FedORCA: Federated Orbit-Aware Residual Contextual Adaptation for Non-Stationary Remote Sensing Classification

Gengjun Huang[1], Yi Zhang[2](✉), Zuming Lin[2], and Yifan Liu[2]

[1] Institute of Artificial Intelligence, Xiamen University, Xiamen, China
[2] Department of Information and Communication Engineering, School of Informatics, Xiamen University, Xiamen, China
yizhang@xmu.edu.cn

Abstract. Remote sensing classification in low-Earth-orbit constellations learns from geographically distributed image streams with scene mixtures that evolve along orbital ground tracks. We therefore propose FedORCA, a federated orbit-aware residual contextual adaptation framework. Its shared encoder and classifier learn transferable scene semantics, while a lightweight private low-rank adapter is modulated by the Hellinger embedding of the current land-cover state. The private adapter adds only 0.30M onboard parameters to the 11.19M-parameter shared model and remains local. Based on the observed orbital paths, we establish a finite-time dynamic tracking bound and a conditional stationarity theorem for the orbit-average objective. Experiments use a Walker–Delta satellite constellation with samples dynamically assigned by land-cover distribution. FedORCA improves the best final-window results by 11.8 and 10.3 points on EuroSAT and NWPU-RESISC45 and reaches five-round mean targets in 15.3% and 18.8% fewer rounds.

Keywords: Federated learning · Spatiotemporal heterogeneous data · Space science data mining · Satellite constellations

1 Introduction

Remote sensing classification assigns semantic categories to Earth observation images and supports land-use monitoring, environmental assessment, and disaster response [3,7]. Low-Earth-orbit (LEO) constellations extend this task to geographically distributed image streams collected along orbital ground tracks. Transferring raw imagery to a single ground node incurs bandwidth, storage, latency, and energy costs; federated learning (FL) enables satellites to train near their observations and exchange model state across the constellation [14,17].

Orbital observation couples spatial and temporal distribution variation. At a given round, satellites cover different geographic regions and observe different scene mixtures. Across rounds, each satellite traverses successive ground-track segments, so its local scene distribution evolves with orbital motion. The associated land-cover mixture is available from the ground track and provides an observable coordinate for this evolution. Effective classification therefore requires a model that transfers shared visual knowledge across satellites, adapts to the current local mixture, and reaches useful accuracy in few federated rounds.

FedORCA realizes this objective through a shared visual predictor and orbit-aware residual contextual adaptation onboard each satellite. The shared encoder and classifier consolidate scene semantics across the constellation. Each satellite retains a low-rank adapter whose response is modulated by the Hellinger embedding of its current land-cover state. This shared–private partition supports constellation-level transfer and continuous local specialization within one federated model. Our contributions are listed below.

- We formulate non-stationary federated remote sensing classification under orbit-induced spatial and temporal distribution heterogeneity and develop FedORCA, whose Hellinger-embedded land-cover state drives an onboard low-rank residual adapter while the reusable encoder and classifier are shared through federated aggregation.
- For the cold-start Adam and FedAvg semantics, we prove a finite-time tracking bound along an orbital q-path; a diminishing-step-size theorem establishes conditional orbit-average stationarity under a positive descent margin.
- In a 200-satellite Walker–Delta simulation, samples are dynamically assigned by land-cover distribution. FedORCA improves final-window accuracy by 11.8 and 10.3 percentage points and reaches five-round trailing-mean targets in 15.3% and 18.8% fewer rounds.

2 Related Work

2.1 Remote Sensing Classification

Remote sensing classification maps Earth observation images to semantic scene categories. The NWPU-RESISC45 benchmark established a diverse 45-class evaluation setting and a systematic account of remote sensing classification methods [3]. EuroSAT complements this setting with Sentinel-2 imagery for land-use and land-cover recognition [7]. Together, these benchmarks provide two distinct visual taxonomies for evaluating classification under orbit-induced scene evolution. Our study embeds their samples in geographically conditioned satellite streams and evaluates prediction on successive ground-track segments.

2.2 Federated Learning for Distributed Earth Observation

FedAvg coordinates distributed learning through local optimization and weighted model averaging [15], while FedProx adds a proximal term for heterogeneous federated environments [12]. Satellite FL extends this coordination to orbital networks and Earth observation workloads [14,17]. FedSN combines submodel training with pseudo-synchronous aggregation [13], while FedLEO supports decentralized offloading [20]. These mechanisms establish a foundation for onboard coordination, but assume that local data distributions remain fixed during training and thus model static client distributions over a time-varying network, whereas FedORCA accounts for the dynamic evolution of orbital data.

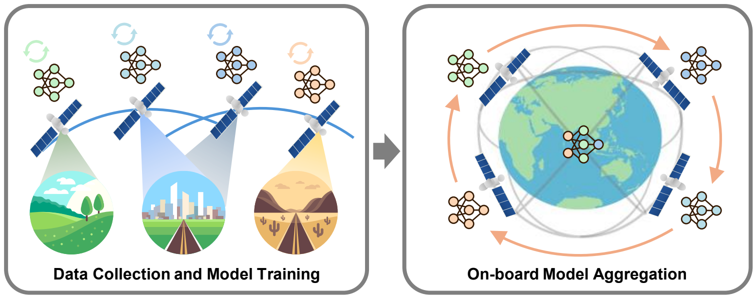

Fig. 1. One-round satellite federated learning workflow.

2.3 Personalized and Adaptive Federated Learning

Personalized federated learning uses selective parameter sharing to preserve client-specific modeling capacity while promoting cross-client knowledge transfer. FedRep shares a representation and retains a local classifier [4]. When client distributions evolve over time, personalization also requires continual adaptation to distribution change. Concept-drift research characterizes abrupt, incremental, gradual, and recurring distribution changes [6], while GLFC addresses federated class-incremental stages through semantic relation distillation and server model selection [5]. For orbital observation streams, FedORCA uses current land-cover context to modulate each satellite's private residual adapter for present- and next-segment classification.

3 Non-Stationary Federated Remote Sensing Classification

3.1 Orbital Observation Streams

Federated learning follows orbital observation in rounds. Each satellite collects current ground-track images, trains the classifier locally, and shares the reusable model block for constellation-level aggregation. Figure 1 shows this workflow, with raw images and the scene-conditioned adapter retained onboard.

3.2 Ground-Track State and Dynamic Scene Distributions

We consider a Walker–Delta constellation of N satellites. At time t, the orbital phase of satellite n evolves as

$$u_n(t) = u_n(0) + \omega_{\text{orbit}} t, \tag{1}$$

where $u_n(0)$ is the orbital phase of satellite n at initialization, and ω_{orbit} is the orbital angular velocity. The corresponding orbital state determines the sub-satellite latitude $\varphi_n(t)$ and longitude $\lambda_n(t)$. Its accumulated coverage of land-cover type j is

$$G_{n,j}^{(r)} := \int_{t_r^-}^{t_r^+} g_j(\varphi_n(t), \lambda_n(t))\, dt. \tag{2}$$

where $G_{n,j}^{(r)}$ is the accumulated coverage of type j by satellite n in round r, $[t_r^-, t_r^+]$ is that round's acquisition interval, and $g_j(\varphi, \lambda)$ is the local proportion of land-cover type j at coordinate (φ, λ). In round r, the trajectory-induced normalized proportion of land-cover type j and its Hellinger state are

$$q_{n,j}^{(r)} := \frac{G_{n,j}^{(r)}}{\sum_{\kappa=1}^{J} G_{n,\kappa}^{(r)}}, \qquad \boldsymbol{s}_n^{(r)} := \sqrt{\boldsymbol{q}_n^{(r)}} \in \mathbb{R}^J. \tag{3}$$

The input space is $\mathcal{X} \subseteq \mathbb{R}^p$, and the label space is $\mathcal{Y} = \{1, \ldots, K\}$. Here, p is the input dimension and K is the number of classes. Let $\Gamma(j) \subseteq \mathcal{Y}$ be the dataset labels compatible with land-cover type j. The fixed semantic map used by the data generator converts orbital coverage into the estimated label marginal

$$\widetilde{P}_n^{(r)}(y = c) := \sum_{j=1}^{J} w_{j\to c} q_{n,j}^{(r)}, \qquad w_{j\to c} := \begin{cases} 1/|\Gamma(j)|, & c \in \Gamma(j), \\ 0, & c \notin \Gamma(j). \end{cases} \tag{4}$$

where y is the label variable, c is a class index, $\widetilde{P}_n^{(r)}$ is the estimated label marginal, and $w_{j\to c}$ is the semantic weight from land-cover type j to class c. From the same segment, client n observes

$$\mathcal{D}_n^{(r)} = \{(\boldsymbol{x}_{n,i}^{(r)}, y_{n,i}^{(r)})\}_{i=1}^{|\mathcal{D}_n^{(r)}|} \sim P_n^{(r)}. \tag{5}$$

where $\mathcal{D}_n^{(r)}$ is the round-r sample set, $\boldsymbol{x}_{n,i}^{(r)}$ and $y_{n,i}^{(r)}$ are the corresponding input and label, $|\mathcal{D}_n^{(r)}|$ is the sample count, and $P_n^{(r)}$ is the underlying joint data distribution. At a fixed round, satellites traverse different regions, so typically $\boldsymbol{q}_n^{(r)} \neq \boldsymbol{q}_m^{(r)}$, $\widetilde{P}_n^{(r)} \neq \widetilde{P}_m^{(r)}$, and $P_n^{(r)} \neq P_m^{(r)}$ for some $n \neq m$; this is the spatial component. For one satellite, successive ground-track segments change the same quantities, so typically $P_n^{(r)} \neq P_n^{(r+1)}$; this is the temporal component. Orbital motion couples the spatial and temporal components into a single continuous process. Spatial overlap supports cross-satellite knowledge transfer, whereas temporal variation calls for adaptation conditioned on the current state $\boldsymbol{s}_n^{(r)}$. Figure 2 shows the orbital data stream of a representative satellite. Its ground-track land-cover mixture changes across rounds, producing continuous temporal variation in addition to spatial heterogeneity across satellites.

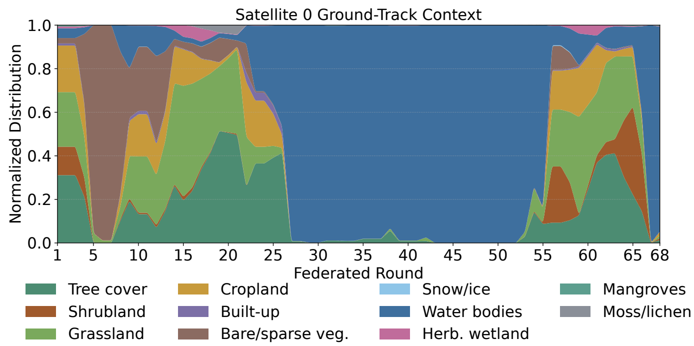

Fig. 2. A representative satellite's orbital data stream.

3.3 Dynamic Federated Classification Objective

The shared parameters are $\theta^{(r)}$, and the private parameters of client n are $\phi_n^{(r)}$. Let

$$z^{(r)} := (\theta^{(r)}, \phi_1^{(r)}, \dots, \phi_N^{(r)}). \tag{6}$$

where $z^{(r)}$ concatenates the shared block and all N private blocks. For classification loss ℓ, the local empirical risk is

$$\widehat{\mathcal{L}}_n^{(r)}(\theta, \phi_n) := \frac{1}{|\mathcal{D}_n^{(r)}|} \sum_i \ell\Big(f(\boldsymbol{x}_{n,i}^{(r)}; \theta, \phi_n), y_{n,i}^{(r)}\Big). \tag{7}$$

where $f(\boldsymbol{x}; \theta, \phi_n)$ uses the shared and private blocks, and the sum covers client n's round-r samples. With empirical weights

$$\widehat{\omega}_n^{(r)} := \frac{|\mathcal{D}_n^{(r)}|}{\sum_{m=1}^N |\mathcal{D}_m^{(r)}|}, \qquad \sum_{n=1}^N \widehat{\omega}_n^{(r)} = 1, \tag{8}$$

the round-r objective is

$$\widehat{\mathcal{L}}^{(r)}(z) := \sum_{n=1}^N \widehat{\omega}_n^{(r)} \widehat{\mathcal{L}}_n^{(r)}(\theta, \phi_n). \tag{9}$$

Therefore, the learning goal is to track approximate stationary points of the changing sequence $\{\widehat{\mathcal{L}}^{(r)}\}_{r\geq 0}$ while preserving a useful shared representation.

4 FedORCA

4.1 Overview

FedORCA provides distribution-aware local adaptation for evolving satellite distributions. As an orbit progresses, each satellite crosses different regions and its suitable model changes across rounds. Satellites in different orbital planes should nevertheless share semantic representations when observing similar land-cover types. FedORCA therefore uses the current ground-track distribution as the basic unit of adaptation.

Static local classification heads conflict with this objective in two respects. First, all satellites use the same land-cover category system, so fully privatizing decision boundaries discards valuable cross-satellite supervision. Second, an individual satellite moves along a continuously evolving data stream, causing a context-invariant module to average the responses required by mutually conflicting ground-track segments. The predictor must therefore satisfy three requirements. (i) retain a constellation-wide jointly trained visual representation and classifier; (ii) continuously vary its local response according to the observable ground-track mixture; and (iii) adapt directly from the observable state. FedORCA writes the predictor of satellite n in round r as

$$f_n^{(r)}(\boldsymbol{x}) = C\Big(\mathcal{E}(\boldsymbol{x}) + \mathcal{A}_n\Big(\mathcal{E}(\boldsymbol{x}); \boldsymbol{s}_n^{(r)}\Big)\Big). \tag{10}$$

where $f_n^{(r)}$ is the predictor of satellite n in round r, $\mathcal{E}$ is the shared encoder, C is the shared classifier, and $\mathcal{A}_n$ is the private adapter conditioned on the current state. Client identity determines where the private parameters reside, and the current state determines their response.

4.2 Hellinger State Representation

The conditioning variable is constructed from land-cover metadata. The square-root embedding in Eq. (3) maps normalized histograms to the unit sphere and gives low-frequency land-cover types greater resolution than raw proportions. For normalized histograms $\bar{\boldsymbol{q}}$ and $\bar{\boldsymbol{q}}'$, the embedding satisfies

$$H^2(\bar{\boldsymbol{q}}, \bar{\boldsymbol{q}}') = \frac{1}{2}\left\|\sqrt{\bar{\boldsymbol{q}}} - \sqrt{\bar{\boldsymbol{q}}'}\right\|_2^2 = \frac{1}{2}\|\boldsymbol{s} - \boldsymbol{s}'\|_2^2, \qquad \|\boldsymbol{s}\|_2 = 1, \tag{11}$$

where H is the Hellinger distance, and $\boldsymbol{s} = \sqrt{\bar{\boldsymbol{q}}}$ and $\boldsymbol{s}' = \sqrt{\bar{\boldsymbol{q}}'}$ are the corresponding square-root embeddings. Thus, Euclidean distance in the conditioning space is proportional to the Hellinger distance, continuously representing orbital distribution changes while separating distinct land-cover mixtures.

4.3 Orbit-Aware Residual Contextual Adaptation

The architecture combines a shared prediction path with a private residual correction path. Its shared–private parameter partition is

$$\theta = \{\mathcal{E}, C\}, \qquad \phi_n = \{\mathcal{A}_n\}. \tag{12}$$

where θ is the shared parameter block composed of the shared encoder $\mathcal{E}$ and shared classifier C, while ϕ_n is the private parameter block of satellite n, composed of its private adapter $\mathcal{A}_n$.

The encoder accumulates visual primitives from diverse observations, while the classifier consolidates evidence for the common land-cover taxonomy. The context-dependent adapter is inserted between them, where it can recalibrate the representation before a shared decision boundary is applied. This placement assigns global parameters to recurring semantics and reserves private capacity for the part that changes with the ground track. For input $\boldsymbol{x}$,

$$\boldsymbol{z} = \mathcal{E}(\boldsymbol{x}) \in \mathbb{R}^d. \tag{13}$$

where $\boldsymbol{z}$ is the shared feature representation and d is its dimension.

Satellite n owns an onboard private adapter $\mathcal{A}_n = \{W_{\mathrm{dn},n}, W_{\mathrm{up},n}, \mathcal{H}_n, \boldsymbol{b}_n\}$. The hypernetwork $\mathcal{H}_n$ maps the current context to FiLM parameters [16].

$$(\boldsymbol{\gamma}_n^{(r)}, \boldsymbol{\beta}_n^{(r)}) = \mathcal{H}_n(\boldsymbol{s}_n^{(r)}). \tag{14}$$

where $\boldsymbol{\gamma}_n^{(r)}$ and $\boldsymbol{\beta}_n^{(r)}$ are the feature-wise scale and shift generated for satellite n in round r, respectively. The scale selects which latent directions should be emphasized under the current mixture, whereas the shift supplies a context-specific offset. The hypernetwork interpolates between mixtures over a continuous context space, replacing discrete region tables and predefined tasks.

The adapter follows a low-rank bottleneck inspired by parameter-efficient adaptation [8,9].

$$\boldsymbol{\delta}_n = W_{\mathrm{up},n}\,\mathrm{GELU}\Big(\boldsymbol{\gamma}_n^{(r)} \odot W_{\mathrm{dn},n}\boldsymbol{z} + \boldsymbol{\beta}_n^{(r)}\Big) + \boldsymbol{b}_n, \tag{15}$$

where $\boldsymbol{\delta}_n$ is the state residual, $W_{\mathrm{dn},n}$ and $W_{\mathrm{up},n}$ the projections, GELU the activation, $\odot$ elementwise multiplication, and $\boldsymbol{b}_n$ the bias.

$$\widetilde{\boldsymbol{z}}_n = \boldsymbol{z} + \boldsymbol{\delta}_n, \qquad \widehat{\boldsymbol{y}}_n = C(\widetilde{\boldsymbol{z}}_n). \tag{16}$$

where $\widetilde{\boldsymbol{z}}_n$ is the adapted feature and $\widehat{\boldsymbol{y}}_n$ is the prediction vector. The two projections have complementary roles. $W_{\mathrm{dn},n}$ compresses the d-dimensional shared representation into a rank-h adaptation space, so the FiLM variables act only on directions that the local client can afford to specialize. $W_{\mathrm{up},n}$ then maps the conditioned correction back to the shared feature space. The bottleneck replaces an unrestricted $d \times d$ transformation with two thin projections, concentrating local capacity on a learnable subspace while keeping the onboard state small. The context acts on $\boldsymbol{\delta}_n$, while the full semantic backbone $\boldsymbol{z}$ remains available.

The residual form in Eq. (16) preserves the shared predictor at initialization. The residual initialization sets $W_{\mathrm{up},n} = 0$, initializes the FiLM scale and shift to one and zero, and sets $\boldsymbol{b}_n = 0$. The initial model therefore exactly equals the shared baseline. Local training then develops a scene-conditioned correction from the shared starting point.

4.4 Federated Training and Selective Aggregation

At the start of round r, every satellite receives the same shared block $\theta^{(r)}$ and retains its own $\phi_n^{(r)}$. Satellite n updates both parameter blocks on $\mathcal{D}_n^{(r)}$. If $\theta_n^{(r,\mathrm{loc})}$ is the locally updated shared block, the next shared state is

$$\theta^{(r+1)} = \sum_{n=1}^{N} \widehat{\omega}_n^{(r)} \theta_n^{(r,\mathrm{loc})}, \tag{17}$$

where $\theta^{(r+1)}$ is the next-round shared state; the private state $\phi_n^{(r+1)}$ instead remains at client n. Averaging the shared block pools repeated visual and label evidence. Retaining each private block preserves correction functions learned from its orbital history, including ground-track-specific amplification and suppression of latent directions. One round therefore follows four operations. The server broadcasts $\theta^{(r)}$, each client sets the current state $s_n^{(r)}$ and jointly optimizes (θ, ϕ_n) on the new segment, and the server aggregates only the returned $\theta_n^{(r,\mathrm{loc})}$.

5 Theoretical Analysis

5.1 Exact Dynamic Model

All N satellites participate in every completed round.[3] For sample weights $\widehat{\omega}_n^{(r)}$ and orbital objectives $f_{n,r} := f_{q_{n,r}}$, define

$$\begin{aligned} z_r &= (\theta_r, \phi_{1,r}, \dots, \phi_{N,r}), \qquad F_r(z) = \sum_n \widehat{\omega}_n^{(r)} f_{n,r}(\theta, \phi_n), \\ G_r^{(\lambda)} &= \|\nabla_\theta F_r(z_r)\|^2 + \lambda \sum_n \widehat{\omega}_n^{(r)} \|\nabla_{\phi_n} f_{n,r}(z_r)\|^2. \end{aligned} \tag{18}$$

Write $G_r^{(1)} = \|\nabla_{\widehat{\omega}} F_r(z_r)\|^2_{\widehat{\omega},r}$ and $c_\lambda = \max\{1, \lambda\}$, so $G_r^{(\lambda)} \le c_\lambda G_r^{(1)}$.

For $x_{n,r,k} = (\theta_{n,r,k}, \phi_{n,r,k})$, every round reconstructs standard Adam [11] from $m_{n,r,0} = v_{n,r,0} = 0$.

$$\begin{aligned} m_{n,r,k} &= \beta_1 m_{n,r,k-1} + (1-\beta_1) g_{n,r,k}, \quad v_{n,r,k} = \beta_2 v_{n,r,k-1} + (1-\beta_2) g_{n,r,k}^{\odot 2}, \\ \widehat{m}_{n,r,k} &= \frac{m_{n,r,k}}{1-\beta_1^k}, \quad \widehat{v}_{n,r,k} = \frac{v_{n,r,k}}{1-\beta_2^k}, \quad x_{n,r,k} = x_{n,r,k-1} - \eta_r \frac{\widehat{m}_{n,r,k}}{\sqrt{\widehat{v}_{n,r,k}} + \varepsilon}, \end{aligned} \tag{19}$$

where $x_{n,r,0} = (\theta_r, \phi_{n,r})$, $0 \le \beta_1 < 1$, $0 < \beta_2 < 1$, and $\varepsilon > 0$. For $d_{n,r,k} := \widehat{m}_{n,r,k} / (\sqrt{\widehat{v}_{n,r,k}} + \varepsilon)$, the exact FedAvg round map is

$$\begin{aligned} D_r &= \left(\sum_n \widehat{\omega}_n^{(r)} \sum_{k=1}^{K_{n,r}} d_{n,r,k}^{\theta}, \left\{ \sum_{k=1}^{K_{n,r}} d_{n,r,k}^{\phi_n} \right\}_{n=1}^{N} \right), \\ z_{r+1} &= z_r - \eta_r D_r, \end{aligned} \tag{20}$$

and $\mathcal{H}_r$ denotes the round-start history.

[3] Detailed proofs of the theoretical results are available in the repository at `https://anonymous.4open.science/r/FedORCA-2F4B/`.

Table 1. Observed 3° orbital q-path statistics derived from ESA WorldCover 2021.

Statistic	Value	Statistic	Value
Satellites	200	Zero-change fraction	0.4467
Median	0.0397	95th percentile	0.5339
Maximum	0.8565	Largest-decile share	0.3727

5.2 Orbital Stream Analysis

We analyze the 10 m global land-cover map from ESA WorldCover 2021 [19] at 3° resolution. The statistics are computed over one complete orbital period and are summarized in Table 1.

The 44.67% zero-change rate, together with the large 95th percentile, maximum, and largest-decile contribution, identifies stable plateaus punctuated by jumps and supports a round-frozen, piecewise-constant model. We measure these changes by $h_{n,r} = H(q_{n,r+1}, q_{n,r})$ and $V_R^{\mathrm{orb}} = \sum_{r<R} \sum_n \widehat{\omega}_n^{(r)} h_{n,r}$.

5.3 Convergence Results

We use the following assumptions. A1, $f_{n,r}$ is fixed within each round; A2, $F_r, f_{n,r}$ are L-smooth and $F_{\mathrm{inf}} := \inf_{r,z} F_r(z) > -\infty$, with bounded component losses/gradients B_q, G_q and locally Lipschitz dependence on $\sqrt{q}$ with constants L_s^0, L_s^1; A3, $\|g_{n,r,k}\|_\infty \le G_\infty$, $\mathbb{E}[\|D_r\|^2_{\widehat{\omega},r} \mid \mathcal{H}_r] \le M_r^2$, and $K_- \le K_{n,r} \le K_+$; and A4,

$$\begin{aligned} \overline{D}_r &= A_r \nabla_{\widehat{\omega}} F_r(z_r) + \sum_{j \in \mathcal{J}} E_r^j, \\ |\langle \nabla_{\widehat{\omega}} F_r, E_r^j \rangle_{\widehat{\omega},r}| &\le \rho_r^j G_r^{(1)} + b_r^j, \quad \mu_r = p_- - \sum_j \rho_r^j, \quad b_r = \sum_j b_r^j. \end{aligned} \tag{21}$$

Here $(G_\infty + \varepsilon)^{-1} I \preceq P_{n,r,k} \preceq \varepsilon^{-1} I$ gives $p_- I \preceq A_r \preceq p_+ I$, with $p_- = K_-/(G_\infty + \varepsilon)$ and $p_+ = K_+/\varepsilon$; $\mathcal{J} = \{\mathrm{mom}, \mathrm{samp}, \mathrm{cov}, \mathrm{pre}, \mathrm{local}, \mathrm{orb}\}$. The cold start sets $E_r^{\mathrm{mom}} = 0$, while E_r^{cov} retains the $g_{n,r,k}$–$v_{n,r,k}$ dependence. A2 also yields $|f_{n,r+1} - f_{n,r}| \le L_0 h_{n,r}$ and $\|\nabla f_{n,r+1} - \nabla f_{n,r}\| \le L_q h_{n,r}$, where $L_0 = \sqrt{2}(2B_q + L_s^0)$ and $L_q = \sqrt{2}(2G_q + L_s^1)$.

Theorem 1 (Finite-time tracking). *Under A1–A4, let* $\mu_0 = \inf_{r<R} \mu_r > 0$ *and* $S_R = \sum_{r<R} \eta_r$. *Then*

$$\begin{aligned} \frac{1}{S_R} \sum_{r<R} \eta_r \mathbb{E}[G_r^{(\lambda)}] \le \frac{c_\lambda}{\mu_0 S_R} \Bigg[& F_0(z_0) - F_{\mathrm{inf}} + L_0 V_R^{\mathrm{orb}} \\ & + \sum_{r<R} \eta_r b_r + \frac{L}{2} \sum_{r<R} \eta_r^2 M_r^2 \Bigg]. \end{aligned} \tag{22}$$

For $\eta_r \equiv \eta$, the $O(\eta)$ smoothness term, residual floor, and $V_R^{\mathrm{orb}}/(R\eta)$ imply finite-radius dynamic tracking. Asymptotic convergence requires the additional conditions below.

Theorem 2 (Orbit-average stationarity).
If the weighted geometry converges, $\bar{F} := \lim_{R\to\infty} R^{-1} \sum_{r<R} F_r$ has locally uniform gradient convergence, A4 holds relative to $\nabla_{\widehat{\omega}} \bar{F}$ with $(\widetilde{\rho}_r^j, \widetilde{b}_r^j)$, and

$$\inf_r \left(p_- - \sum_j \widetilde{\rho}_r^j \right) > 0, \qquad \sum_r \eta_r = \infty, \qquad \sum_r \eta_r^2 < \infty,$$
$$\sum_r \eta_r \sum_j \widetilde{b}_r^j < \infty, \qquad \sum_r \eta_r^2 M_r^2 < \infty, \tag{23}$$

then

$$\lim_{R\to\infty} \frac{1}{S_R} \sum_{r<R} \eta_r \mathbb{E} \|\nabla_{\widehat{\omega}} \bar{F}(z_r)\|_{\widehat{\omega}}^2 = 0. \tag{24}$$

Hence the corresponding lim inf *is zero.*

6 Evaluation

6.1 Experimental Setup

Constellation and ground-track streams. We simulate 200 LEO satellite clients in a Walker-Delta constellation with 20 orbital planes, 10 satellites per plane, altitude 850 km, and inclination 50°.

Datasets. EuroSAT [7] and NWPU-RESISC45 [3] are assigned using the 10 m global map from ESA WorldCover 2021 [19], which is also used to instantiate $w_{j\to c}$ in Eq. (4).

Models and training. All methods use a ResNet-18 backbone with $d = 512$. The FedORCA adapter rank is $h = 256$, and its context hypernetwork has hidden width 64. We use Adam with learning rate 10^{-3}, one local epoch, and batch size 16.

Baselines. We compare with nine baselines covering independent, conventional, drift-corrected, and personalized training. LocalOnly [2] trains each satellite independently without aggregation. FedAvg [15] performs weighted model averaging, FedProx [12] adds a proximal regularizer, and SCAFFOLD [10] corrects local gradients with server and client control variates. FedLC [21] calibrates local logits using class counts, while FedRep [4] shares the feature extractor and retains a local classifier. FLoRAL [1] combines a shared backbone and low-rank adapter bank with client-specific routing, and pFedHN [18] generates personalized model parameters from client embeddings through a server hypernetwork. GLFC [5] combines local forgetting compensation with server model selection.

Evaluation. The round-r model is trained on $\mathcal{D}_n^{(r)}$ with context $\boldsymbol{s}_n^{(r)}$, and evaluated on $\mathcal{D}_n^{(r+1)}$ before the next local update. We report the mean next-segment accuracy over the last 5 and 10 rounds as $\bar{A}_5$ and $\bar{A}_{10}$, respectively.

Table 2. Mean next-segment accuracy at selected rounds and over the final and full-stream windows. Best means are bold; second best are underlined.

Method	R=10	R=20	R=30	R=40	R=50	R=60	$\bar{A}_5$	$\bar{A}_{10}$	$\bar{A}_{\text{all}}$
EuroSAT									
LocalOnly	0.624	0.639	0.663	0.669	0.660	0.674	0.667	0.667	0.635
FedAvg (2017)	0.503	0.589	0.631	0.658	0.699	0.720	0.732	0.729	0.626
FedProx (2020)	0.554	0.586	0.608	0.624	0.642	0.674	0.685	0.680	0.605
SCAFFOLD (2020)	0.253	0.261	0.251	0.223	0.243	0.237	0.225	0.231	0.239
FedRep (2021)	0.621	0.713	0.749	0.741	0.739	0.720	0.736	0.736	0.690
pFedHN (2021)	0.091	0.140	0.199	0.219	0.253	0.151	0.189	0.169	0.175
FedLC (2022)	0.546	0.582	0.629	0.676	0.706	0.726	0.742	0.736	0.624
GLFC (2022)	0.524	0.614	0.658	0.685	0.686	0.709	0.717	0.713	0.626
FLoRAL (2025)	0.490	0.580	0.615	0.637	0.681	0.699	0.724	0.719	0.613
FedORCA	**0.646**	**0.749**	**0.816**	**0.824**	**0.853**	**0.849**	**0.857**	**0.854**	**0.755**
NWPU-RESISC45									
LocalOnly	0.440	0.456	0.475	0.502	0.505	0.517	0.518	0.517	0.472
FedAvg (2017)	0.411	0.436	0.455	0.471	0.580	0.619	0.611	0.601	0.461
FedProx (2020)	0.405	0.451	0.485	0.521	0.607	0.643	0.663	0.649	0.491
SCAFFOLD (2020)	0.128	0.141	0.131	0.128	0.145	0.141	0.134	0.138	0.137
FedRep (2021)	0.482	0.553	0.582	0.609	0.624	0.637	0.629	0.629	0.562
pFedHN (2021)	0.022	0.031	0.160	0.164	0.110	0.065	0.058	0.062	0.091
FedLC (2022)	0.242	0.257	0.280	0.327	0.314	0.360	0.368	0.358	0.283
GLFC (2022)	0.403	0.494	0.509	0.522	0.570	0.586	0.592	0.585	0.497
FLoRAL (2025)	0.396	0.419	0.471	0.466	0.561	0.603	0.576	0.571	0.447
FedORCA	**0.500**	**0.579**	**0.635**	**0.688**	**0.725**	**0.746**	**0.754**	**0.752**	**0.621**

6.2 Overall Performance and Round Efficiency

Table 2 shows that FedORCA ranks first at every reported round and averaging window on both streams. For $\bar{A}_{10}$, FedORCA improves over the strongest final-window baselines by 11.8 percentage points on EuroSAT and 10.3 points on NWPU-RESISC45; the corresponding $\bar{A}_5$ gains are 11.5 and 9.1 points. pFedHN and SCAFFOLD remain below the other methods under the same protocol. These two low results reflect a mismatch between their adaptation mechanisms and the dynamic evaluation protocol. pFedHN maps a fixed satellite-identity embedding to a complete personalized model, whereas the class distribution associated with a satellite changes from segment to segment along its ground track. Because the embedding contains no current orbital context, successive, different local objectives are compressed into one client representation, so the generated model can lag the next-segment distribution. By contrast, SCAFFOLD uses historical control variates to correct client-specific gradient drift around a shared objective, but local objectives in our setting change across orbital segments, so a variate estimated on the previous segment can become stale and partially offset the update needed to adapt to the next distribution.

Figure 3 traces next-segment accuracy over the orbital streams, while Table 2 reports the selected-round and full-stream means. At round 10, LocalOnly is the closest EuroSAT comparator and FedRep the closest NWPU-RESISC45 comparator; by round 60, the closest methods are FedLC and FedProx. Accordingly, the margin of FedORCA grows from 2.2 to 12.3 percentage points on EuroSAT and from 1.8 to 10.2 points on NWPU-RESISC45. FedAvg, FLoRAL, and GLFC improve gradually over the stream, while

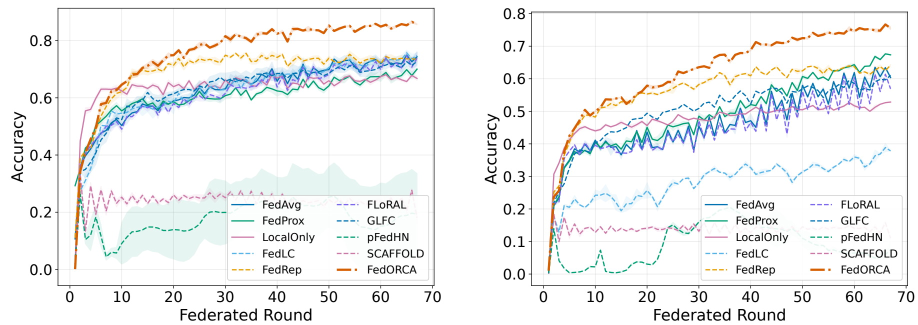

Fig. 3. Mean next-segment accuracy over federated rounds on EuroSAT (left) and NWPU-RESISC45 (right).

Table 3. Ablation results for the final tracking windows.

Variant	EuroSAT		NWPU-RESISC45	
	$\bar{A}_5$	$\bar{A}_{10}$	$\bar{A}_5$	$\bar{A}_{10}$
FedORCA	**0.857**	**0.854**	**0.755**	**0.752**
w/o adapter $\mathcal{A}$	0.732	0.729	0.615	0.606
w/o context $\boldsymbol{s}$	0.820	0.819	0.721	0.717
aggregate low-rank basis	0.834	0.825	0.660	0.653
private classifier	0.842	0.844	0.737	0.733

the independent LocalOnly models plateau. Across all 67 rounds, FedORCA exceeds FedRep, the strongest full-stream baseline, by 6.6 and 5.9 percentage points on EuroSAT and NWPU-RESISC45, respectively. For five-round mean accuracy targets of 0.70 and 0.60, FedORCA requires 15.3% and 18.8% fewer federated rounds than FedRep on EuroSAT and NWPU-RESISC45, respectively, showing that the accuracy gains also translate into faster attainment of the target performance.

6.3 Ablation Analysis

Table 3 reports the ablation results for the final tracking windows. The full adapter contributes 12.5 and 14.6 percentage points to $\bar{A}_{10}$ on EuroSAT and NWPU-RESISC45, identifying the private residual path as the primary source of scene-specific adaptation. Context conditioning contributes a further 3.5 percentage points on each dataset by indexing the residual response with the current scene mixture. Private retention of the low-rank basis contributes 2.9 and 9.9 percentage points, with the larger NWPU-RESISC45 gain reflecting the value of preserving specialized correction subspaces in a 45-class stream. Sharing the classifier contributes 1.0 and 1.9 percentage points by pooling decision-boundary evidence for the common taxonomy. These results support the design.

Table 4. Sensitivity to local epochs E, adapter rank h, and client count N.

Factor	Setting	EuroSAT		NWPU-RESISC45	
		$\bar{A}_5$	$\bar{A}_{10}$	$\bar{A}_5$	$\bar{A}_{10}$
E	**1**	**0.857**	**0.854**	**0.755**	**0.752**
	2	0.842	0.833	0.680	0.671
	4	0.747	0.705	0.603	0.564
h	128	0.856	**0.855**	0.753	0.749
	256	**0.857**	0.854	**0.755**	**0.752**
	512	**0.857**	**0.855**	**0.755**	0.751
N	100	0.859	0.858	0.758	0.753
	200	0.857	0.854	0.755	0.752
	400	0.846	0.841	0.757	0.755

Table 5. Trainable parameters and mean local training time per client in seconds.

Method	M	EuroSAT	NWPU	Method	M	EuroSAT	NWPU
LocalOnly	11.19	2.68	14.78	pFedHN	11.19	2.71	14.83
FedAvg	11.19	2.74	15.01	FedLC	11.19	2.73	14.77
FedProx	11.19	3.25	16.04	GLFC	11.19	3.81	17.26
SCAFFOLD	11.19	3.11	14.85	FLoRAL	11.59	7.47	42.63
FedRep	11.19	3.54	19.98	**FedORCA**	11.49	2.85	15.29

6.4 Sensitivity Analysis

Table 4 reports the sensitivity results across local epochs, adapter ranks, and client counts. One local epoch gives the highest next-segment accuracy. Relative to $E = 1$, two local epochs change $\bar{A}_{10}$ by 2.1 and 8.1 percentage points on EuroSAT and NWPU-RESISC45; four epochs change it by 14.9 and 18.8 points. This pattern highlights the role of cross-round transfer when each short segment provides a temporally localized sample of the classification problem. Ranks $h \in \{128, 256, 512\}$ change $\bar{A}_{10}$ by at most 0.1 percentage points on EuroSAT and 0.3 points on NWPU-RESISC45, indicating that the low-rank bottleneck reaches sufficient capacity by $h = 128$. Changing the federation from 100 to 400 clients yields a 1.7-point EuroSAT range and a 0.3-point NWPU-RESISC45 range. The narrow ranges across adapter widths and client counts demonstrate stable classification under these configurations.

6.5 Computational Efficiency

Table 5 reports one local epoch on 250 current-segment samples. LocalOnly, pFedHN, and FedLC stay near FedAvg because their client backpropagation paths are unchanged. FedProx, SCAFFOLD, FedRep, and GLFC add proximal, control-variate, staged-update, or history-based operations. These raise FedRep by 29.2%/33.1% and GLFC by 39.1%/15.0%. FLoRAL's four routed, gradient-preconditioned branches make it the most expensive baseline. FedORCA conditions one adapter, adding only 3.7%/1.9% over FedAvg and cutting 61.9%/64.1% versus FLoRAL.

7 Conclusion

FedORCA uses Hellinger-conditioned adaptation for non-stationary LEO classification, improving final-window accuracy by 11.8/10.3 points and reaching the targets in 15.3%/18.8% fewer rounds under the established tracking and stationarity guarantees.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant 62401485 and the Fundamental Research Funds for the Central Universities under Grant 20720250028.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Almansoori, A.J., Horváth, S., Takáč, M.: Collaborative and efficient personalization with mixtures of adaptors. In: Conference on Parsimony and Learning. Proceedings of Machine Learning Research, vol. 280, pp. 1328–1364. PMLR (2025), https://proceedings.mlr.press/v280/almansoori25a.html
2. Chen, S., Zheng, Q., Long, Q., Su, W.J.: Minimax estimation for personalized federated learning: An alternative between FedAvg and local training? Journal of Machine Learning Research **24**(262), 1–59 (2023), https://jmlr.org/papers/v24/21-0224.html
3. Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE **105**(10), 1865–1883 (2017)
4. Collins, L., Hassani, H., Mokhtari, A., Shakkottai, S.: Exploiting shared representations for personalized federated learning. In: International Conference on Machine Learning. pp. 2089–2099. PMLR (2021)
5. Dong, J., Wang, L., Fang, Z., Sun, G., Xu, S., Wang, X., Zhu, Q.: Federated class-incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10164–10173 (2022)
6. Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys **46**(4), 1–37 (2014). https://doi.org/10.1145/2523813
7. Helber, P., Bischke, B., Dengel, A., Borth, D.: EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **12**(7), 2217–2226 (2019)
8. Houlsby, N., Giber, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for NLP. In: International Conference on Machine Learning. pp. 2790–2799. PMLR (2019)
9. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2022)
10. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., Suresh, A.T.: SCAFFOLD: Stochastic controlled averaging for federated learning. In: Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 5132–5143. PMLR (2020), https://proceedings.mlr.press/v119/karimireddy20a.html

11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings (2015), https://arxiv.org/abs/1412.6980
12. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
13. Lin, Z., Chen, Z., Fang, Z., Chen, X., Wang, X., Gao, Y.: FedSN: A federated learning framework over heterogeneous LEO satellite networks. IEEE Transactions on Mobile Computing **24**(3), 1293–1307 (2025). https://doi.org/10.1109/TMC.2024.3481275
14. Matthiesen, B., Razmi, N., Leyva-Mayorga, I., Dekorsy, A., Popovski, P.: Federated learning in satellite constellations. IEEE Network **38**(2), 232–239 (2024). https://doi.org/10.1109/MNET.132.2200504
15. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics. pp. 1273–1282. PMLR (2017)
16. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: FiLM: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
17. Razmi, N., Matthiesen, B., Dekorsy, A., Popovski, P.: Ground-assisted federated learning in LEO satellite constellations. IEEE Wireless Communications Letters **11**(4), 717–721 (2022). https://doi.org/10.1109/LWC.2022.3141120
18. Shamsian, A., Navon, A., Fetaya, E., Chechik, G.: Personalized federated learning using hypernetworks. In: Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 9489–9502. PMLR (2021), https://proceedings.mlr.press/v139/shamsian21a.html
19. Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoino, F., Arino, O.: ESA WorldCover 10 m 2021 v200 (2022). https://doi.org/10.5281/zenodo.7254221, https://doi.org/10.5281/zenodo.7254221
20. Zhai, Z., Wu, Q., Yu, S., Li, R., Zhang, F., Chen, X.: FedLEO: An offloading-assisted decentralized federated learning framework for low earth orbit satellite networks. IEEE Transactions on Mobile Computing **23**(5), 5260–5279 (2024). https://doi.org/10.1109/TMC.2023.3304988
21. Zhang, J., Li, Z., Li, B., Xu, J., Wu, S., Ding, S., Wu, C.: Federated learning with label distribution skew via logits calibration. In: Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 162, pp. 26311–26329. PMLR (2022), https://proceedings.mlr.press/v162/zhang22p.html