\mathcal{L}_{\mathrm{ch}}$ is defined as the mean squared error (MSE) between z_t and z_s, encouraging the student to match the teacher's semantic distribution across channels.

Spatial distillation. The masked features are compressed along the channel dimension to produce a spatial activation map $A \in \{A_t, A_s\} \in \mathbb{R}^{H \times W}$:

$$A(u,v) = \frac{1}{C} \sum_{c=1}^{C} \big(F^c(u,v) \cdot M_i(u,v)\big)^2. \tag{4}$$

The spatial loss $\mathcal{L}_{\text{sp}}$ minimizes the MSE between A_t and A_s, preserving localization-sensitive structures that are critical for dense object detection.

Overall objective. The complete local training loss combines the standard detection loss with the two distillation terms:

$$\mathcal{L}_{\text{local}} = \mathcal{L}_{\text{det}} + \lambda_{\text{rd}} \big(\mathcal{L}_{\text{sp}} + \lambda_{\text{ch}}\, \mathcal{L}_{\text{ch}}\big), \tag{5}$$

where $\mathcal{L}_{\text{det}}$ is the standard detection loss, and λ_{rd}, λ_{ch} are balancing coefficients.

3.3 Server-Side True Energy-Restored Consensus Aggregation

The TERCA strategy decouples the classification head from non-classifier components during global fusion. The server-side update proceeds in three steps: (1) standard weighted aggregation of non-classifier parameters, (2) category-wise consensus modeling, and (3) true energy-restored consensus correction.

3.3.1 Weighted Aggregation of Non-Classifier Parameters

Let $\{W_k\}_{k=1}^{K}$ denote the local models uploaded by K participating clients, and let n_k be the number of local samples on client k. The aggregation weight of client k is $\alpha_k = n_k / \sum_{j=1}^{K} n_j$. Non-classifier parameters (backbone, neck, and RPN) are aggregated via standard sample-size-weighted averaging:

$$\Theta_{\text{new}} = \sum_{k=1}^{K} \alpha_k\, \Theta_k, \tag{6}$$

where Θ_k denotes a non-classifier parameter tensor from client k. This preserves the stability of conventional federated optimization for shared representation layers.

3.3.2 Category-Wise Consensus Modeling

We now turn to the classification head, which directly governs category decision boundaries and is most susceptible to heterogeneous client updates. Let W_k^c denote the classifier weight vector for category c from client k, and W_g^c the corresponding global weight from the previous round. The local update vector is $\Delta_k^c = W_k^c - W_g^c$.

Under standard FedAvg, the global update is the weighted sum of local updates:

$$\Delta_{\text{FA}}^c = \sum_{k=1}^{K} \alpha_k\, \Delta_k^c. \tag{7}$$

When client updates drift in conflicting directions, direct summation causes severe vector cancellation, drastically shrinking $\|\Delta_{\text{FA}}^c\|_2$. To quantify this destructive interference, we define the *true energy* E_{true}^c as the uncompromised sum of individual update norms prior to any cancellation:

$$E_{\text{true}}^c = \sum_{k=1}^{K} \alpha_k \left\|\Delta_k^c\right\|_2. \tag{8}$$

We then measure the normalized consensus ratio:

$$\rho^c = \frac{\|\Delta_{\text{FA}}^c\|_2}{E_{\text{true}}^c + \epsilon}, \tag{9}$$

where ϵ is a small constant to prevent division by zero. The ratio $\rho^c \in [0,1]$ reflects the degree of directional alignment across clients for category c: a value near 1 indicates high consensus, whereas a small value signals severe directional conflict.

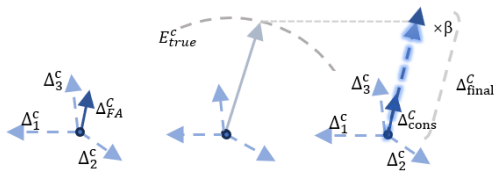


Figure 2: Illustration of TERCA: update direction and magnitude are decoupled to suppress destructive interference, restore true energy, and prevent confidence attenuation.

3.3.3 True Energy-Restored Consensus Correction

Using the consensus ratio, we first derive a consensus-guided update direction that suppresses conflicting gradients:

$$\Delta^c_{\text{cons}} = \rho^c \cdot \Delta^c_{\text{FA}}. \tag{10}$$

While Δ^c_{cons} effectively filters out destructive interference, its magnitude is heavily penalized when consensus is low. To preserve classifier scale and prevent confidence attenuation, we restore the magnitude of this consensus update to match the uncompromised true energy E^c_{true}. The restoration scale factor is:

$$S^c = \min\left(\frac{E^c_{\text{true}}}{\|\Delta^c_{\text{cons}}\|_2 + \epsilon},\ \tau\right), \tag{11}$$

where $\tau = 5.0$ acts as a safety clamp against gradient explosion under extreme inconsistency.

Finally, we introduce a confidence-boosting factor $\beta \geq 1.0$ to endow the global model with calibrated prediction confidence. The final classifier update and weight are:

$$\Delta^c_{\text{final}} = \beta \cdot S^c \cdot \Delta^c_{\text{cons}}, \tag{12}$$

$$W^c_{\text{new}} = W^c_g + \Delta^c_{\text{final}}. \tag{13}$$

As illustrated in Fig. 2, TERCA explicitly decouples direction and magnitude during aggregation. Conflicting directions are suppressed through consensus modeling, while the representational energy of the classifier is fully recovered and boosted. This mechanism addresses the root cause of confidence attenuation. The classifier bias term is aggregated via standard weighted averaging.

4 Experiments

4.1 Experimental Setup

Datasets and federated protocol. We evaluate on DIOR (20 categories) and DOTA (18 categories), with all images resized or cropped to 800×800. Following standard protocols, 20% of the data forms the global test set; the remainder is partitioned across $K = 10$ clients via a Dirichlet distribution with $\alpha = 0.1$, and each client splits its local data 8:2 for training and validation. In each of the $T = 32$ communication rounds, 30% of clients are sampled, each performing $E = 5$ local epochs before uploading updates.

Models and metrics. The student is Faster R-CNN with a ResNet-18 backbone and FPN; the local teacher shares the same detection head but uses a ResNet-101 backbone. All models are ImageNet-pretrained. We report COCO metrics ($\text{mAP}_{50:95}$, AP_{50}, AP_{75}, $\text{AP}_{S/M/L}$) and, to validate TERCA's effect on confidence attenuation, Average Recall (AR_{100}, AR_S). All values are in percentages.

4.2 Comparison with State-of-the-Art

Table 1 presents the global detection performance of CoRD-Fed compared against general FL optimization methods (FedAvg, FedProx, FedAdam, FedAWA) and distillation-based baselines (CWD, FGD, GID, DCSF). Representative qualitative comparisons are visualized in Fig. 3, with additional visualizations provided in Appendix E. We further include a Local-only setting, in which each client

Table 1: Global detection performance on DIOR and DOTA under federated settings. Best and second-best results are in **bold** and underline.

Method	DIOR								DOTA							
	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_{100}	AR_S	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_{100}	AR_S
FedAvg	46.44	75.74	49.33	16.72	35.60	53.07	56.04	23.17	25.58	46.06	24.69	10.47	28.49	36.61	35.24	18.1
FedProx	44.79	73.22	46.49	11.51	31.82	50.81	54.71	18.20	30.39	51.10	31.96	14.77	33.37	45.25	38.81	21.46
FedAdam	44.39	74.57	46.12	12.10	31.69	51.07	53.99	19.48	23.48	43.66	22.61	9.68	27.93	34.71	32.37	16.81
FedAWA	43.08	71.88	43.91	12.90	32.35	48.62	54.38	20.68	24.57	45.05	23.76	10.48	27.88	34.88	34.93	17.33
CWD	47.12	74.68	50.66	10.30	31.72	54.16	58.02	20.25	25.10	45.66	24.30	11.91	28.55	35.68	35.50	19.91
FGD	46.44	74.74	49.05	14.02	33.17	52.94	54.39	21.13	25.67	46.53	24.73	10.55	28.26	36.73	35.80	18.39
GID	31.15	49.25	34.55	7.20	25.31	33.94	39.39	10.45	18.35	31.62	18.88	6.56	22.17	27.11	25.88	10.02
DCSF	48.31	75.80	52.06	14.00	32.06	**55.54**	58.81	23.19	24.00	43.22	23.58	9.52	26.42	34.89	34.97	16.42
Ours	**48.70**	**76.00**	**52.45**	**17.30**	**37.42**	55.27	**59.27**	**25.42**	**30.85**	**51.85**	**32.45**	**15.25**	**33.85**	**45.80**	**39.54**	**23.45**

trains independently without any communication; the detailed per-client results are deferred to Appendix C. The severely degraded and highly biased performance across clients confirms that collaborative federated learning is essential under fragmented remote sensing distributions.



Figure 3: Qualitative comparison with FL baselines (top two rows) and KD baselines (bottom two rows). Red boxes denote correct detections; green boxes highlight missed targets (false negatives).

Overcoming FL Optimization Bottlenecks. On DIOR, standard FL methods exhibit limited gains. While FedProx and FedAdam improve stability via proximal terms and adaptive solvers, they still rely on direct parameter averaging. Under severe label skew, this induces update conflicts and magnitude shrinkage, degrading confidence. FedAWA mitigates this partially via adaptive weighting but lacks explicit category-wise conflict resolution. In contrast, TERCA explicitly decouples update direction and magnitude, restoring true update energy and preserving classifier scale. Consequently, CoRD-Fed achieves substantial AP improvements of 2.26 over FedAvg and 5.62 over FedAWA.

Overcoming KD Background Interference. Compared to distillation baselines, a different limitation emerges. Methods like CWD and GID rely on global feature imitation, introducing substantial background noise that degrades small-object detection. While DCSF partially disentangles features, it lacks strict object-level selectivity. Conversely, our object-centric region distillation explicitly restricts knowledge transfer to informative regions, suppressing background interference and preserving fine-grained spatial cues. This translates to consistent improvements in small-object metrics (AP_S and AR_S).

Robustness under Extreme Heterogeneity. The advantages of CoRD-Fed are most pronounced on the DOTA dataset, which exhibits more severe heterogeneity and scale variation. Here, both optimization- and distillation-based baselines suffer from intensified conflicts, leading to increased missed detections. Notably, CoRD-Fed achieves the most significant gains in AR and AR_S, indicating a substantial reduction in false negatives. This provides strong empirical support for our core hypothesis: by mitigating conflict-induced norm shrinkage and restoring classifier energy, TERCA effectively maintains prediction confidence and recovers suppressed true positives in dense scenes. A formal theoretical analysis of TERCA, including its direction-preservation, exact energy-restoration, and FedAvg-consistency properties, is provided in Appendix D.

4.3 Ablation Study

We conduct comprehensive ablations to evaluate the individual and joint contributions of the Region Distillation (RD) module and the TERCA strategy.

4.3.1 Overall Contribution of RD and TERCA

Table 2 details the progressive integration of our modules into the FedAvg baseline. Applying either module individually yields improvements: RD enhances local representation purity, while TERCA resolves conflict-driven confidence attenuation. Notably, TERCA alone significantly boosts recall (AR_{100} from 56.04 to 58.12, AR_S from 23.17 to 23.56), validating its capacity to salvage true positives discarded by naive averaging. Combining both achieves optimal performance across all metrics, demonstrating that local object-centric distillation and global true-energy restoration are highly complementary.

Table 2: Overall ablation of the proposed modules.

Variant	mAP	AP_{50}	AP_S	AP_M	AP_L	AR_{100}	AR_S
FedAvg	46.45	75.74	16.72	35.60	53.07	56.04	23.17
TERCA	47.41	74.60	16.27	35.08	53.68	58.12	23.56
RD	47.33	75.11	16.47	37.18	54.11	59.28	24.25
RD + TERCA	48.70	76.00	17.30	37.42	59.27	59.27	25.42

4.3.2 Analysis of Local Distillation Variants

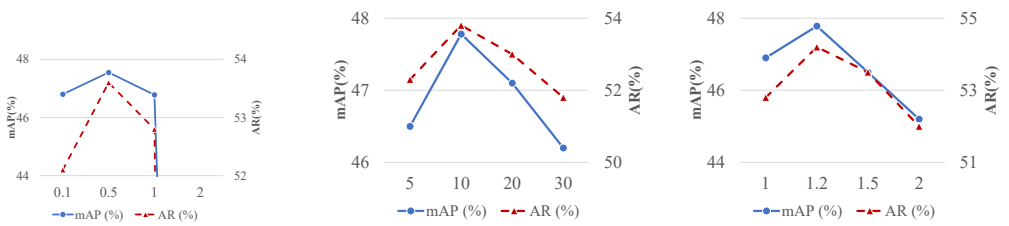
Table 3 investigates the spatial and channel branches of RD. Both individual branches outperform FedAvg, with channel-wise distillation notably improving small-object detection and spatial distillation enhancing localization. When coupled with TERCA, both variants benefit from stabilized classifier aggregation. The full RD configuration outperforms either branch alone, and integrating full RD with TERCA achieves the best overall performance, confirming the synergy between local region selectivity and global energy restoration.

Table 3: Ablation of local distillation variants with/without TERCA.

Variant	mAP	AP_{50}	AP_S	AP_M	AP_L	AR_{100}	AR_S
FedAvg	46.45	75.74	16.72	35.60	53.07	56.04	23.17
RD_{Channel}	47.93	75.60	17.96	37.07	53.92	59.62	25.43
RD_{Channel} + TERCA	47.78	75.18	17.25	37.10	54.05	58.93	25.58
RD_{Spatial}	47.18	74.64	16.38	35.81	53.25	58.33	24.90
RD_{Spatial} + TERCA	47.72	74.62	17.13	36.82	54.67	58.60	24.57
RD + TERCA	48.70	76.00	17.30	37.42	59.27	59.27	25.42

4.4 Hyperparameter Analysis

We investigate the sensitivity of CoRD-Fed to three pivotal hyperparameters on the DIOR dataset (Fig. 4).



(a) Effect of λ_{rd} (b) Effect of λ_{ch} (c) Effect of β

Figure 4: Hyperparameter sensitivity on DIOR. mAP and AR illustrate the accuracy-confidence trade-off across: (a) Region distillation weight λ_{rd}. (b) Channel balance λ_{ch}. (c) Confidence boost β.

Region distillation weight λ_{rd}. This parameter controls the strength of local region distillation (Fig. 4a). A small λ_{rd} provides insufficient regularization, limiting the transfer of discriminative cues. Conversely, an excessively large λ_{rd} (> 1) triggers a sharp performance drop due to objective conflict: the over-weighted distillation term dominates the optimization, disrupting task-specific representation learning. A moderate λ_{rd} strikes the optimal balance.

Channel balance coefficient λ_{ch}. This coefficient governs the trade-off between spatial and channel-wise distillation (Fig. 4b). A small λ_{ch} insufficiently captures channel-level semantic dependencies, while an overly large λ_{ch} dilutes the focus on spatial localization, which is critical for dense objects. The optimal performance at a moderate λ_{ch} confirms that balanced alignment of both semantics is essential.

Confidence-boosting factor β. The factor β in TERCA controls confidence calibration after true energy restoration (Fig. 4c). Setting $\beta = 1.0$ strictly recovers the uncompromised magnitude. A slightly larger β (1.2) provides a necessary confidence boost to combat strict NMS thresholding, improving both mAP and AR. However, excessively large values induce over-confidence and increased false positives, degrading overall performance.

5 Conclusions

We presented CoRD-Fed, a dual-level federated framework for remote sensing object detection under severe non-IID conditions. CoRD-Fed attacks the dual bottleneck of federated detection from both ends: on the client side, an object-centric region distillation module confines knowledge transfer to informative target regions via ground-truth-constrained matching, discrepancy-driven hard instance mining, and CAM-guided spatial–channel alignment, thereby filtering background clutter; on the server side, the TERCA strategy decouples update direction from magnitude, suppressing destructive interference among conflicting client updates while restoring the aggregated classifier norm to its uncompromised true energy. Our theoretical analysis further shows that TERCA preserves the FedAvg update direction, restores update energy exactly, and reduces to standard FedAvg in the perfect-consensus regime (Appendix D).

Experiments on DIOR and DOTA under severe non-IID partitioning demonstrate that CoRD-Fed consistently outperforms federated optimization and knowledge distillation baselines, with the most pronounced gains in average recall and small-object recall—empirically confirming that restoring the uncompromised classifier norm recovers true positives suppressed by naive parameter averaging. Ablations further verify that region distillation and energy-restored aggregation are complementary rather than redundant. Future work will extend the energy-restoration principle to regression heads and other dense prediction tasks such as instance segmentation, and establish convergence guarantees for the direction–magnitude decoupled update rule under extreme heterogeneity.

AI use statement

In this work, we used generative AI tools for writing assistance, specifically to polish the language, correct grammar, and improve the readability of the manuscript. We did not use generative AI tools

for research ideation or execution, for retrieval and discovery (e.g., locating related work), for drafting scientific sections of the paper, for generating synthetic datasets, or for proving mathematical claims, and the remaining mandatory-disclosure tasks are not applicable to this work. Additionally, no recommended-disclosure tasks beyond the aforementioned writing assistance apply to this work. We have reviewed all AI-assisted work: every AI-suggested edit was manually inspected, revised, and approved by the authors, and all technical claims, theoretical derivations, and experimental results were independently produced and verified by the authors against the original code and data. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI. See the ICLR 2027 AI Policy for Authors for more details.

Ethics Statement

This work involves no human subjects, personal data, or IRB-regulated study. Experiments use only the publicly released DIOR and DOTA benchmarks in accordance with their licenses. The proposed federated framework is privacy-preserving by design, as raw imagery never leaves local clients. Intended applications are civilian (e.g., disaster assessment, traffic monitoring, urban management). The authors declare no conflicts of interest, and all results are reported honestly and reproducibly.

Reproducibility Statement

We have made concerted efforts to ensure the reproducibility of our work. The complete algorithmic design of CoRD-Fed, including the object-centric region distillation module and the TERCA aggregation strategy, is specified in Section 3 together with all governing equations and key hyperparameter values (e.g., the safety clamp τ and the boost factor β). All experimental configurations—dataset preprocessing and non-IID partitioning, model architectures, the federated protocol, and evaluation metrics—are detailed in Section 4.1, and hyperparameter sensitivity analyses are provided in Section 4.4. The theoretical properties of TERCA are stated as formal propositions with explicit assumptions and complete proofs in Appendix D. Per-client results under the Local-only setting and additional activation and qualitative visualizations are reported in Appendices C, B, and E. The source code implementing our framework, together with configuration files and scripts for data partitioning, will be made publicly available upon acceptance.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

References

Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. *arXiv preprint arXiv:2111.04263*, 2021.

Weihan Cao, Yifan Zhang, Jianfei Gao, Anda Cheng, Ke Cheng, and Jian Cheng. Pkd: General distillation framework for object detectors via pearson correlation coefficient. *Advances in Neural Information Processing Systems*, 35:15394–15406, 2022.

Pengguang Chen, Shu Liu, Hengshuang Zhao, and Jiaya Jia. Distilling knowledge via knowledge review. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5008–5017, 2021.

Shengchao Chen, Ting Shu, Huan Zhao, Jiahao Wang, Sufen Ren, and Lina Yang. Free lunch for federated remote sensing target fine-grained classification: A parameter-efficient framework. *Knowledge-Based Systems*, 294:111694, 2024. ISSN 0950-7051. doi: https://doi.org/10.1016/j.knosys.2024.111694. URL https://www.sciencedirect.com/science/article/pii/S0950705124003290.

Gong Cheng, Qingyang Li, Guangxing Wang, Xingxing Xie, Lingtong Min, and Junwei Han. Sfrnet: Fine-grained oriented object recognition via separate feature refinement. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–10, 2023. doi: 10.1109/TGRS.2023.3277626.

Peihua Deng, Jiehua Zhang, Xichun Sheng, Chenggang Yan, Yaoqi Sun, Ying Fu, and Liang Li. Multi-granularity class prototype topology distillation for class-incremental source-free unsupervised domain adaptation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 30566–30576, 2025.

Eros Fanì, Raffaello Camoriano, Barbara Caputo, and Marco Ciccone. Accelerating heterogeneous federated learning with closed-form classifiers. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Hangtao Feng, Lu Zhang, Xu Yang, and Zhiyong Liu. Enhancing class-incremental object detection in remote sensing through instance-aware distillation. *Neurocomputing*, 583:127552, 2024.

Liang Gao, Huazhu Fu, Li Li, Yingwen Chen, Ming Xu, and Cheng-Zhong Xu. Feddc: Federated learning with non-iid data via local drift decoupling and correction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10112–10121, 2022.

Bo Guo, Ruixiang Zhang, Haowen Guo, Wen Yang, Huai Yu, Peng Zhang, and Tongyuan Zou. Fine-grained ship detection in high-resolution satellite images with shape-aware feature learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1914–1926, 2023. doi: 10.1109/JSTARS.2023.3241969.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Peter Kairouz and H Brendan McMahan. Advances and open problems in federated learning. *Foundations and trends in machine learning*, 14(1-2):1–210, 2021.

Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. *arXiv preprint arXiv:1910.09217*, 2019.

Sunwoo Lee, Tuo Zhang, and A Salman Avestimehr. Layer-wise adaptive model aggregation for scalable federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8491–8499, 2023.

Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS journal of photogrammetry and remote sensing*, 159:296–307, 2020a.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020b.

Zexi Li, Xinyi Shang, Rui He, Tao Lin, and Chao Wu. No fear of classifier biases: Neural collapse inspired federated learning with synthetic and fixed classifier. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5319–5329, 2023.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

Yifan Liu, Ke Chen, Chris Liu, Zengchang Qin, Zhenbo Luo, and Jingdong Wang. Structured knowledge distillation for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2604–2613, 2019a.

Yufan Liu, Jiajiong Cao, Bing Li, Chunfeng Yuan, Weiming Hu, Yangxi Li, and Yunqiang Duan. Knowledge distillation via instance relationship graph. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7096–7104, 2019b.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. Pmlr, 2017.

Zhiyuan Ning, Chunlin Tian, Meng Xiao, Wei Fan, Pengyang Wang, Li Li, Pengfei Wang, and Yuanchun Zhou. Fedgcs: a generative framework for efficient client selection in federated learning via gradient-based optimization. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, IJCAI '24, 2024. ISBN 978-1-956792-04-1. doi: 10.24963/ijcai.2024/526. URL https://doi.org/10.24963/ijcai.2024/526.

Zhuang Qi, Lei Meng, Zitan Chen, Han Hu, Hui Lin, and Xiangxu Meng. Cross-silo prototypical calibration for federated learning with non-iid data. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, pp. 3099–3107, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701085. doi: 10.1145/3581783.3612481. URL https://doi.org/10.1145/3581783.3612481.

Sashank J. Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečný, Sanjiv Kumar, and Hugh Brendan McMahan. Adaptive federated optimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=LkFG3lB13U5.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

Tim Salimans and Durk P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. *Advances in neural information processing systems*, 29, 2016.

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626, 2017. doi: 10.1109/ICCV.2017.74.

Changlong Shi, He Zhao, Bingjie Zhang, Mingyuan Zhou, Dandan Guo, and Yi Chang. Fedawa: adaptive optimization of aggregation weights in federated learning using client vectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30651–30660, 2025.

Yutian Shi, Guoquan Li, Zhilong Shen, Hongying Meng, and Yu Pang. Banet: Enhancing weakly aligned multimodal object detection via balanced bidirectional alignment network. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–14, 2026. doi: 10.1109/TGRS.2026.3674946.

Xian Sun, Peijin Wang, Zhiyuan Yan, Feng Xu, Ruiping Wang, Wenhui Diao, Jin Chen, Jihao Li, Yingchao Feng, Tao Xu, et al. Fair1m: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184:116–130, 2022.

Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated prototype learning across heterogeneous clients. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8432–8440, 2022.

Jiabao Wang, Yuming Chen, Zhaohui Zheng, Xiang Li, Ming-Ming Cheng, and Qibin Hou. Crosskd: Cross-head knowledge distillation for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16520–16530, 2024.

Ruiqing Wang, Yifan Fu, Yang Xu, Zebin Wu, and Zhihui Wei. Dual prediction-guided distillation for object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3974–3983, 2018.

Xi Yang, Sheng Zhang, and Weichao Yang. Two-way assistant: A knowledge distillation object detection method for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 2024.

Zhendong Yang, Zhe Li, Xiaohu Jiang, Yuan Gong, Zehuan Yuan, Danpei Zhao, and Chun Yuan. Focal and global knowledge distillation for detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4643–4652, 2022.

Xiaokang Zhang, Boning Zhang, Weikang Yu, and Xudong Kang. Federated deep learning with prototype matching for object extraction from very-high-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023.

Zhaohui Zheng, Rongguang Ye, Ping Wang, Dongwei Ren, Wangmeng Zuo, Qibin Hou, and Ming-Ming Cheng. Localization distillation for dense object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9407–9416, 2022.

Zhuangzhuang Zhou, Xuewu Zhao, Xuezhuan Zhao, Yingying Zhu, Zunjin Zhao, and Lingling Li. Gendist: Generative-driven knowledge distillation for small object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–16, 2026. doi: 10.1109/TGRS.2026.3676015.

Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE geoscience and remote sensing magazine*, 5(4):8–36, 2017.

A Visualization of Non-IID Category Distribution

To provide an intuitive understanding of the severe data heterogeneity in our federated setting, we visualize the object category distribution across clients. Fig. 5 illustrates the category distribution on the DOTA dataset under a Dirichlet partition with concentration parameter $\alpha = 0.1$. The highly fragmented and imbalanced distributions across clients clearly demonstrate the challenges of non-IID data in federated remote sensing object detection, which motivates both our client-side region distillation and server-side consensus-aware aggregation designs.

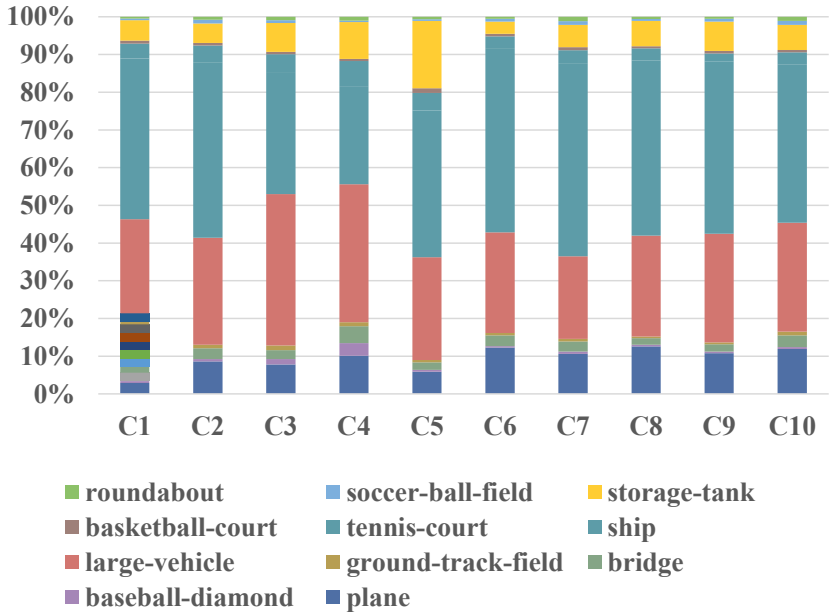


Figure 5: Category distribution across clients on the DOTA dataset under a severe non-IID setting ($\alpha = 0.1$).

B Analysis of Object Activation across FPN Levels

This section provides visual evidence for the design choice of extracting object-centric masks at the P_2 level. Fig. 6 visualizes the object-specific activation responses at different FPN levels for the same target instance. The first row shows the P2-level results: the feature map (a), the corresponding activation mask (b), and the overlaid visualization (c); the second row shows the corresponding P3-level results (d–f). Compared with P3, the P2-level responses are markedly more compact and concentrated around the target instance, confirming that higher-resolution features provide cleaner spatial priors for mask generation and justifying our choice of P_2 as the mask extraction level.

C Local-Only Training Baseline

To establish a lower performance bound and to highlight the necessity of collaboration, we evaluate a *Local-only* setting in which each client trains its detector independently without any communication with the server. Tables 4 and 5 report the per-client detection performance on DIOR and DOTA, respectively. Two observations stand out. First, isolated training yields severely degraded performance on both datasets: for instance, client-wise AP on DIOR ranges only from 30.25 to 35.46, far below the federated results reported in Table 1. Second, the performance is highly biased across clients, directly reflecting the fragmented category and scene distributions that each client observes in isolation. Taken together, these results confirm that, under severely fragmented remote sensing data distributions, collaborative federated learning is not merely beneficial but essential.

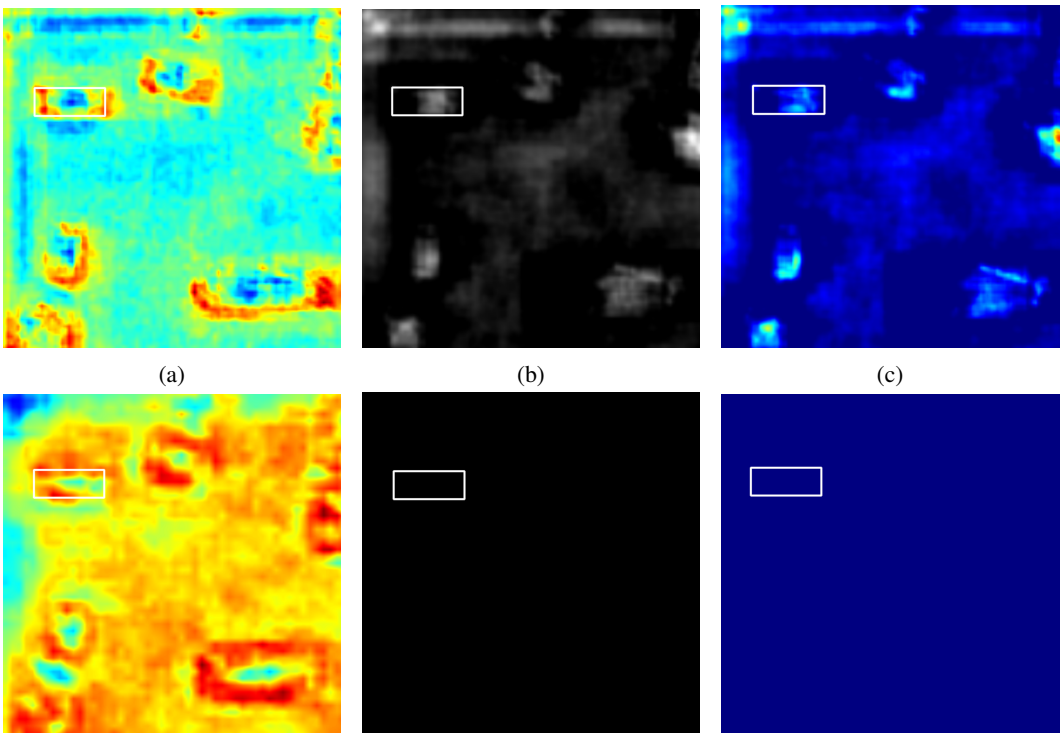


Figure 6: Visualization of object-specific activation responses at different FPN levels for the same target instance. Top row: P2-level feature map (a), activation mask (b), and overlaid visualization (c). Bottom row: corresponding P3-level results (d–f). Compared with P3, P2 yields more compact and discriminative object responses.

Table 4: Client-wise detection performance on the DIOR dataset under the Local-only setting.

Client	**AP**	**AP**$_{50}$	**AP**$_{75}$	**AP**$_S$	**AP**$_M$	**AP**$_L$
C0	33.75	55.25	34.41	3.43	20.38	39.88
C1	32.18	54.60	32.76	3.51	20.12	40.00
C2	31.95	53.07	32.97	2.45	17.75	38.50
C3	32.83	54.97	34.03	3.84	20.12	39.52
C4	31.76	53.82	33.25	3.19	20.17	37.99
C5	32.25	53.75	33.63	3.38	20.45	39.68
C6	31.63	54.20	32.58	2.62	17.98	39.32
C7	32.06	54.40	32.67	3.11	20.38	37.98
C8	35.46	58.61	36.92	3.90	22.36	42.16
C9	30.25	50.98	30.97	3.36	20.23	35.71

D THEORETICAL ANALYSIS OF THE TERCA AGGREGATION

We provide a formal analysis of the True Energy-Restored Consensus Aggregation (TERCA) strategy. For clarity, we restate the key quantities. For category c, let $\Delta_k^c = W_k^c - W_g^c$ denote the local classifier update of client k, and let $\alpha_k \geq 0$ with $\sum_{k=1}^K \alpha_k = 1$ be the aggregation weights. The FedAvg update, true energy, consensus ratio, consensus update, restoration factor, and final update are defined as:

$$\Delta_{\mathrm{FA}}^c = \sum_{k=1}^K \alpha_k \Delta_k^c, \quad E_{\mathrm{true}}^c = \sum_{k=1}^K \alpha_k \left\|\Delta_k^c\right\|_2, \quad \rho^c = \frac{\left\|\Delta_{\mathrm{FA}}^c\right\|_2}{E_{\mathrm{true}}^c + \epsilon}, \tag{14}$$

Table 5: Client-wise detection performance on the DOTA dataset under the Local-only setting.

Client	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
C0	21.86	41.38	20.81	6.28	22.49	20.14
C1	19.63	37.33	18.29	9.24	26.38	17.06
C2	19.70	37.67	18.91	6.20	24.87	17.26
C3	18.67	37.08	17.62	6.44	26.07	14.40
C4	19.83	36.89	19.70	6.67	24.70	17.81
C5	16.48	32.11	14.73	8.96	20.96	15.00
C6	16.69	32.87	14.46	8.17	23.95	16.88
C7	21.45	39.47	21.41	6.62	23.09	20.07
C8	23.81	43.42	22.80	8.92	25.99	21.59
C9	20.25	38.73	18.80	7.95	26.13	18.69

$$\Delta^c_{\text{cons}} = \rho^c \Delta^c_{\text{FA}}, \quad S^c = \min\left(\frac{E^c_{\text{true}}}{\|\Delta^c_{\text{cons}}\|_2 + \epsilon},\ \tau\right), \quad \Delta^c_{\text{final}} = \beta\, S^c\, \Delta^c_{\text{cons}}. \tag{15}$$

Proposition 1 (Boundedness of the consensus ratio) *For every category c with $E^c_{\text{true}} > 0$, the consensus ratio satisfies $\rho^c \in [0, 1]$. Moreover, in the limit $\epsilon \to 0$, $\rho^c = 1$ if and only if all nonzero local updates $\{\Delta^c_k\}$ are positively collinear, i.e., perfectly aligned in direction.*

Proof. Since $\alpha_k \geq 0$, the triangle inequality yields $\|\Delta^c_{\text{FA}}\|_2 = \|\sum_k \alpha_k \Delta^c_k\|_2 \leq \sum_k \alpha_k \|\Delta^c_k\|_2 = E^c_{\text{true}} \leq E^c_{\text{true}} + \epsilon$, which gives $\rho^c \leq 1$; nonnegativity of the norm gives $\rho^c \geq 0$. Equality in the triangle inequality holds if and only if all nonzero vectors $\alpha_k \Delta^c_k$ share a common direction, which establishes the claim.

Proposition 1 justifies interpreting ρ^c as a normalized measure of directional agreement: $\rho^c \to 1$ indicates full consensus, whereas $\rho^c \to 0$ indicates near-complete vector cancellation.

Proposition 2 (Direction preservation) *For any $\rho^c > 0$, the final TERCA update Δ^c_{final} is positively collinear with the FedAvg update Δ^c_{FA}, i.e., $\Delta^c_{\text{final}} / \|\Delta^c_{\text{final}}\|_2 = \Delta^c_{\text{FA}} / \|\Delta^c_{\text{FA}}\|_2$.*

Proof. By construction, $\Delta^c_{\text{cons}} = \rho^c \Delta^c_{\text{FA}}$ with $\rho^c > 0$, and both $S^c > 0$ and $\beta \geq 1 > 0$ are positive scalars. Hence $\Delta^c_{\text{final}} = (\beta S^c \rho^c)\, \Delta^c_{\text{FA}}$ is a positive scalar multiple of Δ^c_{FA}, which preserves its direction.

Proposition 2 shows that TERCA never introduces directional bias: the global classifier continues to move along the consensus (weighted-average) direction, and TERCA solely recalibrates the step magnitude. This property is essential for maintaining alignment with the federated optimization objective.

Proposition 3 (Exact energy restoration) *Suppose the safety clamp is inactive, i.e., $E^c_{\text{true}} / (\|\Delta^c_{\text{cons}}\|_2 + \epsilon) < \tau$. Then the norm of the final update satisfies $\|\Delta^c_{\text{final}}\|_2 = \beta\, E^c_{\text{true}} \cdot \frac{\|\Delta^c_{\text{cons}}\|_2}{\|\Delta^c_{\text{cons}}\|_2 + \epsilon}$, which converges to $\beta\, E^c_{\text{true}}$ as $\epsilon \to 0$. In particular, whenever $\epsilon \ll \|\Delta^c_{\text{cons}}\|_2$, the restored magnitude is independent of the consensus ratio ρ^c.*

Proof. Substituting $S^c = E^c_{\text{true}} / (\|\Delta^c_{\text{cons}}\|_2 + \epsilon)$ into $\Delta^c_{\text{final}} = \beta S^c \Delta^c_{\text{cons}}$ and taking norms yields the result directly.

Proposition 3 is the core guarantee of TERCA: even under severe directional conflict ($\rho^c \to 0$), the magnitude of the aggregated classifier update is restored to the uncompromised energy E^c_{true} (up to the boost β), rather than being shrunk by vector cancellation as in FedAvg. This directly breaks the chain "directional conflict $\to$ norm shrinkage $\to$ confidence attenuation $\to$ missed detections."

Proposition 4 (Consistency with FedAvg) *In the perfect-consensus regime (all nonzero Δ^c_k positively collinear) and with $\beta = 1$, TERCA reduces exactly to standard FedAvg, i.e., $\Delta^c_{\text{final}} = \Delta^c_{\text{FA}}$.*

Proof. Perfect consensus implies $\|\Delta^c_{\text{FA}}\|_2 = E^c_{\text{true}}$ (the equality case of Proposition 1), hence $\rho^c = 1$ as $\epsilon \to 0$, $\Delta^c_{\text{cons}} = \Delta^c_{\text{FA}}$, and $\|\Delta^c_{\text{cons}}\|_2 = E^c_{\text{true}}$. Therefore $S^c = \min\big(E^c_{\text{true}} / (E^c_{\text{true}} + \epsilon),\ \tau\big) = 1$ for sufficiently small ϵ (since $\tau \geq 1$), and $\Delta^c_{\text{final}} = \beta S^c \Delta^c_{\text{cons}} = \Delta^c_{\text{FA}}$ when $\beta = 1$.

Proposition 4 establishes that TERCA is a *conservative* extension of FedAvg: it intervenes only when directional conflict is present, and leaves the aggregation unchanged in benign (IID or directionally aligned) regimes.

Remark 1 (Role of the safety clamp) *As $\rho^c \to 0$, the direction of Δ^c_{FA} itself becomes unreliable, being dominated by near-cancelling residuals that may amplify noise. Without the clamp, $S^c \to \infty$ would arbitrarily amplify this unreliable direction. The clamp τ enforces a trust region: $\|\Delta^c_{\mathrm{final}}\|_2 \leq \beta\,\tau\,\|\Delta^c_{\mathrm{cons}}\|_2 = \beta\,\tau\,\rho^c\,\|\Delta^c_{\mathrm{FA}}\|_2$, guaranteeing bounded updates even under extreme inconsistency.*

Remark 2 (Connection to confidence preservation) *The scale of classification logits is positively correlated with the classifier weight norm (Salimans & Kingma, 2016; Kang et al., 2019). Under FedAvg, conflict-induced cancellation yields $\|\Delta^c_{\mathrm{FA}}\|_2 \ll E^c_{\mathrm{true}}$, stunting the growth of $\|W^c_{\mathrm{new}}\|_2$ and thereby attenuating logit scales and prediction confidence. In dense detection, where confidence thresholds and NMS govern the survival of proposals, this suppresses valid detections and degrades recall. By Proposition 3, TERCA maintains $\|\Delta^c_{\mathrm{final}}\|_2 \approx \beta\,E^c_{\mathrm{true}}$ regardless of the consensus level, ensuring healthy evolution of weight norms and hence stable confidence. This provides a theoretical explanation for the consistent AR and AR_S gains observed in Table 1.*

Remark 3 (Effect of the boost factor) *The factor $\beta \geq 1$ provides a controlled confidence margin on top of exact restoration. Empirically (Fig. 4c), a mild boost ($\beta = 1.2$) compensates for strict NMS thresholding in dense scenes, whereas excessive β induces over-confidence and additional false positives. The clamp τ and the validation-based selection of β jointly keep the mechanism within a stable regime.*

E Additional Qualitative Comparisons

We provide additional qualitative results to further demonstrate the effectiveness of CoRD-Fed in dense and heterogeneous remote sensing scenes. As shown in Fig. 7, our method consistently reduces missed detections (highlighted in green) compared to both FL optimization and KD baselines, validating its superior recall capabilities.

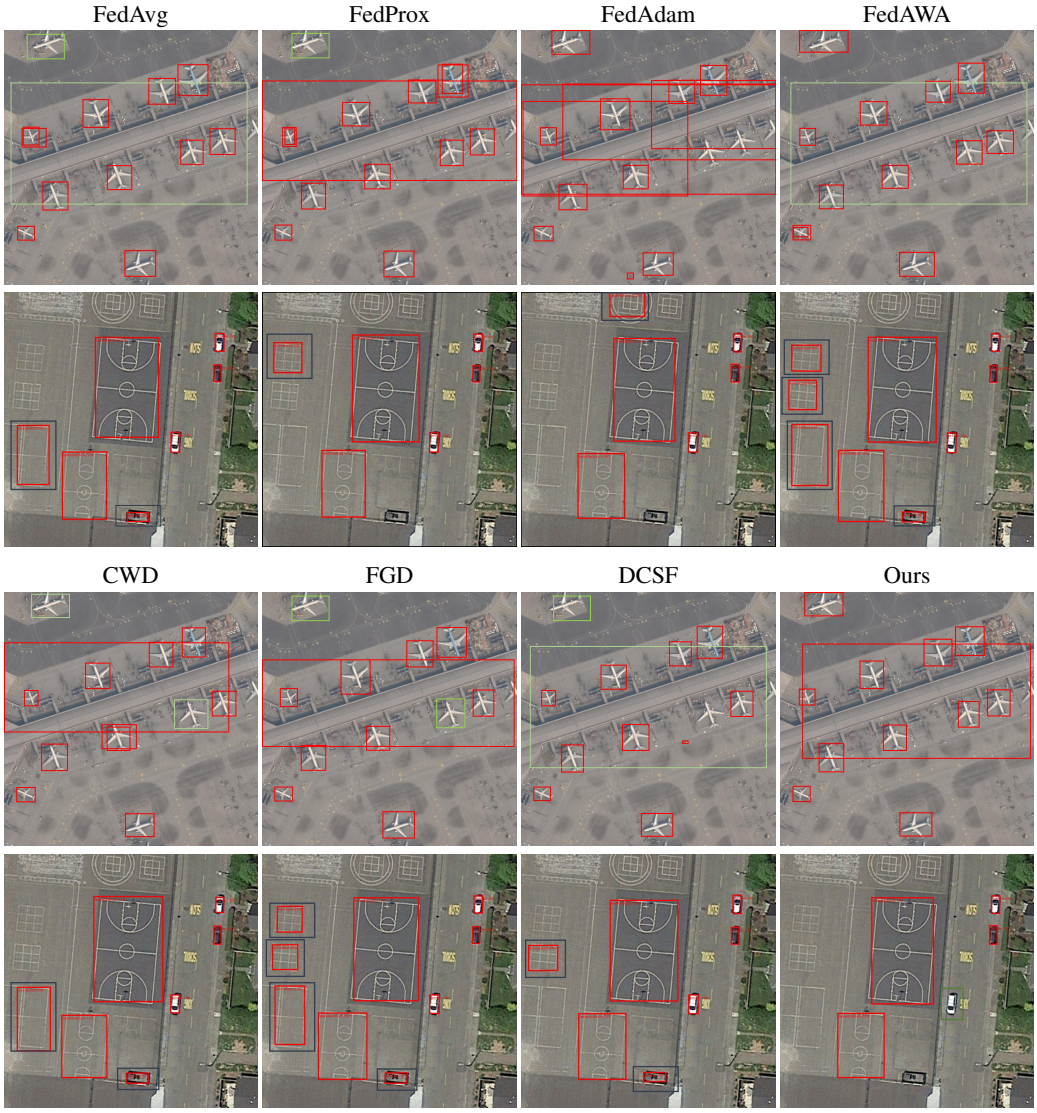


Figure 7: Additional qualitative comparisons across baselines. The top row presents the detection results of traditional federated learning methods, while the bottom row illustrates the performance of representative knowledge distillation approaches alongside our proposed method. Red bounding boxes denote correctly detected objects, whereas green bounding boxes highlight missed targets (false negatives). CoRD-Fed demonstrates superior recall in dense scenes.