ANALYTIC FEDERATED LEARNING IS VIABLE FOR REGION-BASED OBJECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Federated object detection typically requires repeated exchanges of model updates, while one-shot AFL promises a single upload followed by a closed-form solve. Whether this promise extends beyond fixed-output classification is unclear because detection must handle variable proposal sets, severe class imbalance, and coupled classification and localization. We introduce an analytically solved-head framework that compiles each client's supervision into additive Gram, cross-moment, and count statistics over a shared frozen proposal representation. Aggregating these statistics exactly recovers the pooled classification and box-regression objectives for any horizontal partition, without transmitting raw images, proposal features, gradients, or locally trained detector weights. With one configuration selected on VOC2007 train/validation, the resulting detector achieves 78.48 ± 0.51 AP50 on VOC2007 test and 75.32 ± 0.24 on VOC2012 validation without retuning; its float32 sufficient statistics are $52.88\times$ smaller than the corresponding raw JPEGs. Additional full-split controls show stable prior calibration for $\alpha \in [4, 8]$ and a 22.2-ms server solve for the selected $q = 193$ system. These results show that partition-invariant, one-shot AFL can support a structured, multi-function vision task while retaining useful detection accuracy across dataset releases, establishing object detection as a viable target for AFL.

1 INTRODUCTION

Federated learning enables collaborative training without sharing raw data, but mainstream methods repeatedly exchange gradients or model updates (McMahan et al., 2017; Li et al., 2020). Analytic federated learning (AFL) instead lets each client upload additive normal-equation statistics once, from which the server reconstructs a centralized closed-form solution (He et al., 2025). Recent extensions add layer-wise analytic blocks or structured nonlinear heads (Tang et al., 2026; Bacellar et al., 2026), yet evidence remains concentrated on fixed-output tasks, principally image classification. Exact analytic recovery for object detection remains open.

Detection is a stricter test. Each image induces a variable number of candidate regions. Background proposals greatly outnumber foreground classes. Prediction couples categorical ranking with geometric localization. Local class balancing can also change the pooled objective under non-IID partitions. A useful protocol must handle all four properties without adding a preliminary round or transmitting proposal features.

Here we make partition-invariant AFL concrete for object detection. The analytically solved-head framework freezes proposal and region extraction and represents local supervision by additive Gram, cross-moment, and count statistics. Any horizontal split of the same data therefore yields the same pooled classification and localization heads, independent of client count, label skew, or sample assignment. Each client uploads once; the server applies global balancing and solves the readouts in closed form without transferring raw images, RoI features, gradients, or locally trained detector weights.

The framework supports two inference modes from the same compilation law. The *independent analytic mode* solves classification and localization without frozen prediction logits or box deltas; the *prior-calibrated region mode* combines the recovered classifier with a public detector's scoring and decoding prior. To our knowledge, the analytically solved-head framework is the first one-upload

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.

analytic-federated construction that exactly reconstructs pooled classification and box-regression readouts for a 2D detector under a fixed proposal mechanism.

Our contributions are:

- A one-upload detection compiler for variable proposal sets, globally deferred class balancing, and separate analytic classification and localization heads.
- A proof of exact pooled-objective recovery for every horizontal partition, separated from finite-precision summation and wire quantization.
- VOC evidence across two releases, 15 heterogeneous partitions, and three feature seeds, with sufficient statistics $52.88\times$ smaller than the corresponding raw images.
- A prior-calibrated mode that reaches 89.75 AP50 on VOC2007 test and 86.92 on VOC2012 validation from the same uploaded statistics.
- A functionally specialized three-branch readout: Conv3 and Conv4 attention memories contribute complementary intermediate evidence, the FC ridge branch performs final class discrimination, and the class-specific box decoder performs localization. Controlled depth, DINOv2, FCOS, and routing studies establish that these roles are complementary rather than interchangeable.

2 Related Work

Iterative federated optimization. FedAvg and FedProx exchange parameters over multiple rounds (McMahan et al., 2017; Li et al., 2020). Federated detection systems and broader benchmarks establish the task's relevance (Liu et al., 2020; He et al., 2021; Kim et al., 2023; Lai et al., 2022; Zhuang et al., 2024), but retain iterative training rather than recovering a pooled detector readout in one upload. FedSTO, for example, addresses semi-supervised domain shift through selective and full-parameter training on unlabeled clients (Kim et al., 2023); its supervision and datasets differ from our labeled cross-silo setting, so we compare protocol communication structurally rather than treating its reported AP as a matched baseline.

One-shot and analytic federated learning. One-shot methods aggregate local models, distill knowledge, or transfer synthetic and statistical surrogates (Li et al., 2021; Beitollahi et al., 2025; Zeng et al., 2025; How Shi Qi et al., 2025; Qian et al., 2026). Fed3R and federated ridge establish additive normal-equation recovery for linear heads (Fanì et al., 2024; Alsulaimawi, 2026); FedCGS and GH-OFL compile feature moments (Guan et al., 2025; Turazza et al., 2026); AFL, DeepAFL, and SAFLe develop exact single-round analytic aggregation and more expressive readouts (He et al., 2025; Tang et al., 2026; Bacellar et al., 2026). Their evaluations focus mainly on image classification, with one-shot structured vision also studied in segmentation (Lian et al., 2025). The analytically solved-head framework contributes the detection-specific construction: variable proposals, deferred balancing, separate classification and localization heads, and two server inference modes from the same statistics.

Frozen-region adaptation. Region detectors combine shared representations with task-specific classification and localization heads (Girshick et al., 2014; Ren et al., 2015); frozen features and calibration are also established (Qiao et al., 2021; Kuo et al., 2023; Pham et al., 2024). Random-feature methods freeze nonlinear transformations and solve linear outputs (Rahimi & Recht, 2007; Huang et al., 2006). Our contribution is their integration into a partition-invariant, one-shot, multi-function detector readout with independent analytic and prior-calibrated deployment.

3 Problem Setting

Let $m \geq 1$ clients hold a horizontal partition $D = \bigsqcup_{s=1}^{m} D_s$ of labeled images. A shared configuration identifies a frozen detector representation g_θ, proposal limit, RoI schema, feature-map identifier and, when applicable, random seed, class vocabulary, regression encoding, and solver settings. The parameter θ is fixed and common: the analytically solved-head framework adapts only the readout heads.

Table 1: The two operating modes of the analytically solved heads share compilation and aggregation but use different inference information.

Mode	Analytically fitted functions	Frozen outputs used at inference
Independent analytic	Classification and box regression	Proposals and RoI representation only
Prior-calibrated region	Classification correction; analytic box head retained for comparison	Class probabilities and class-specific box deltas

For image x, g_θ produces proposals b_i and RoI embeddings $z_i \in \mathbb{R}^d$. Proposal matching assigns a class $y_i \in \{1, \dots, C, \mathrm{bg}\}$ and, for positive proposals, a box transform $t_i \in \mathbb{R}^4$. We require:

1. **one upload:** no pre-round, iterative gradient exchange, or client-to-client coordination;
2. **pooled equivalence:** aggregation yields the head obtained from the union of all proposals;
3. **partition invariance:** the recovered head depends on the data union, not its client assignment;
4. **bounded statistic size:** transmitted statistic dimensions depend on task configuration, not the number of images or proposals; and
5. **useful detection:** both classification and localization are evaluated on sealed data.

The setting is cross-silo and synchronous only at delivery time. Availability, privacy, and adversarial verification are orthogonal to the present claim.

4 Analytically Solved Detection Heads

4.1 One compiler, two operating modes

Both modes begin with a frozen proposal and region representation and end with additive sufficient statistics. They do not require different aggregation protocols. The independent analytic mode solves and deploys both W^{cls} and W^{box}. It uses a detector-derived representation but does not use the frozen detector's output logits or box deltas at inference. The prior-calibrated region mode deploys the recovered classification correction together with the frozen detector's prediction outputs. It therefore depends on a detection prior, while the independent mode depends only on the shared representation and proposal mechanism. Table 1 fixes this distinction.

4.2 Shared proposal representation and random-affine map

Let $\psi(x, b) \in \mathbb{R}^d$ be any deterministic representation of proposal b in image x, and let $\phi : \mathbb{R}^d \to \mathbb{R}^q$ be fixed independently of client data. Both are identified by the shared configuration, so all clients compile identical functions and only their proposal rows differ.

Our compact VOC instantiation uses a COCO-pretrained Faster R-CNN with a ResNet-50-FPN-v2 backbone (Ren et al., 2015; He et al., 2016; Lin et al., 2017). The RPN and RoI box head remain frozen. We keep at most 300 RPN proposals per image and obtain $z = \psi(x, b) \in \mathbb{R}^{1024}$. For a width r, a public seed generates independent Gaussian matrices $R_1, R_2 \in \mathbb{R}^{1024 \times r}$ with entries scaled by $1024^{-1/2}$:

$$\phi_{\mathrm{dual}}(z) = [zR_1 \mid \mathrm{softplus}(zR_2) \mid 1] \in \mathbb{R}^{2r+1}. \tag{1}$$

4.3 Class-conditional and localization statistics

Let $\Phi_{s,c}$ contain the design rows of client s assigned to class c, including background. The client compiles

$$P_{s,c} = \Phi_{s,c}^{\mathsf{T}} \Phi_{s,c}, \qquad \ell_{s,c} = \Phi_{s,c}^{\mathsf{T}} \mathbf{1}, \qquad n_{s,c} = |\Phi_{s,c}|. \tag{2}$$

Let $\Phi_{s,+}$ and $T_{s,+}$ denote positive-proposal designs and encoded box targets. The localization statistics are

$$P_s^{\mathrm{box}} = \Phi_{s,+}^{\mathsf{T}} \Phi_{s,+}, \qquad L_s^{\mathrm{box}} = \Phi_{s,+}^{\mathsf{T}} T_{s,+}. \tag{3}$$

The server first sums all statistics and counts. For every globally present class, it computes the inverse-frequency weight

$$a_c = \frac{\sum_j n_j}{|\{j : n_j > 0\}|\, n_c}, \qquad n_c = \sum_s n_{s,c}. \tag{4}$$

It then forms

$$\begin{aligned} P^{\mathrm{cls}} &= \sum_c a_c \sum_s P_{s,c}, & L^{\mathrm{cls}}_{:,c} &= a_c \sum_s \ell_{s,c}, \\ W^{\mathrm{cls}} &= (P^{\mathrm{cls}} + \lambda I)^{-1} L^{\mathrm{cls}}, & W^{\mathrm{box}} &= \left(\sum_s P_s^{\mathrm{box}} + \lambda I \right)^{-1} \sum_s L_s^{\mathrm{box}}. \end{aligned} \tag{5}$$

The weight in Equation 4 gives every globally present class equal total mass while preserving the mean weight of one. It is computed from global counts because local inverse frequencies define a different objective for each partition. Other count-based weighting rules remain partition invariant whenever the weight is a deterministic function of the aggregated counts; we use inverse frequency to expose the detection imbalance directly without another client field or coordination round.

Proposition 1 (Exact pooled recovery)**.** *If all clients use the identified configuration and the server applies Equation 4 after summation, Equations 2–5 equal the class-balanced ridge classification and box-regression systems formed from the union of all proposal rows. The recovered heads are therefore invariant to every horizontal partition in exact arithmetic.*

Proof. Both Gram and cross-moment terms are sums over proposal rows. Summing client statistics reconstructs the corresponding pooled terms; the pooled counts then determine the same a_c as centralized compilation. Substitution into Equation 5 gives identical normal equations. □

4.4 Sufficient statistics and communication

Each symmetric Gram matrix transmits only its upper triangle. With $K = C + 1$ classes including background, the transmitted statistics contain

$$N_{\mathrm{float}} = (K+1)\frac{q(q+1)}{2} + q(K+4), \qquad N_{\mathrm{count}} = K + 1. \tag{6}$$

Here the K class counts plus one positive-proposal count give $N_{\mathrm{count}} = K + 1$. Communication and server statistic storage are $O(Kq^2)$, while forming the systems costs $O(Kq^2)$ and dense factorization costs $O(q^3)$, followed by $O(q^2(K+4))$ solves. These costs are independent of local image and proposal counts but grow linearly with the class vocabulary and quadratically or cubically with feature width. For the selected $K = 21, q = 193$ configuration, each client sends 1.67 MB in float32 and the server factors only 193×193 systems; Figure 2 measures the resulting width–accuracy trade-off.

4.5 Server-side prior-calibrated region mode

The same compiled statistics also support a prior-calibrated mode. Let $p_c^{(0)}(z)$ and $\delta_c^{(0)}(z)$ be the frozen detector's probability and class-specific box transform. For analytic logits $u(z) = \phi(z)W^{\mathrm{cls}}$, define

$$s_c(z) = u_c(z) - u_{\mathrm{bg}}(z) + \alpha p_c^{(0)}(z), \qquad \widehat{c} = \arg\max_{c \le C} s_c(z), \qquad \widehat{b} = \mathrm{Decode}\Big(b, \delta_{\widehat{c}}^{(0)}(z)\Big). \tag{7}$$

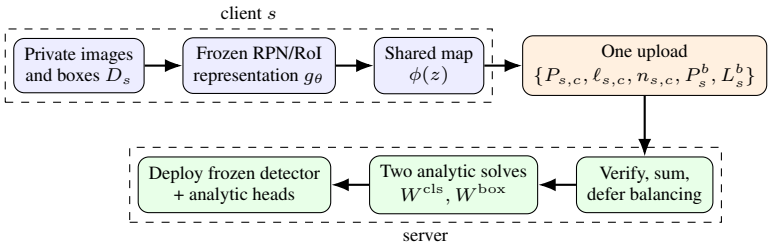

Figure 1: Execution of the analytically solved-head framework. Each client performs deterministic feature extraction locally and uploads Gram, cross-moment, and count statistics once. No client receives an intermediate global state.

After thresholding, the server applies deterministic NMS. The analytic margin $u_c - u_{\mathrm{bg}}$ removes a shared score offset, while the bounded probability $p_c^{(0)}$ retains the detector's calibrated ranking without fitting another model. The scalar α is selected once on VOC2007 train/validation and then frozen for VOC2007 test, VOC2012 validation, and the untuned MobileNet transfer. Because $p^{(0)}$ and $\delta^{(0)}$ come from the identified detector and u from the original class statistics, this mode adds no client field or upload cost; matched controls isolate its contribution in Section 6.

4.6 Additive linear-attention score memories

The controlled extension exposes Conv3/Conv4/FC readouts as a functionally specialized three-branch architecture. At each tap, a public positive random map lets clients compile class-conditional $K^\top V$ and $K^\top \mathbf{1}$ sums; the server applies globally deferred balancing and normalized linear-attention scoring without a normal-matrix inverse. A decoder branch D remains the class-balanced ridge score from Equation 5. Routing strings list Conv3/Conv4/FC in order. The branches form one coordinated decision rather than three independent detectors: Conv3 captures shallow region evidence, Conv4 contributes complementary mid-level evidence, and the final FC ridge branch anchors semantic class discrimination. Thus AAD combines the two attention memories with the FC ridge score before a single class decision. A separate class-specific decoder supplies geometric localization, completing the division of labor between scoring and box prediction. The attention memory is additive and partition invariant under the same shared-map assumption; Appendix I gives the exact compilation rule.

Figure 1 emphasizes the execution boundary. Raw images, RoI embeddings, and gradients remain with clients. The server receives only Gram, cross-moment, count, and optional attention-sum statistics, reconstructs the analytic heads, and combines them with the already identified frozen detector representation.

5 Experimental Protocol

Because of its canonical detection task, compact label space, and stable public protocol, we choose PASCAL VOC (Everingham et al., 2010) to test whether partition-invariant, one-shot AFL is viable before moving to larger detection settings. The paired releases also support a stronger check than a single benchmark split: VOC2007 train/val selects the configuration, VOC2007 trainval compiles confirmatory statistics, and VOC2007 test plus the disjoint later VOC2012 validation release evaluate the frozen setting without retuning. The validation grid covers $r \in \{16, 32, 64, 96, 128, 192\}$, $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$, and thresholds $\{0.02, 0.05, 0.1, 0.2\}$; the smallest float32 upload within 95% of the best validation AP50 gives $r = 96$, $q = 193$, $\lambda = 0.01$, and threshold 0.05. The prior-calibrated mode uses the same statistics and selects $\alpha = 4$ and threshold 0.02 on train/val. We evaluate IID, dominant-label, Dirichlet $\alpha \in \{1, 0.1, 0.01\}$, and $m \in \{2, 5, 10, 20\}$ partitions; pooled recovery, local-head averages, the frozen detector, matched prior controls, feature-map and head ablations, six readout depths, A/D routing, three DINOv2 encoders, and frozen FCOS (Tian et al., 2019). Metrics are VOC AP50, AP75, AP@[.50:.95], and VOC2007 11-point AP50; the averaged metric uses VOC rather than COCO conventions. Float64/float32 audits measure statistic, head, prediction, and AP differences. Confirmatory results average feature seeds 42–44; the additional

Table 2: Frozen confirmatory results (mean ± population standard deviation, three seeds).

Dataset	AP50	AP75	AP@[.50:.95]	$AP50_{07}$
VOC07 test	78.48 ± 0.51	55.62 ± 0.89	49.76 ± 0.62	75.57 ± 0.42
VOC12 val	75.32 ± 0.24	53.71 ± 0.98	48.12 ± 0.57	72.84 ± 0.21

Table 3: Published VOC2007 test reference points. AP50 values follow each cited paper; training data are shown because these conventional detectors use different supervision and optimization settings from our one-upload analytic setting.

Method	Training data	VOC07 AP50
Faster R-CNN (Ren et al., 2015)	VOC07+12 trainval	73.2
R-FCN, multi-scale (Dai et al., 2016)	VOC07+12+COCO	83.6
Analytic two-head (ours)	VOC07 trainval	78.48 ± 0.51
Prior-calibrated pooled (ours)	VOC07 trainval + frozen COCO prior	89.75 ± 0.07

studies use seed 42 except for the stated second attention seed. Runs use Python 3.12.3, NumPy 1.26.4, and an NVIDIA RTX A6000. Communication is measured from serialized statistics against the corresponding trainval JPEG bytes. The appendix provides the selection criteria, partition definitions, accounting, and implementation details.

6 Results

6.1 Confirmatory detection quality

Table 2 reports means and population standard deviations across the three frozen random-feature seeds. The analytically solved-head framework remains stable across both sealed sets. The VOC2012 decrease is 3.16 AP50 points relative to VOC2007 test, while seed variation narrows to 0.24 points.

Table 3 places the reported VOC2007 performance alongside a classical region-based baseline and a strong published detector. The analytic two-head result exceeds the Faster R-CNN reference while using a single sufficient-statistic upload, and prior-calibrated deployment reaches 89.75 AP50 by combining the partition-invariant analytic correction with the shared frozen detector prior. The R-FCN result uses additional VOC2012 and COCO training data and is included as a high-performance literature reference rather than a matched federated baseline.

The result highlights the distinct operating point of the analytically solved-head framework. The unadapted COCO detector obtains 89.50 AP50 on VOC07 and 86.77 on VOC12 because its pretraining vocabulary already contains all VOC categories, whereas the analytically solved-head framework evaluates whether a newly compiled two-head detector can be recovered exactly from one upload. The comparison therefore isolates analytic recovery under a strict communication constraint rather than competing with a detector whose prediction prior is already available.

6.2 Prior-calibrated region mode

Prior calibration recovers 11.27/11.60 AP50 points over independent heads without changing uploaded bytes, but most accuracy comes from the frozen detector. Relative to matched prior-only controls, gains are 0.45 points on VOC07 (paired bootstrap 95% CI [0.20, 0.73]) and 0.50 on VOC12 ([0.30, 0.72]); Mobile gains 0.67. VOC07 AP@[.50:.95] is 64.70 ± 0.27, below prior-only 65.45, and VOC12 is 63.78 ± 0.30, below 64.50. We therefore claim AP50 calibration, not metric-wide dominance.

Untuned architecture transfer. The preregistered MobileNetV3-Large-FPN transfer reuses every compiler and solver choice without tuning. The analytic two-head mode obtains 58.35 AP50. The prior-calibrated rule applied to the same sufficient statistics yields 83.92 AP50, 60.29 AP75, and 55.04 AP@[.50:.95]; prior-only obtains 83.25 and the frozen detector output 83.58 AP50. Pooled recovery remains exact, and float32 changes prior-calibrated AP50 by 3.89×10^{-9}.

Table 4: AP50 comparison from identical compiled sufficient statistics. ResNet entries use mean $\pm$ population standard deviation over seeds 42–44; MobileNet uses frozen seed 42.

Mode	VOC07 test	VOC12 val	Mobile VOC07
Analytic two-head	78.48 ± 0.51	75.32 ± 0.24	58.35
Detector prior only	89.30	86.42	83.25
Prior-calibrated pooled	89.75 ± 0.07	86.92 ± 0.07	83.92
Frozen detector output	89.50	86.77	83.58

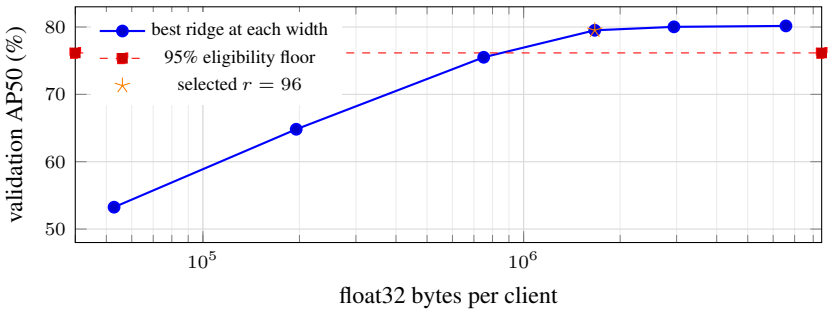

Figure 2: Measured communication–accuracy frontier on VOC2007 validation. The selection rule was fixed before confirmatory evaluation.

6.3 Pooled recovery under heterogeneity

Prior-calibrated region mode repeats three representative partitions: five dominant-label clients, ten Dirichlet clients at $\alpha = 0.01$, and twenty Dirichlet clients at $\alpha = 0.1$. All exactness checks pass, and pooled validation AP50 is exactly 0.9012975366 in every run (computed standard deviation zero). Weighted and uniform local averages vary across assignments. Hence Equation 7 preserves pooled partition invariance because its prior term is server-fixed and its only client-dependent input is the exactly aggregated analytic head.

Local head averaging is not pooled recovery. On VOC12, the pooled head obtains 75.32 AP50, versus 76.50 for sample-weighted and 76.85 for uniform local averaging. The sign is informative: heuristic averaging can regularize favorably on a particular metric even though it solves no declared centralized objective. We consequently claim invariance and predictability, not universal accuracy dominance.

6.4 Communication–accuracy frontier

The validation optimum is 80.16 AP50. Figure 2 shows that $r = 64$ misses the 95% eligibility floor (76.15 AP50), whereas $r = 96$ reaches 79.51 AP50 and is the smallest qualifying statistic set. Increasing to $r = 192$ improves AP50 by only 0.65 points while increasing upload from 1.67 to 6.58 MB per client.

For five clients, the selected float32 sufficient statistics total 8,334,620 bytes. The corresponding 5,011 VOC2007 trainval JPEG files total 440,719,007 bytes, so the transmitted statistics are $52.88\times$ smaller (98.109%) than the corresponding raw images. For an iterative protocol that uploads a B_{model}-byte detector update for R rounds, client upload is RB_{model} before server broadcasts; our measured upload is 1,666,924 bytes once. Thus our upload is smaller whenever $RB_{\text{model}} > 1{,}666{,}924$ bytes, while eliminating round synchronization. This accounting is protocol-level rather than an AP comparison: FedSTO and related systems optimize different supervision and adaptation objectives. Float64 statistics retain a $26.44\times$ raw-image reduction. Across confirmatory seeds, the largest relative float32 statistic difference is below 3.00×10^{-8} and mean VOC12 AP50 changes by $+1.59 \times 10^{-7}$. Prior calibration uses the same statistics, adds zero measured upload, and changes float32 VOC12 AP50 by -3.65×10^{-9}.

6.5 WHAT MAKES ANALYTIC DETECTION WORK?

Table 5: Matched validation ablations at the frozen threshold.

Method	AP50	AP75	AP@[.50:.95]
Dual map, full heads	79.51	57.45	50.52
Linear only, matched q	78.75	56.70	50.27
Softplus only, matched q	77.69	55.14	49.07
No analytic box regression	59.82	11.64	24.64
No deferred class balancing	38.08	34.43	28.14

Table 5 identifies the detection-specific requirements. Removing box regression costs 19.69 AP50 and 45.81 AP75 points; omitting balancing costs 41.43 AP50 points. By contrast, the dual random map improves only 0.76 AP50 points over a matched-size linear map. The central advance is therefore the one-upload multi-head compilation and deferred balancing, not a claim that the chosen random feature map is uniquely necessary.

The prior-calibrated ablation is complementary. At the selected validation operating point, analytic reconstruction gives 79.51 AP50; substituting frozen class-specific boxes gives 80.48; prior-only gives 89.52; fused scoring with analytic boxes gives 89.79; and full Equation 7 gives 90.13. Thus frozen localization contributes most of the high-IoU recovery, while the analytic score contributes a smaller AP50 calibration gain. Selection on AP50 explains why full fusion (65.07 AP@[.50:.95]) is below prior-only (65.76) on the stricter aggregate.

A full-split sensitivity study (Appendix F) varies α from 0 to 8. AP50 rises from 83.11 at $\alpha = 0$ to 87.33 at $\alpha = 4$, then remains within 0.13 points through $\alpha = 8$, showing a broad operating plateau.

6.6 CONTROLLED READOUT-DEPTH AND ROUTING STUDY

This VOC2007 train/validation study uses the same Faster R-CNN ResNet-50-FPN-v2 proposal mechanism, 300 proposals, $q = 193$, $\lambda = 0.01$, and seed 42. Its single-FC reference is 74.94 AP50, and all comparisons below use this implementation.

The controlled depth sweep (Appendix Figure 3) resolves the layer-selection question in favor of the final FC tap. AP50 rises from 19.95 at RoIAlign to 74.94 at FC, while AP75 rises from 2.35 to 53.42. The FC result is unchanged to the reported precision for $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$: AP50 remains in $[74.9424, 74.9435]$, so the ordering is not an artifact of the ridge coefficient.

Table 6: Linear-attention (A) and ridge-decoder (D) routing across Conv3/Conv4/FC. Frozen-box rows use the shared class-specific decoder.

Routing	Box source	AP50	AP75	AP@[.50:.95]
DDD	analytic D branches	70.63	46.67	43.47
ADD	analytic D branches	73.12	49.75	45.25
AAD	analytic FC D branch	74.97	53.38	47.38
DDD	frozen decoder	72.29	54.16	49.76
ADD	frozen decoder	74.46	57.04	51.71
AAD	frozen decoder	**76.09**	**58.32**	**53.27**
AAA	frozen decoder	71.67	55.48	50.05

Table 6 shows a consistent division of labor. The best AAD/frozen-decoder route improves over the single-FC result by 1.15 AP50, 4.90 AP75, and 5.92 AP@[.50:.95], and a second attention seed gives 76.08/58.31/53.24. Conv3 and Conv4 attention memories supply complementary intermediate evidence, the FC ridge branch anchors class discrimination, and the frozen class-specific decoder localizes boxes. Replacing the FC branch (AAA) or predicting boxes from attention memory (54.93/12.14/22.76 in a 500-image screen) weakens this specialization. The coordinated branches are therefore unequal and complementary, not independent predictors. Appendix J reports the DINOv2 and FCOS results.

7 Discussion and Limitations

Partition-invariant one-shot recovery and specialized readouts. Statistic dimensions are fixed by the task configuration rather than local sample count, and one aggregation recovers the declared pooled ridge or additive-memory objective under identical configurations. The depth and routing studies reveal a functional hierarchy: Conv3/Conv4 memories contribute intermediate evidence, the FC ridge branch anchors class discrimination, and the box decoder handles localization. Their unequal roles explain why neither all-attention scoring nor attention-based localization matches the coordinated route.

Scaling regime and deployment boundary. The method is most favorable for cross-silo clients with many local samples: payload is independent of local data volume, but total traffic grows as mKq^2 and dense server factorization as q^3. Appendix F measures this boundary through $q = 769$ and an 81-label vocabulary: the selected $K = 21, q = 193$ system solves in 22.2 ms in the local CPU benchmark, while per-client upload grows from 1.67 MB to 97.37 MB at $K = 81, q = 769$. Larger vocabularies increase payload linearly, and wider readouts can use block or sketched statistics. Secure aggregation is directly compatible with the additive fields; differential-privacy noise instead trades exact recovery for protection (Alsulaimawi, 2026).

Evaluation breadth. VOC provides the controlled paired-release test targeted by this study, but its classes overlap the frozen detectors' COCO vocabulary. MobileNetV3-FPN transfer shows that standalone recovery depends on detector representation, while reuse of the frozen α shows that prior calibration transfers without retuning. DINOv2 patch pooling localizes the open interface question: proposal-aware token aggregation is the next test, whereas the FCOS control shows that the evaluator and anchor-free detection itself are not the bottleneck. COCO-scale vocabularies and matched iterative-FL accuracy comparisons remain outside the present evidence.

8 Conclusion

The analytically solved-head framework demonstrates that analytic federated learning can extend beyond fixed-output prediction to object detection. Frozen proposal representations, class-conditional compilation, deferred balancing, and separate analytic heads let a server reconstruct a useful pooled detector from one upload of sufficient statistics per client. A configuration selected on VOC2007 train/validation remains effective on both VOC2007 test and the disjoint VOC2012 validation release, providing evidence of cross-release generalization without retuning. The same statistics support prior-calibrated deployment, while the controlled follow-up identifies a concrete readout design: late RoI features carry the strongest standalone signal, additive attention supplies complementary shallow/mid-level evidence, and precise boxes are recovered through a ridge or frozen detector decoder. The resulting framework combines exact pooled recovery, a clear communication–accuracy trade-off, and a principled division of labor across detector functions.

Ethics Statement

The experiments use public benchmark images and do not involve human-subject interaction. The method reduces raw-data movement but does not, by itself, guarantee privacy: Gram and cross-moment statistics can reveal aggregate information. Secure aggregation is algebraically compatible because recovery uses sums, whereas differential-privacy noise changes the exact pooled objective. No membership-inference or reconstruction-resistance claim is made. Deployment would require an explicit threat model, access control, integrity checks, and an independently evaluated protection layer.

Reproducibility Statement

Sections 4–6 specify compilation, recovery, evaluation, and model selection. The appendix provides client and server procedures, per-seed results, partition definitions, communication accounting, and architecture studies. We will release the implementation, configurations, and execution records upon acceptance.

AI USE STATEMENT

Large language models assisted with literature triage, manuscript drafting, code scaffolding, and consistency checks. The authors selected the research question and claims and verified all reported values against execution records. No result was inferred by a language model.

REFERENCES

Zahir Alsulaimawi. One-shot federated ridge regression: Exact recovery via sufficient statistic aggregation. *arXiv preprint arXiv:2601.08216*, 2026.

Alan T. L. Bacellar, Mustafa Munir, Felipe M. G. França, Priscila M. V. Lima, Radu Marculescu, and Lizy K. John. Single-round scalable analytic federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Mahdi Beitollahi, Alex Bie, Sobhan Hemati, Leo Maxime Brunswic, Xu Li, Xi Chen, and Guojun Zhang. Parametric feature transfer: One-shot federated learning with foundation models. *Transactions on Machine Learning Research*, 2025.

Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88 (2):303–338, 2010.

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*, volume 235 of *Proceedings of Machine Learning Research*, pp. 13029–13048, 2024.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

Zenghao Guan, Yucan Zhou, and Xiaoyan Gu. Capture global feature statistics for one-shot federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 16942–16950, 2025.

Chaoyang He, Alay Dilipbhai Shah, Zhenheng Tang, Di Fan, Adarshan Naiynar Sivashunmugam, Keerti Bhogaraju, Mita Shimpi, Li Shen, Xiaowen Chu, Mahdi Soltanolkotabi, and Salman Avestimehr. FedCV: A federated learning framework for diverse computer vision tasks. *arXiv preprint arXiv:2111.11066*, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Run He, Kai Tong, Di Fang, Han Sun, Ziqian Zeng, Haoran Li, Tianyi Chen, and Huiping Zhuang. AFL: A single-round analytic approach for federated learning with pre-trained models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4988–4998, 2025.

Shannon How Shi Qi, Jagmohan Chauhan, Geoff V. Merrett, and Jonathon Hare. FedTMOS: Efficient one-shot federated learning with tsetlin machine. In *International Conference on Learning Representations*, 2025.

Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1314–1324, 2019.

Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3):489–501, 2006.

Taehyeon Kim, Eric Lin, Junu Lee, Christian Lau, and Vaikkunth Mugunthan. Navigating data heterogeneity in federated learning: A semi-supervised federated object detection. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Weicheng Kuo, Yin Cui, Xiuye Gu, A. J. Piergiovanni, and Anelia Angelova. F-VLM: Open-vocabulary object detection upon frozen vision and language models. In *International Conference on Learning Representations*, 2023.

Fan Lai, Yinwei Dai, Sanjay Singapuram, Jiachen Liu, Xiangfeng Zhu, Harsha Madhyastha, and Mosharaf Chowdhury. FedScale: Benchmarking model and system performance of federated learning at scale. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 11814–11827, 2022.

Qinbin Li, Bingsheng He, and Dawn Song. Practical one-shot federated learning for cross-silo setting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 1484–1490, 2021.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems*, volume 2, pp. 429–450, 2020.

Tao Lian, Jose L. Gómez, and Antonio M. López. FedS2R: One-shot federated domain generalization for synthetic-to-real semantic segmentation in autonomous driving. *arXiv preprint arXiv:2507.19881*, 2025.

Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, 2017.

Yang Liu, Anbu Huang, Yun Luo, He Huang, Youzhi Liu, Yuanyuan Chen, Lican Feng, Tianjian Chen, Han Yu, and Qiang Yang. FedVision: An online visual object detection platform powered by federated learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(8): 13172–13179, 2020.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282, 2017.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Chau Pham, Truong Vu, and Khoi Nguyen. LP-OVOD: Open-vocabulary object detection by linear probing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 779–788, 2024.

Xuwei Qian, Jinghui Zhang, Yuchuan Tan, Wenbo Huang, Zhen Wu, Shen Zhou, LiSha Gao, Ding Ding, and Fang Dong. OS-Fed: One snapshot is all you need. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 31758–31768, 2026.

Limeng Qiao, Yuxuan Zhao, Zhiyuan Li, Xi Qiu, Jianan Wu, and Chi Zhang. DeFRCN: Decoupled faster R-CNN for few-shot object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8681–8690, 2021.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Jianheng Tang, Yajiang Huang, Kejia Fan, Feijiang Han, Jiaxu Li, Jinfeng Xu, Run He, Anfeng Liu, Houbing Song, Huiping Zhuang, and Yunhuai Liu. DeepAFL: Deep analytic federated learning. In *International Conference on Learning Representations*, 2026.

Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. FCOS: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9627–9636, 2019.

Fabio Turazza, Marco Picone, and Marco Mamei. The gaussian-head OFL family: One-shot federated learning from client global statistics. In *International Conference on Learning Representations*, 2026.

Hui Zeng, Wenke Huang, Tongqing Zhou, Xinyi Wu, Guancheng Wan, Yingwen Chen, and Zhiping Cai. Does one-shot give the best shot? mitigating model inconsistency in one-shot federated learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 74080–74097, 2025.

Weiming Zhuang, Jian Xu, Chen Chen, Jingtao Li, and Lingjuan Lyu. COALA: A practical and vision-centric federated learning platform. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 62723–62742, 2024.

A Complete Protocol

A.1 Client compilation

Algorithm 1: ClientCompile(D_s, cfg)
Require: shared detector, proposal schema, vocabulary, feature map, and target encoding.
1. Initialize class-conditional $P_{s,c} \leftarrow 0$, $\ell_{s,c} \leftarrow 0$, $n_{s,c} \leftarrow 0$.
2. Initialize $P_s^{\mathrm{box}} \leftarrow 0$, $L_s^{\mathrm{box}} \leftarrow 0$.
3. For each local image, run the frozen detector, retain the configured proposal cap, match proposals at IoU ≥ 0.5, and compute Equation 1.
4. Accumulate Equation 2 for every proposal class.
5. Accumulate Equation 3 for positive proposals.
6. Pack upper Gram triangles, full moment matrices, counts, and configuration identifier.
7. Upload exactly once.

Algorithm 2: Recover$(\{\mathsf{stats}_s\}_{s=1}^m, \lambda)$
1. Check configuration, shape, vocabulary, dtype, and finite values.
2. Reconstruct symmetric Grams and sum all fields.
3. Compute global weights a_c from summed counts using Equation 4.
4. Form the class and box systems in Equation 5.
5. Solve in float64 and attach the resulting heads to the identified frozen detector.
6. Record the solver configuration and output heads.

For the compact VOC configuration, logits $\phi(z)W^{\mathrm{cls}}$ are normalized by softmax. The highest-scoring non-background class is retained when its score exceeds 0.05. The decoded class-agnostic box correction $\phi(z)W^{\mathrm{box}}$ is applied to the RPN proposal; deterministic per-class NMS uses IoU 0.5 and retains at most 100 detections. Prior-calibrated region mode replaces this final scoring and decoding step with Equation 7, $\alpha = 4$, and threshold 0.02; client compilation and server recovery are unchanged.

Table 7: Selection criteria fixed before confirmatory evaluation.

Item	Criterion
Selection	Smallest float32 sufficient-statistic upload with validation AP50 ≥ 0.95 of the best grid result.
Confirmatory quality	Mean AP50 ≥ 0.70 on both sealed evaluations.
Seed stability	AP50 population standard deviation ≤ 0.01.
Local baseline	Pooled AP50 no more than 0.02 below sample-weighted local head averaging.
Wire precision	float32 AP50 loss ≤ 0.002 versus float64.
Exactness	Every float64 partition reconstruction passes registered tolerances.
Invalid run	Split leakage, incomplete union, empty client, configuration mismatch, or nonfinite solve.

B Preregistered Selection Criteria

The criteria are satisfied. The prior-calibrated mode gains 0.11269/0.11602 AP50 over analytic reconstruction, with maximum seed standard deviation 0.000708, VOC12 float32 change -3.65×10^{-9}, exact pooled recovery, and no added upload.

C Per-Seed Confirmatory Results

Table 8: Pooled float64 results for every frozen seed.

Seed	Dataset	AP50	AP75	AP@[.50:.95]	$AP50_{07}$
42	VOC07 test	79.16	56.87	50.61	76.11
43	VOC07 test	78.36	54.93	49.19	75.54
44	VOC07 test	77.92	55.05	49.46	75.07
42	VOC12 val	75.65	55.10	48.91	73.12
43	VOC12 val	75.11	52.91	47.63	72.61
44	VOC12 val	75.19	53.13	47.81	72.78

The largest confirmatory float64 statistic residual is 7.28×10^{-12}; the largest box-head residual is 7.22×10^{-15}; class-head residuals are zero. The largest float32 statistic absolute difference is 0.04198, but its largest relative difference is below 3.00×10^{-8}.

D Non-IID Matrix

The 15-run matrix contains two controls (IID hashing and dominant-label assignment), nine ten-client Dirichlet partitions ($\alpha \in \{1, 0.1, 0.01\}$, seeds 701–703), and four client-count runs ($m \in \{2, 5, 10, 20\}$ at $\alpha = 0.1$, seed 701). The pooled head is constant because the union is constant. Local weighted AP50 ranges approximately from 78.78 to 81.43 and uniform local averaging from 77.93 to 81.86, illustrating that averaging local optima is partition dependent even when it sometimes scores higher.

The prior-calibrated extension freezes three representative runs: dominant-label $m = 5$, Dirichlet $\alpha = 0.01, m = 10$, and Dirichlet $\alpha = 0.1, m = 20$, all at partition seed 701 and random-feature seed 42. Pooled AP50 is 0.9012975366 in all three; exactness passes in all three. Total float32 upload is 8,334,620, 16,669,240, and 33,338,480 bytes, respectively, confirming fixed per-client cost and linear total cost.

E Communication Accounting

For $K = 21$ proposal labels and $q = 193$, Equation 6 gives 416,687 transmitted floating values after symmetric packing and 22 int64 counts. This is 1,666,924 bytes in float32 and 3,333,672 bytes in float64 per client. Five clients transmit 8,334,620 float32 bytes in total. The corresponding raw-image

reference is 440,719,007 bytes, the exact size of the 5,011 VOC2007 trainval JPEG files, making the float32 statistics $52.88\times$ smaller than the raw images. This raw-image reference quantifies payload scale rather than serving as a federated-learning baseline; iterative FL ordinarily exchanges gradients or model parameters. Annotation bytes and distribution of the already shared detector are excluded.

Total upload scales linearly with client count because each client emits fixed-dimensional statistics; per-client upload does not scale with local sample count. This regime is attractive when each client owns many samples. It can be unattractive for many tiny clients, where mq^2 exceeds transmitting the samples or a smaller model update.

F Sensitivity and Scaling

The following studies use the complete VOC2007 train/validation split (2,501/2,510 images), 300 proposals per image, the FC readout, $q = 193$, and seed 42.

Class-weighting rule. Every rule in Table 9 is computed after aggregation from global counts, preserving exact partition invariance without an additional client field. Effective-number weighting gives the strongest AP50 in this study, showing that the one-upload compiler supports alternative global balancing rules.

Table 9: Global class-weighting sensitivity on VOC2007 validation.

Global rule	AP50	AP75	AP@[.50:.95]
Uniform	80.89	61.78	53.59
Square-root inverse frequency	81.43	60.50	52.98
Effective number ($\beta = 0.9999$)	**83.16**	**62.02**	**54.30**
Inverse frequency (frozen rule)	74.94	53.42	47.35

Prior strength. Table 10 evaluates Equation 7 with class-specific boxes and threshold 0.02. The selected $\alpha = 4$ gains 4.23 AP50 over $\alpha = 0$, and values from 4 to 8 form a broad plateau.

Table 10: Prior-strength sensitivity on VOC2007 validation.

α	0	1	2	4	6	8
AP50	83.11	86.59	87.07	87.33	87.40	**87.46**
AP75	65.58	69.83	70.23	70.88	71.05	**71.24**
AP@[.50:.95]	58.85	62.31	62.82	63.34	63.54	**63.65**

Server scaling. Table 11 uses the exact packed-upload formula in Equation 6 and dense positive-definite systems with five repeated NumPy solves. The timing is the median local CPU wall time and includes factorization and all $K + 4$ right-hand sides. Increasing the vocabulary from 21 to 81 primarily affects communication and stored class-conditional Grams; the dense factorization remains governed by q. The selected $q = 193$ operating point is small, whereas the $q = 769, K = 81$ row quantifies why larger vocabularies should use block or sketched statistics.

G Second-Backbone Transfer

Before execution, we froze an untuned Faster R-CNN MobileNetV3-Large-FPN check (Howard et al., 2019). It reuses $r = 96$, $q = 193$, seed 42, $\lambda = 0.01$, five dominant-label clients, VOC2007 trainval, and VOC2007 test. The analytic mode uses threshold 0.05; prior-calibrated region mode reuses the ResNet-selected $\alpha = 4$ and threshold 0.02. Passing requires AP50 ≥ 0.65, exact pooled recovery, and float32 AP50 loss ≤ 0.002.

The analytic mode obtains 58.35 AP50. Prior-calibrated region mode reaches 83.92 AP50, 60.29 AP75, and 55.04 AP@[.50:.95], compared with 83.25 AP50 for prior-only and 83.58 for the frozen detector. Exact recovery is preserved, and float32 changes AP50 by 3.89×10^{-9}. Extraction,

Table 11: Measured dense server solve and exact per-client payload scaling.

q	K	solve (ms)	float32 upload (MB)
97	21	0.05	0.43
193	21	22.23	1.67
385	21	35.53	6.58
769	21	60.21	26.13
97	81	0.10	1.59
193	81	15.96	6.21
385	81	31.80	24.50
769	81	58.93	97.37

compilation, and evaluation take 179.6, 24.3, and 66.5 seconds, respectively. The sufficient statistics remain 1,666,924 bytes per client, with no additional upload for prior calibration.

H Independent High-Dimensional Region Realization

To test whether the prior-calibrated construction depends on the compact map in Equation 1, we separately implement a higher-dimensional region compiler:

$$x_R = \left[\bar{f}_R \mid g(b) \mid p^{(0)}_{1:C} \mid s^{(0)}\right] \in \mathbb{R}^{539}, \qquad \phi_R(x_R) = \left[x_R \mid \sqrt{2/p}\cos(x_R\Omega + \beta) \mid 1\right] \in \mathbb{R}^{2588}, \tag{8}$$

where $\bar{f}_R \in \mathbb{R}^{512}$ is pooled RoI information, $g(b) \in \mathbb{R}^6$ is box geometry, $p = 2048$, and Ω, β are fixed random Fourier parameters. It compiles classification and optional box-regression moments through the same additive law. Its selected inference retains frozen proposal boxes and uses Equation 7.

The realization uses a COCO-pretrained Faster R-CNN ResNet-50-FPN-v2, at most 100 class-agnostic proposals per 640-pixel image, $\gamma = 1$, seed 42, $\lambda_{\text{cls}} = 0.1$, $\alpha = 4$, and threshold 0.02. VOC2007 train/val selects the operating point; trainval compilation yields 193,458 labeled regions and VOC2007 test is evaluated once. It reaches 85.60 AP50, compared with 84.73 for its matched zero-shot prior. Its default analytic configuration obtains 82.29; additive score blending raises this to 85.00, disabling analytic box refinement raises it to 85.51, and the held-out selected regularization and threshold produce 85.60. The maximum federated-versus-pooled statistic difference is 2.3×10^{-13}, at float64 summation precision.

This independent implementation reproduces the prior-calibration mechanism with a different feature dimension, proposal count, and evaluator. It is therefore corroborating evidence rather than a direct comparison with the compact mode's 89.75 AP50, and it is neither a third protocol nor part of the compact sufficient-statistic configuration.

I Linear-Attention Compilation

For tap ℓ, a public random projection followed by softplus, an appended bias coordinate, and row normalization defines $\kappa_\ell(z) \in \mathbb{R}^{q_A}_+$. Client s transmits $S_{s,\ell,c} = \sum_{i:y_i=c} \kappa_\ell(z_i)$ and $n_{s,c} = |\{i : y_i = c\}|$. After summation and the weights a_c from Equation 4, the server forms $M_{\ell,:,c} = a_c \sum_s S_{s,\ell,c}$ and $m_\ell = \sum_c M_{\ell,:,c}$, then evaluates

$$p^A_{\ell,c}(z) = \frac{\kappa_\ell(z)^\top M_{\ell,:,c}}{\max\{\kappa_\ell(z)^\top m_\ell, \epsilon\}}. \tag{9}$$

Because every transmitted field is a sum over proposal rows, recovery is exactly invariant to horizontal partitioning in real arithmetic. Unlike the ridge statistics, this memory contains no Gram matrix and requires no linear solve.

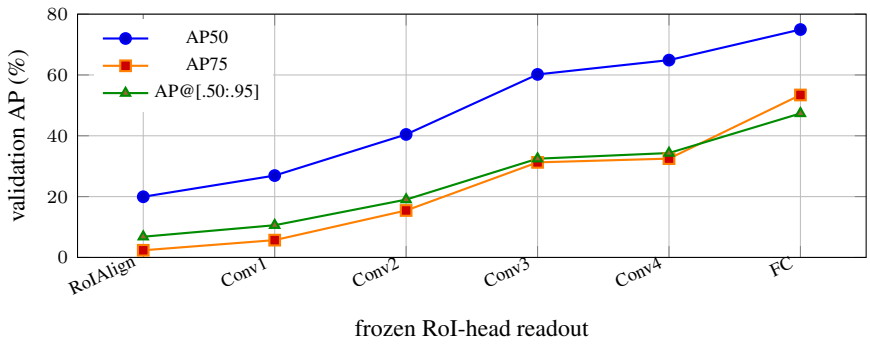

Figure 3: Controlled readout-depth sweep. Performance increases monotonically through the frozen RoI head, with the largest strict-IoU gain at the final FC representation.

J Controlled Architecture and Representation Checks

The DINOv2 study retains the Faster R-CNN RPN proposals but replaces the RoI-head representation with proposal-aligned pooling from a frozen final patch-token map (Oquab et al., 2024). Images are resized to a 518-pixel long side. Table 12 reports all three scales together with the frozen FCOS result.

Table 12: Controlled VOC2007 validation architecture checks. DINOv2 rows use the shared Faster R-CNN proposals and analytic compiler; FCOS is a complete frozen detector.

Model or frozen region encoder	AP50	AP75	AP@[.50:.95]
Analytic FC readout	74.94	53.42	47.35
DINOv2-S/14 (12 blocks, 384-d)	8.83	0.81	2.86
DINOv2-B/14 (12 blocks, 768-d)	8.87	1.08	3.05
DINOv2-L/14 (24 blocks, 1024-d)	8.44	0.84	2.79
Frozen FCOS-ResNet50-FPN	**78.57**	**61.99**	**55.86**

Increasing DINOv2 scale does not improve the direct patch-map interface. A 500-image Small-model study across blocks 3/6/9/12 obtains AP50 0.74/2.17/7.39/9.98, respectively, with the final block performing best. The frozen FCOS result confirms the evaluator on a conventional detector. Together, these results motivate proposal-aware token aggregation or a detector-trained transformer interface for ViT backbones.