Capturing Spectral and Spatial Patterns for Federated Remote Sensing Segmentation

Yixin Xue[1,2,3,4,5*], Haonan Guo[6,7*,†], Wenke Huang[8], and Bo Du[1,2,3,4†]

[1] School of Computer Science, Wuhan University, Wuhan, China
[2] National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan, China
[3] Institute of Artificial Intelligence, Wuhan University, Wuhan, China
[4] Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, Wuhan, China
[5] National Center for Applied Mathematics in Hubei, Wuhan University, Wuhan, China
[6] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
[7] Zhongguancun Academy, Beijing, China
[8] College of Computing and Data Science, Nanyang Technological University, Singapore

xueyixin@whu.edu.cn, haonan.guo@whu.edu.cn, wenke.huang@ntu.edu.sg, dubo@whu.edu.cn

Abstract. Federated remote sensing (RS) semantic segmentation enables privacy-preserving collaborative training of dense prediction models. However, its performance is severely hindered by the structured heterogeneity inherent in RS image: sensor diversity causes distinct spectral shifts, while geographic variability induces spatial shifts. Existing methods treat client heterogeneity as an unstructured distributional mismatch, failing to accommodate these coupled, RS-specific variations. To address this, we propose **FedDap**, a federated prototype adaptation framework that disentangles domain-conditioned variations from invariant class semantics via Domain-Adaptive Prototypes (DAPs). DAPs transform globally shared semantic anchors through two specialized modulation branches: a spectral branch for radiometric alignment and a spatial branch for multi-scale structural adaptation. A shift-aware gating mechanism dynamically fuses these branches, producing class-conditional prototypes tailored to local sensing and regional contexts. By employing a pixel-to-prototype contrastive objective, we align local representations with both the shared semantic references and the DAPs, effectively mitigating cross-client mismatch. Extensive experiments across diverse federated RS benchmarks demonstrate that our method consistently outperforms state-of-the-art baselines.

1 Introduction

Remote sensing (RS) semantic segmentation produces fine-grained land-cover maps for critical applications such as urban planning, disaster response, and envi-

* Equal contribution.
† Corresponding authors.

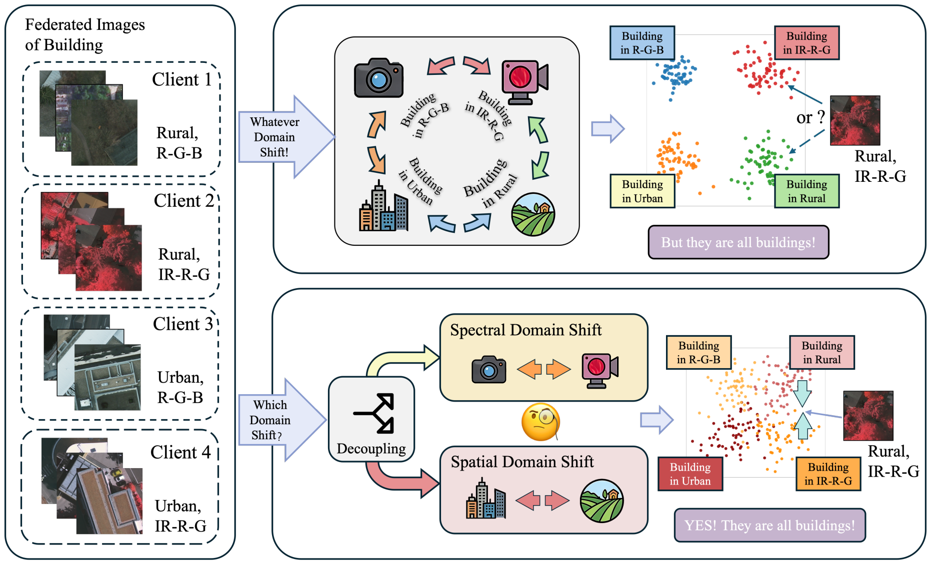

Fig. 1: Motivation of structured domain shifts. Across federated clients, the same class may appear under different sensors and regions. Conventional prototype methods treat this as a single distribution shift, causing class representations to split by domain (top). We instead decouple domain shifts into *spectral* and *spatial* components to restore consistent semantic alignment (bottom).

ronmental monitoring [16,21,38]. However, annotated RS data is often restricted by privacy, policy, and ownership constraints, making centralized training impractical. Federated learning (FL) emerges as a natural solution, enabling collaborative model training on distributed data while keeping raw images strictly local.

Despite its promise, federated RS segmentation suffers from severe client heterogeneity under non-IID data. This heterogeneity leads to inconsistent class representations at the pixel level, exacerbating client drift. Prototype-based FL methods address this by sharing class-wise semantic prototypes [25, 28, 34, 36], which has proven effective in aligning cross-client features and mitigating drift. However, most existing prototype-based methods implicitly treat client heterogeneity as an unstructured distributional mismatch, which is inherently suboptimal for RS image. As illustrated in Fig. 1, samples belonging to the same semantic category (e.g., buildings) may appear under different sensors and geographic regions across clients, leading to multiple domain-dependent clusters in the feature space.

This observation suggests that domain shifts in federated RS segmentation are not arbitrary but highly structured. In particular, they are primarily driven by two factors: **sensor diversity** (reflecting varied radiometric responses and imaging pipelines) and **regional variability** (reflecting distinct landscapes and scene organizations). These factors induce two coupled domain gaps: **a spec-**

tral shift that perturbs channel-wise appearance, and **a spatial shift** that alters multi-scale texture, layout, and morphology. Consequently, relying on a single shared prototype per class inevitably conflates these region- and sensor-dependent cues. This forces an over-averaged representation that fails to capture structured intra-class variances, ultimately leading to conflicting pixel-level supervision across heterogeneous clients.

Motivated by this structured heterogeneity, we propose a federated prototype adaptation framework termed **FedDap**, which disentangles domain-specific variations from invariant class semantics by performing adaptation directly in the prototype space. In FedDap, the server maintains globally shared class prototypes as stable semantic references. Meanwhile, each client learns a prototype modulator to transform these shared anchors into Domain-Adaptive Prototypes (DAPs), tailored to local data contexts.

To explicitly address RS-specific domain gaps, the modulator features two complementary branches: **a spectral branch** that captures shallow, channel-wise statistics to correct sensor-induced radiometric shifts, and **a spatial branch** that extracts deep, multi-scale structural cues to accommodate region-dependent texture and morphology. A shift-aware gating module dynamically fuses these branches by estimating the relative intensity of spectral versus spatial shifts, yielding the final class-conditional DAPs.

For robust optimization, we introduce a pixel-to-prototype contrastive learning objective [2] that aligns local representations with both the global anchors and the local DAPs, alongside a drift regularizer to prevent uncontrolled prototype deformation. Finally, we exploit the cross-client similarity of DAPs to discover latent domains. This enables the domain-wise sharing of modulators, substantially enhancing model stability under extreme non-IID RS data scenarios.

The main contributions of this paper are summarized as follows:

- **Structured Heterogeneity Formulation.** We formalize federated RS segmentation under domain shifts as a *structured heterogeneity* problem, explicitly decomposing client-level domain shifts into spectral and spatial shifts. This conceptual decomposition provides a principled basis for class-conditional alignment, moving beyond generic non-IID assumptions.
- **Shift-Aware Prototype Adaptation Framework.** We propose **FedDap**, a prototype-centric federated framework that decouples invariant semantics from domain-conditioned variations. Specifically, clients learn DAPs by transforming globally shared semantic anchors through two complementary modulation branches (spectral and spatial branches), which are dynamically fused via a shift-aware gating mechanism.
- **Extensive Empirical Validation.** We evaluate our method across diverse federated RS benchmarks featuring severe sensor- and region-induced shifts. Our approach consistently achieves state-of-the-art performance, with comprehensive ablations validating the efficacy of our proposed structured shift decomposition, prototype adaptation, and drift-aware fusion.

2 Related work

Federated Semantic Segmentation under Heterogeneity: Federated learning enables collaborative model training across distributed clients without sharing raw data [17]. While extensive research addresses statistical heterogeneity in classification tasks through proximal constraints [12], dynamic regularization [1], local normalization [13], or adaptive heterogeneous collaboration [32], federated semantic segmentation introduces severe pixel-level challenges, such as dense prediction drift and foreground–background imbalance [14, 18]. Recent semantic segmentation studies also show that heterogeneous structures and reliable pixel-level supervision can improve dense prediction under limited supervision [24]. Furthermore, existing clustered or personalized FL methods [3, 11, 23, 31] predominantly treat client heterogeneity as an unstructured distribution mismatch. In contrast, our work recognizes that heterogeneity in remote sensing (RS) is fundamentally structured, motivating a framework that explicitly models these distinct heterogeneous factors rather than treating them as generic noise.

Domain Shift in Remote Sensing Image: RS semantic segmentation natively suffers from strong domain shifts across different sensors and geographic regions. Benchmarks like LoveDA [29] and ISPRS Potsdam/Vaihingen [22] highlight the massive structural and radiometric gaps between urban/rural and multi-sensor datasets. Traditional domain adaptive segmentation methods [4, 8, 26, 39] address these gaps by aligning structured outputs or class-centered features. However, they rely on centralized access to source and target data streams, which blatantly violates FL privacy constraints. Our setting strictly isolates local domains; therefore, we must perform domain adaptation implicitly without ever exchanging cross-domain image.

Prototype-Centric Alignment and Modulation: Sharing lightweight class prototypes serves as an effective privacy-preserving strategy for semantic alignment across isolated domains [6, 25, 35]. Yet, existing prototype-based FL methods typically enforce a single global prototype per class. In RS segmentation, this rigidly conflates diverse spectral and spatial shifts, leading to over-averaged representations. To untangle these factors, we draw inspiration from conditional modulation layers (e.g., FiLM [19], AdaIN [7]). However, rather than performing standard feature-level modulation, our core innovation lies in adapting the prototypes themselves to generate DAPs. By designing specialized spectral and spatial modulation branches and dynamically fusing them via a shift-aware gating mechanism, we tailor global semantic anchors to local RS contexts. To optimize this structured adaptation, we employ a pixel-to-prototype contrastive objective [9, 30] to stably align local dense features with our dynamically fused DAPs.

3 Method

An overview of the proposed framework is illustrated in Fig. 2. In this section, we first formulate federated RS semantic segmentation under *structured* domain

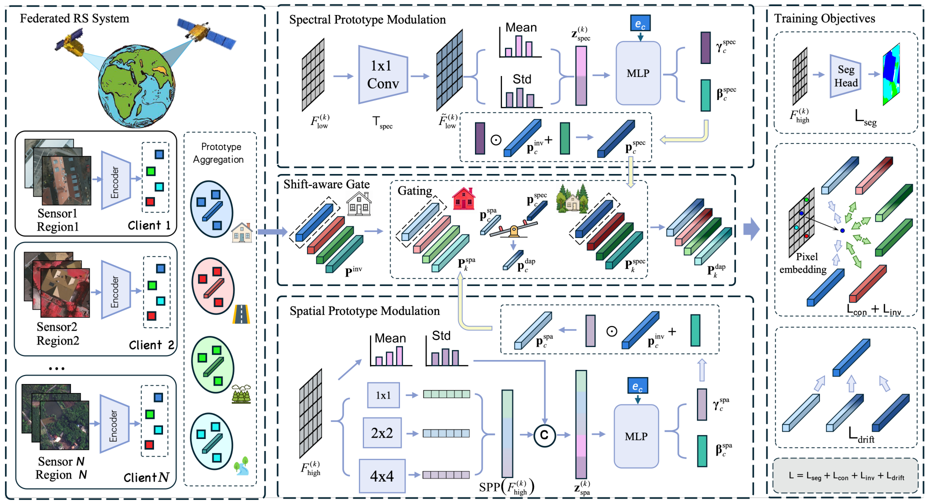

Fig. 2: Overview of the proposed **FedDap** framework.

shifts. We then present a shift-aware prototype modulation module that generates Domain-Adaptive Prototypes (DAPs) from server-side global class prototypes. The spectral and spatial modulation branches are introduced in Sections 3.2 and 3.3, followed by a shift-aware gating mechanism to fuse them into final DAPs (Section 3.4). Finally, we integrate DAPs into federated optimization via pixel-to-prototype contrastive supervision and drift regularization (Section 3.5).

3.1 Problem Formulation

We study federated semantic segmentation for RS image under domain shifts. There are N clients indexed by $k \in \{1, \ldots, N\}$. Client k privately holds a labeled dataset $\mathcal{D}_k = \{(x_i^{(k)}, y_i^{(k)})\}_{i=1}^{n_k}$, where $x_i^{(k)}$ is an RS image acquired under client-specific sensing conditions and $y_i^{(k)} \in \{1, \ldots, C\}^{H \times W}$ is the pixel-wise annotation over C land-cover classes. Raw images remain local and are never shared.

Federated objective: Federated learning aims to optimize a global segmentation model $\mathbf{w}$ by minimizing the weighted sum of local objectives:

$$\min_{\mathbf{w}} \ \mathcal{L}(\mathbf{w}) = \sum_{k=1}^{N} \pi_k \, \mathcal{L}_k(\mathbf{w}), \qquad \mathcal{L}_k(\mathbf{w}) = \mathbb{E}_{(x,y)\in\mathcal{D}_k}\big[\ell(x, y; \mathbf{w})\big], \tag{1}$$

where $\pi_k = \frac{|\mathcal{D}_k|}{|\mathcal{D}|}$ and $\mathcal{D} = \bigcup_{k=1}^{N} \mathcal{D}_k$. At round t, each client updates locally from $\mathbf{w}^{(t)}$ to $\mathbf{w}_k^{(t+1)}$, and the server aggregates via FedAvg [17]:

$$\mathbf{w}^{(t+1)} = \sum_{k=1}^{N} \pi_k \, \mathbf{w}_k^{(t+1)}. \tag{2}$$

RS domain shifts under client heterogeneity: A key challenge in federated RS segmentation is that client data are collected from different sensors and geographic regions, yielding substantial domain shifts. Consequently, samples on different clients follow distinct distributions:

$$(x, y) \sim \mathcal{P}_k, \qquad \mathcal{P}_k \neq \mathcal{P}_j \quad (k \neq j), \tag{3}$$

which leads to inconsistent pixel-level class representations and exacerbates client drift in FL.

Importantly, RS domain shifts are not arbitrary. We observe that they are largely structured by two factors: *spectral shift* induced by sensor/radiometric variations, and *spatial shift* induced by regional variability in texture, morphology, and layout. We therefore explicitly decompose the client domain into two latent shift components:

$$\mathbf{s}^{(k)} = \left(\mathbf{s}^{(k)}_{\text{spec}},\ \mathbf{s}^{(k)}_{\text{spa}}\right), \qquad \mathcal{P}_k = \mathcal{P}\Big(\cdot \mid \mathbf{s}^{(k)}_{\text{spec}}, \mathbf{s}^{(k)}_{\text{spa}}\Big). \tag{4}$$

Our goal is to learn a federated model robust to both components and achieve consistent class-conditional alignment across clients.
Global prototypes as semantic anchors: We maintain a server-side global prototype bank $\mathbf{P}^{\text{inv}} = \{\mathbf{p}^{\text{inv}}_c \in \mathbb{R}^D\}_{c=1}^C$, where each class prototype is defined as the ℓ_2-normalized mean of pixel embeddings and serves as a shared semantic anchor.
Client-side sufficient statistics: On client k, given pixel embeddings $\mathrm{u}_i \in \mathbb{R}^D$ and labels $y_i \in \{1, \ldots, C\}$, we accumulate per-class sufficient statistics (sum and count) over labeled pixels:

$$\mathbf{s}^{(k)}_c = \sum_{i \in \Omega^{(k)}_c} \mathbf{u}_i, \qquad n^{(k)}_c = |\Omega^{(k)}_c|, \qquad \Omega^{(k)}_c = \{i \mid y_i = c\}. \tag{5}$$

Only $\{(\mathbf{s}^{(k)}_c, n^{(k)}_c)\}_{c=1}^C$ are uploaded to the server, avoiding transmission of raw embeddings.
Server-side prototype aggregation: At round t, the server aggregates class prototypes by a count-weighted mean over participating clients $\mathcal{K}_t$:

$$\mathbf{p}^{\text{inv}}_c = \text{Norm}\left(\frac{\sum_{k \in \mathcal{K}_t} \mathbf{s}^{(k)}_c}{\sum_{k \in \mathcal{K}_t} n^{(k)}_c + \epsilon}\right), \tag{6}$$

where $\text{Norm}(\cdot)$ denotes ℓ_2 normalization and ϵ avoids numerical instability. The resulting $\mathbf{P}^{\text{inv}}$ is broadcast to clients and will be locally modulated into domain-adaptive prototypes in Secs. 3.2–3.4.

3.2 Spectral Prototype Modulation

Spectral shift is a primary source of domain discrepancy in federated RS segmentation, arising from sensor diversity and radiometric variations [27]. Such shifts mainly perturb channel-wise appearance statistics and shallow representations. Motivated by this observation, we design a *spectral modulation branch* that adapts the server-side global prototypes $\mathbf{P}^{\text{inv}} = \{\mathbf{p}_c^{\text{inv}}\}_{c=1}^{C}$ to the client-specific spectral condition, producing *spectral-adaptive prototypes* $\mathbf{P}_k^{\text{spec}} = \{\mathbf{p}_c^{\text{spec}}\}_{c=1}^{C}$.

Spectral mapping to a common representation space: Given an input image on client k, the encoder produces a shallow feature map $F_{\text{low}}^{(k)} \in \mathbb{R}^{H_l \times W_l \times D}$. To mitigate sensor-induced spectral discrepancy and align low-level appearance statistics, we first project $F_{\text{low}}^{(k)}$ to a *common spectral representation space* using a lightweight 1×1 convolution:

$$\tilde{F}_{\text{low}}^{(k)} = \mathcal{T}_{\text{spec}}\left(F_{\text{low}}^{(k)}\right), \qquad \mathcal{T}_{\text{spec}} : \mathbb{R}^{H_l \times W_l \times D} \to \mathbb{R}^{H_l \times W_l \times D}. \tag{7}$$

Spectral domain descriptor: We summarize the client-specific spectral condition by channel-wise mean/std over spatial dimensions on $\tilde{F}_{\text{low}}^{(k)}$, and average them over the mini-batch to yield a compact, $2D$-dimensional shift signature:

$$\mathbf{z}_{\text{spec}}^{(k)} = \Big[\ \mathsf{Mean}_{b,h,w}\big(\tilde{F}_{\text{low}}^{(k)}\big), \ \mathsf{Std}_{b,h,w}\big(\tilde{F}_{\text{low}}^{(k)}\big) \ \Big] \in \mathbb{R}^{2D}. \tag{8}$$

FiLM-based prototype modulation: For each semantic class c, we generate class-conditional FiLM parameters $(\boldsymbol{\gamma}_c^{\text{spec}}, \boldsymbol{\beta}_c^{\text{spec}}) \in \mathbb{R}^D \times \mathbb{R}^D$ from the concatenation of the lightweight spectral descriptor $\mathbf{z}_{\text{spec}}^{(k)}$ and a learnable class embedding $\mathbf{e}_c$:

$$(\boldsymbol{\gamma}_c^{\text{spec}}, \boldsymbol{\beta}_c^{\text{spec}}) = \text{MLP}_{\text{spec}}\left([\mathbf{z}_{\text{spec}}^{(k)}, \mathbf{e}_c]\right). \tag{9}$$

We then adapt the global prototype $\mathbf{p}_c^{\text{inv}}$ via channel-wise affine transformation followed by ℓ_2 normalization:

$$\mathbf{p}_c^{\text{spec}} = \text{Norm}\big(\boldsymbol{\gamma}_c^{\text{spec}} \odot \mathbf{p}_c^{\text{inv}} + \boldsymbol{\beta}_c^{\text{spec}}\big), \tag{10}$$

where $\odot$ denotes channel-wise multiplication and $\text{Norm}(\cdot)$ indicates ℓ_2 normalization. The resulting $\mathbf{p}_c^{\text{spec}}$ is a domain-conditioned transformation of $\mathbf{p}_c^{\text{inv}}$, correcting spectral discrepancies while preserving shared semantics. Crucially for federated training, this modulation process is highly efficient and privacy-preserving. The resulting spectral-adaptive prototypes are instantiated locally and never transmitted; the optimization and synchronization of modulators are detailed in Sec. 3.5.

3.3 Spatial Prototype Modulation

Spatial shift in RS image is driven by regional variability in texture, morphology, and scene layout [37], which is more prominently encoded in high-level representations. Accordingly, we design a *spatial modulation branch* that adapts the global prototypes $\mathbf{P}^{\text{inv}} = \{\mathbf{p}_c^{\text{inv}}\}_{c=1}^{C}$ to the client-specific spatial condition, yielding *spatial-adaptive prototypes* $\mathbf{P}_k^{\text{spa}} = \{\mathbf{p}_c^{\text{spa}}\}_{c=1}^{C}$.

Multi-scale spatial descriptor: Given a deep feature map $F_{\text{high}}^{(k)} \in \mathbb{R}^{H_h \times W_h \times D_{\text{in}}}$, we extract a compact spatial shift signature by concatenating multi-scale Spatial

Pyramid Pooling (SPP) with global mean/std of $F_{\text{high}}^{(k)}$, and averaging over the mini-batch:

$$\mathbf{z}_{\text{spa}}^{(k)} = \Big[\,\text{SPP}\big(F_{\text{high}}^{(k)}\big),\ \mathsf{Mean}_{b,h,w}\big(F_{\text{high}}^{(k)}\big),\ \mathsf{Std}_{b,h,w}\big(F_{\text{high}}^{(k)}\big)\, \Big] \in \mathbb{R}^{|\mathcal{S}|D_{\text{in}}+2D_{\text{in}}}, \tag{11}$$

where $\mathcal{S}$ denotes the set of SPP scales.

FiLM-based prototype modulation: Following Sec. 3.2, we generate class-conditional FiLM parameters from $\mathbf{z}_{\text{spa}}^{(k)}$ and a learnable class embedding $\mathbf{e}_c$. To ensure stable modulation during training and prevent extreme feature scaling, we predict a raw scaling factor $\hat{\boldsymbol{\gamma}}_c^{\text{spa}}$ and constrain it using a scaled logistic sigmoid function, i.e., $\boldsymbol{\gamma}_c^{\text{spa}} = 2\,\text{Sigmoid}(\hat{\boldsymbol{\gamma}}_c^{\text{spa}})$. The parameter generation is formulated as:

$$(\boldsymbol{\gamma}_c^{\text{spa}}, \boldsymbol{\beta}_c^{\text{spa}}) = \text{MLP}_{\text{spa}}\Big([\mathbf{z}_{\text{spa}}^{(k)}, \mathbf{e}_c]\Big)\,. \tag{12}$$

We then adapt the global prototype via channel-wise affine transformation and ℓ_2 normalization:

$$\mathbf{p}_c^{\text{spa}} = \text{Norm}\big(\boldsymbol{\gamma}_c^{\text{spa}} \odot \mathbf{p}_c^{\text{inv}} + \boldsymbol{\beta}_c^{\text{spa}}\big)\,. \tag{13}$$

3.4 Shift-aware Gated Fusion for DAPs

The spectral and spatial branches in Secs. 3.2–3.3 capture complementary shift factors, producing $\mathbf{P}_k^{\text{spec}}$ and $\mathbf{P}_k^{\text{spa}}$, respectively. However, the relative importance of spectral versus spatial cues varies across clients and classes. Therefore, we introduce a *shift-aware gating module* to adaptively fuse the two sets of prototypes into the final DAPs for client k: $\mathbf{P}_k^{\text{dap}} = \{\mathbf{p}_c^{\text{dap}}\}_{c=1}^{C}$.

Class-wise shift-aware gate: We predict a class-wise gate vector from compact shift descriptors:

$$\boldsymbol{\alpha}^{(k)} = \text{Sigmoid}\Big(\text{MLP}_{\text{gate}}\Big([\mathbf{z}_{\text{spec}}^{(k)}, \mathbf{z}_{\text{spa}}^{(k)}]\Big)\Big) \in (0,1)^C, \tag{14}$$

where $\alpha_c^{(k)}$ controls the spectral versus spatial contribution for class c on client k.

Gated fusion to obtain DAPs: Given $\mathbf{p}_c^{\text{spec}}$ and $\mathbf{p}_c^{\text{spa}}$, we form the fused prototype as

$$\mathbf{p}_c^{\text{dap}} = \text{Norm}\Big(\alpha_c^{(k)}\,\mathbf{p}_c^{\text{spec}} + (1-\alpha_c^{(k)})\,\mathbf{p}_c^{\text{spa}}\Big)\,, \tag{15}$$

and collect $\mathbf{P}_k^{\text{dap}} = \{\mathbf{p}_c^{\text{dap}}\}_{c=1}^{C}$. The normalization ensures all prototypes lie on a consistent hypersphere for subsequent contrastive learning. The fused DAPs are instantiated locally and never transmitted; the corresponding modulator optimization and synchronization are described in Sec. 3.5.

3.5 Federated Optimization with DAPs

Given the DAPs $\mathbf{P}_k^{\text{dap}} = \{\mathbf{p}_c^{\text{dap}}\}_{c=1}^{C}$ constructed in Sec. 3.4, we integrate them into federated training via pixel-to-prototype contrastive supervision and prototype drift regularization.

Pixel embeddings: Let $\mathbf{u}_i \in \mathbb{R}^D$ be the ℓ_2-normalized embedding of pixel i and y_i its label.
Pixel-to-prototype contrastive supervision: We perform pixel-to-prototype classification over the C prototypes using cosine similarity. Let τ be a temperature. For each labeled pixel i with embedding u_i and label $c = y_i$:

$$\mathcal{L}_{\mathrm{con}}^{(k)} = \frac{1}{|\Omega^{(k)}|} \sum_{i \in \Omega^{(k)}} \mathrm{CE}\left(\frac{\mathbf{u}_i^\top \mathbf{P}^{\mathrm{dap}}}{\tau},\ y_i \right), \tag{16}$$

where $\mathbf{P}^{\mathrm{dap}} \in \mathbb{R}^{C \times D}$ stacks class DAPs and CE is cross-entropy.
Invariant-prototype (anchor) supervision: To further stabilize cross-client semantics, we also classify pixels against the invariant prototype bank $\mathbf{P}^{\mathrm{inv}}$:

$$\mathcal{L}_{\mathrm{inv}}^{(k)} = \frac{1}{|\Omega^{(k)}|} \sum_{i \in \Omega^{(k)}} \mathrm{CE}\left(\frac{\mathbf{u}_i^\top \mathbf{P}^{\mathrm{inv}}}{\tau},\ y_i \right). \tag{17}$$

Prototype drift regularization: To prevent uncontrolled deformation, we penalize cosine drift of modulated prototypes from the global anchors:

$$\mathcal{L}_{\mathrm{drift}}^{(k)} = \frac{1}{C} \sum_{c=1}^{C} \Big(1 - \cos(\mathbf{p}_c^{\mathrm{spec}}, \mathbf{p}_c^{\mathrm{inv}})\Big) + \Big(1 - \cos(\mathbf{p}_c^{\mathrm{spa}}, \mathbf{p}_c^{\mathrm{inv}})\Big) + \eta \Big(1 - \cos(\mathbf{p}_c^{\mathrm{dap}}, \mathbf{p}_c^{\mathrm{inv}})\Big), \tag{18}$$

with η controlling the strength on the fused DAPs.
Overall local objective: The final training objective on client k combines the standard segmentation loss (e.g., pixel-wise cross-entropy or cross-entropy + Dice) with the two additional terms:

$$\mathcal{L}_k(\mathbf{w}) = \mathcal{L}_{\mathrm{seg}}^{(k)}(\mathbf{w}) + \lambda_{\mathrm{dap}} \mathcal{L}_{\mathrm{con}}^{(k)} + \lambda_{\mathrm{inv}} \mathcal{L}_{\mathrm{inv}}^{(k)} + \lambda_{\mathrm{drift}} \mathcal{L}_{\mathrm{drift}}^{(k)}. \tag{19}$$

Federated training: In each communication round, the server broadcasts the current model $\mathbf{w}^{(t)}$ and global prototypes $\mathbf{P}^{\mathrm{inv}}$. Each client computes $\mathbf{P}_k^{\mathrm{dap}}$ using Secs. 3.2–3.4 and optimizes Eq. (19) locally to obtain $\mathbf{w}_k^{(t+1)}$. The server aggregates client models via FedAvg (Eq. (2)) and updates global prototypes using uploaded local prototype statistics (Eq. (6)). Importantly, DAPs $\mathbf{P}_k^{\mathrm{dap}}$ are instantiated on-device and are never transmitted.
Optimization and cluster-wise synchronization of modulators. The spectral and spatial modulators, including $\mathcal{T}_{\mathrm{spec}}$, $\mathcal{T}_{\mathrm{spa}}$, $\mathrm{MLP}_{\mathrm{spec}}$, and $\mathrm{MLP}_{\mathrm{spa}}$, are optimized locally with the segmentation network using Eq. (19). Each client maintains EMA-smoothed shift signatures $(\bar{\mathbf{z}}_{\mathrm{spec}}^{(k)}, \bar{\mathbf{z}}_{\mathrm{spa}}^{(k)})$ and uploads only the modulator parameters, compact signatures, and class-wise prototype statistics; raw images, dense feature maps, pixel embeddings, and local DAPs are never shared. Every R rounds, the server clusters clients according to the uploaded signatures, aggregates spectral/spatial modulators within the corresponding clusters in a FedAvg-style manner, and averages the gating network globally. The aggregated modulators are then broadcast back according to cluster assignments, allowing clients with similar shift patterns to share modulation knowledge.

4 Experiments

4.1 Experimental Settings

Datasets To evaluate our approach in federated RS semantic segmentation under structured domain shifts, we construct federated learning scenarios using three established RS datasets: LoveDA, Potsdam, and Vaihingen.

LoveDA [29]: A multi-domain RS semantic segmentation dataset consisting of 5987 high-resolution images (1024×1024 pixels) from three cities in China (Nanjing, Changzhou, and Wuhan). The dataset naturally exhibits urban-rural domain shifts due to differences in scene layout, texture, and land-use patterns, making it suitable for simulating spatial shifts in federated settings.

Potsdam [22]: A high-resolution aerial image dataset from Potsdam, Germany, with 38 images (6000×6000 pixels, 5 cm GSD) in IRRG (infrared-red-green) format. Its annotations cover 6 classes: impervious surfaces (Surf), building (Bldg), low vegetation (Vegt), tree, car, and clutter/background. We exclude clutter for evaluation, focusing on 5 foreground classes.

Vaihingen [22]: This dataset contains 33 aerial images (average 2494×2064 pixels, 9 cm GSD) from Vaihingen, Germany, available in both IRRG and RGB formats. It shares the same 6 classes as Potsdam. To align the spatial resolution with Potsdam for a controlled comparison, all Vaihingen images are resampled from 9 cm to 5 cm GSD prior to training.

All images are preprocessed to standard resolutions and normalized. We split each dataset into train/val/test sets following standard protocols.

Using LoveDA, Potsdam, and Vaihingen, we construct three main federated learning scenarios to simulate structured domain shifts in RS segmentation. Each scenario contains $N = 10$ clients divided into two domain groups with five clients per group: (i) LoveDA urban vs. rural, which mainly reflects region-induced spatial shifts; (ii) Potsdam(IRRG) vs. Vaihingen(IRRG), which evaluates spatial shifts across different geographic regions under the same spectral modality; and (iii) Potsdam(IRRG) vs. Vaihingen(RGB), which introduces compound spectral and spatial shifts. Local datasets are balanced in size but heterogeneous in domain. Following the experimental protocol of FedSeg [18], we conduct $T = 150$ communication rounds with $E = 2$ local epochs per round. Unless otherwise specified, we adopt a ResNet-50 [5] encoder as the shared backbone for all compared methods to ensure a fair comparison.

In addition to the above main comparisons, we conduct two complementary validation experiments. First, we construct a Potsdam RGB vs. IRRG setting to isolate spectral shifts while minimizing geographic variation. Second, we evaluate a more challenging cross-dataset setting, LoveDA-RGB + Potsdam-IRRG, using SegFormer-B0 [33] to verify whether FedDap remains effective with a modern transformer-based backbone and under stronger spectral–spatial heterogeneity.

4.2 Main Results

We compare FedDap with representative federated learning baselines, including FedAvg [17], MOON [10], FedProto [25], FedProx [12], FedDyn [1], FedOpt [20],

Table 1: Comparison on LoveDA under the Urban/Rural shift. mIoU is averaged over 7 classes.

Method	mIoU	Acc	Bg	Bldg	Road	Water	Barren	Forest	Agri
FedAvg	49.89	68.40	52.37	60.27	46.95	67.29	32.21	39.07	51.06
MOON	49.45	68.02	51.50	61.25	45.81	65.66	31.35	38.89	51.66
FedProto	49.25	67.40	50.52	59.93	47.72	66.27	32.13	37.53	50.66
FedProx	49.68	68.91	53.02	61.75	45.28	67.64	27.92	39.42	52.70
FedDyn	45.75	64.12	50.26	52.90	45.65	66.07	31.19	33.05	41.13
FedOpt	48.38	68.68	50.67	60.89	40.72	65.31	19.71	41.67	59.69
FPL	49.85	68.49	52.24	62.62	47.04	63.91	28.89	38.68	54.18
FedSeg	50.04	**69.90**	53.58	60.54	46.68	66.50	26.74	40.46	55.81
FedDap (Ours)	**50.68**	69.52	53.34	62.23	47.66	65.53	29.89	41.84	54.30

FPL [6], and FedSeg [18], under the three structured shift scenarios described in Sec. 4.1. We report pixel accuracy (Acc) and mean IoU (mIoU), together with per-class IoU. For Potsdam→Vaihingen, mIoU is computed over the 5 foreground classes (Surf, Bldg, Vegt, Tree, Car), excluding clutter/background.

LoveDA (Urban-Rural spatial shift). Table 1 shows the results under the urban–rural (primarily spatial) shift. Our FedDap achieves the best mIoU of **50.68**, outperforming the strongest baseline FedSeg (50.04) by +**0.64 mIoU points**. Notably, FedDap improves several shift-sensitive categories, including *Bldg* (+1.69 IoU points), *Road* (+0.98 IoU points), *Barren* (+3.15 IoU points), and *Forest* (+1.38 IoU points) compared with FedSeg, indicating better cross-client class alignment under urban/rural layout differences. While FedSeg attains the highest overall Acc, FedDap yields the best mIoU, suggesting our method improves class-balanced performance rather than being dominated by frequent classes.

Potsdam(IRRG) vs. Vaihingen(IRRG) (spatial shift). Under spatial shift with consistent IRRG modality (Table 2), FedDap reaches the best mIoU of **73.43** and the best Acc of **84.41**. Compared with the second-best method FPL (73.08), FedDap improves mIoU by +**0.35** mIoU points, with clear gains on *Bldg* (84.02 IoU) and competitive performance across the remaining classes. These results confirm that modeling region-dependent spatial variability in the prototype space improves robustness even when spectral conditions are aligned.

Potsdam(IRRG) vs. Vaihingen(RGB) (spectral + spatial shift). The IRRG→RGB setting introduces the most challenging compound shift (spatial + spectral). As shown in Table 2, FedDap achieves a mIoU of **70.42**, substantially surpassing the best baseline FedSeg (68.82) by +**1.60** mIoU points and FedAvg (67.17) by +**3.25** mIoU points. Per-class IoU indicates that the largest gains occur on *Bldg* (+3.48 IoU points) and *Vegt* (+4.11 IoU points) compared to FedSeg, which are particularly sensitive to cross-sensor appearance changes and region-specific texture/morphology. Overall, the performance gap widens as the shift becomes more complex, supporting our motivation to explicitly decompose RS client heterogeneity into spectral and spatial components and to perform shift-aware prototype adaptation.

Table 2: Comparison with per-class IoU on Potsdam→Vaihingen benchmarks. mIoU is averaged over 5 foreground classes, including Surf, Bldg, Vegt, Tree, and Car. Best and second-best results are highlighted in each column.

Method	mIoU	Acc	Surf	Bldg	Vegt	Tree	Car
Potsdam(IRRG) vs. Vaihingen(IRRG) (Spatial shift).							
FedAvg	70.64	83.02	75.13	79.34	62.98	70.03	65.72
MOON	70.29	82.95	74.99	79.48	63.06	69.70	64.22
FedProto	70.46	83.12	75.07	79.55	63.38	70.20	64.10
FedProx	70.79	83.04	75.18	79.86	63.25	69.50	66.15
FedDyn	71.32	83.20	75.09	79.54	62.96	71.34	67.67
FedOpt	72.17	84.05	77.32	83.78	63.65	69.15	66.93
FPL	<u>73.08</u>	<u>84.16</u>	76.24	81.93	64.30	72.17	70.74
FedSeg	71.06	83.37	75.67	81.21	63.55	69.33	65.52
FedDap (Ours)	**73.43**	**84.41**	76.71	84.02	64.38	71.07	70.96
Potsdam(IRRG) vs. Vaihingen(RGB) (Spectral + Spatial shift).							
FedAvg	67.17	79.88	70.52	74.61	53.32	69.63	67.77
MOON	67.17	80.95	71.25	76.65	58.50	69.31	60.13
FedProto	69.71	80.99	71.28	77.56	58.45	68.43	65.81
FedProx	67.99	80.51	70.23	76.04	56.73	69.73	67.23
FedDyn	65.78	79.09	69.60	69.75	53.96	70.80	64.79
FedOpt	65.77	80.62	70.57	78.43	55.44	55.44	68.97
FPL	68.52	80.85	69.36	77.54	57.99	70.78	66.95
FedSeg	<u>68.82</u>	<u>81.42</u>	72.52	77.01	57.64	70.43	66.51
FedDap (Ours)	**70.42**	**82.84**	73.54	80.49	61.75	71.11	65.21

Additional validation. Table 3 further evaluates FedDap from two complementary perspectives. First, under the isolated spectral-shift setting of Potsdam RGB vs. IRRG, FedDap achieves 75.80 mIoU and 84.86 Acc, outperforming FedAvg and FedSeg. This result verifies that the spectral modulation branch is effective even when the spatial/geographic factor is largely controlled. Second, under the more challenging LoveDA-RGB + Potsdam-IRRG setting with SegFormer-B0, FedDap obtains the best global mIoU of 39.25, improving over FedSeg by 1.21 points, and also achieves the best domain-wise mIoU on both LoveDA and Potsdam. Although the cross-dataset gap causes an overall performance drop, FedDap remains more robust than FL baselines and also outperforms isolated local training on both domains, suggesting that federated sharing is still beneficial under severe but structured spectral–spatial heterogeneity.

4.3 Visualization Analysis

To provide intuitive evidence beyond quantitative metrics, we visualize (i) the learned feature embeddings and (ii) qualitative segmentation predictions under structured domain shifts.

Embedding space visualization. Fig. 3 presents joint t-SNE projections [15] of pixel embeddings learned by FedAvg, FedSeg, and our method, with light/dark shades marking the two domains. Compared with baselines, our method yields

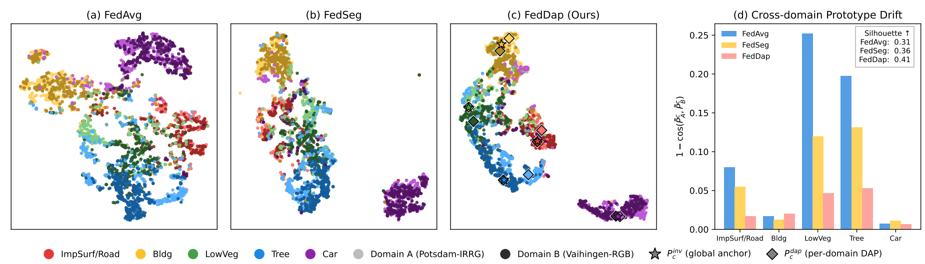

Fig. 3: t-SNE visualization of learned embeddings (FedAvg / FedSeg / FedDap) on the spectral+spatial shift scenario. Colors encode semantic classes; *light/dark* shades distinguish Domain A (Potsdam-IRRG) and Domain B (Vaihingen-RGB). On (c) we overlay global prototypes $\mathbf{p}_c^{\text{inv}}$ ($\star$) and per-domain DAPs $\mathbf{p}_c^{\text{dap}}$ ($\Diamond$). Panel (d) reports per-class cross-domain drift $1 - \cos(\bar{\mathbf{p}}_A^c, \bar{\mathbf{p}}_B^c)$ with the silhouette inset.

Table 3: Additional validation under isolated spectral shift and cross-dataset spectral–spatial shift. All mIoU and Acc values are reported in percentage.

Potsdam RGB vs. IRRG, isolated spectral shift, ResNet-50.

Method	Acc	mIoU
FedAvg	84.17	73.71
FedSeg	84.30	74.70
FedDap (Ours)	**84.86**	**75.80**

LoveDA-RGB + Potsdam-IRRG, cross-dataset spectral–spatial shift, SegFormer-B0.

Method	Global Acc	Global mIoU	$\text{mIoU}_{\text{LoveDA}}$	$\text{mIoU}_{\text{Potsdam}}$
FedAvg	59.34	36.47	30.44	49.34
FedSeg	**61.92**	38.04	32.52	47.86
FedDap (Ours)	61.44	**39.25**	**33.54**	**51.88**
Local-only	–	–	24.07	50.41

more compact class clusters with light-shade (Domain A) and dark-shade (Domain B) pixels tightly intermixed, while the prototypes visualize how each DAP deviates slightly from $\mathbf{p}_c^{\text{inv}}$ toward its own domain mass (Sec. 3.4). Quantitatively, FedDap raises the cosine silhouette from 0.31/0.36 (FedAvg/FedSeg) to **0.41**.

Qualitative segmentation results. Fig. 4 visualizes representative predictions on Potsdam and Vaihingen samples. FedAvg tends to produce noisier masks with fragmented regions and inconsistent boundaries across domains, while FedSeg partially alleviates such errors. In contrast, our method generates cleaner segmentation maps with sharper object boundaries (e.g., buildings/impervious surfaces) and fewer spurious predictions in confusing areas, indicating improved robustness to spatial and spectral variations.

4.4 Ablation Study

We ablate core components of FedDap under the same backbone and federated schedule as Sec. 4.1. Table 4 reports mIoU (%) on the three scenarios.

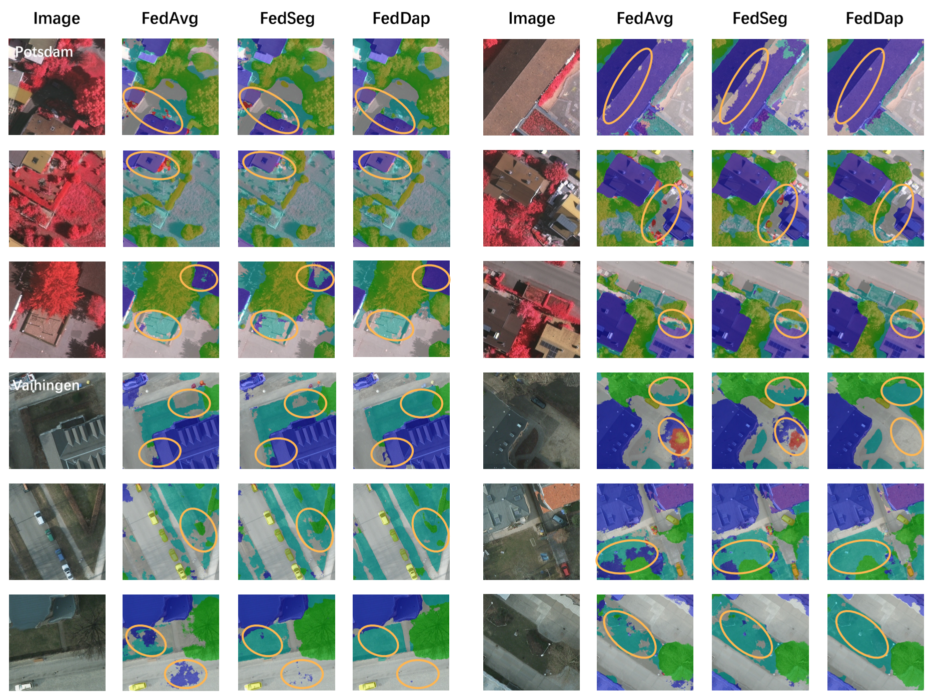

Fig. 4: Qualitative comparisons on Potsdam and Vaihingen samples. From left to right: input image, FedAvg, FedSeg, and our method. Our method produces cleaner masks with fewer artifacts and more consistent boundaries under domain shifts.

Settings and variants. We remove (i) the invariant-prototype supervision (**w/o Invariant Prototype**), (ii) the DAP-based supervision (**w/o Dynamic Prototype**), and (iii) one modulation branch (**Spectral/Spatial Branch Only**). We also compare fusion strategies by replacing the learned gate with a fixed α=0.5 (**Fixed Gate**) or using cascaded (serial) fusion (**Cascaded Fusion**).

Invariant prototypes and DAPs both matter. Removing invariant-prototype supervision consistently degrades performance, with the largest drop on the hardest shift P(IRRG) vs. V(RGB) (70.42→68.35, −2.07), confirming $\mathbf{P}^{\text{inv}}$ is a cru-

Table 4: Ablation study of FedDap. All results are reported as mIoU (%). Best and second-best results are highlighted in each column.

Variant	LoveDA	P(IRRG) vs. V(IRRG)	P(IRRG) vs. V(RGB)
FedDap (Full)	**50.68**	**73.43**	**70.42**
w/o Invariant Prototype	50.38	72.10	68.35
w/o Dynamic Prototype	50.43	72.44	69.05
Spectral Branch Only	50.26	72.18	69.41
Spatial Branch Only	50.34	<u>73.31</u>	69.13
Fixed Gate ($g = 0.5$)	49.83	73.20	69.25
Cascaded Fusion (Serial)	<u>50.43</u>	69.75	<u>69.81</u>

cial semantic anchor. Disabling DAP-based supervision also hurts (70.42→69.05, −1.37), verifying the benefit of shift-conditioned prototypes.

Two branches are complementary and gated fusion is most robust. On spatial-only shift P(IRRG) vs. V(IRRG), the spatial-only variant is close to full (73.31 vs. 73.43), while spectral-only is worse (72.18), matching the expected shift structure. On compound shift P(IRRG) vs. V(RGB), both single-branch variants lag behind full, indicating the need to model both factors jointly. Learned gated fusion is more reliable than fixed or cascaded fusion, especially under spatial shift (73.43→69.75 for cascaded), suggesting sequential modulation can accumulate deformation.

5 Conclusion

We presented **FedDap**, a federated remote sensing semantic segmentation framework designed for the *structured* heterogeneity of RS image. FedDap maintains server-side invariant class prototypes as shared semantic anchors, and derives *Domain-Adaptive Prototypes (DAPs)* via two complementary modulation branches that respectively capture spectral (sensor-induced) and spatial (region-induced) shifts. A shift-aware gating module fuses these branches into class-conditional DAPs, and federated optimization is guided by pixel-to-prototype supervision against both invariant anchors and DAPs, together with a drift regularizer to stabilize prototype deformation. Extensive experiments on LoveDA and ISPRS Potsdam/Vaihingen benchmarks demonstrate consistent improvements over strong federated baselines, especially under compound spatial and spectral shifts, and ablations validate the necessity of invariant anchors, dynamic prototypes, and gated fusion.

This work suggests that explicitly modeling RS-specific heterogeneity is critical for robust federated dense prediction. In future work, we will explore extending the structured shift modeling to more diverse sensors and resolutions, reducing dependence on labeled supervision, and improving scalability to more than two latent domains and highly imbalanced client participation.

Acknowledgements

This work was supported by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (No. JYB2025XDXM101), the New Cornerstone Science Foundation through the XPLORER PRIZE, the Innovative Research Group Project of Hubei Province under Grants 2024AFA017, National Natural Science Foundation of China under Grants (623B2079,623B2080), and NTU AI-for-X Postdoctoral Fellowship.

References

1. Acar, D.A.E., Zhao, Y., Navarro, R.M., Mattina, M., Whatmough, P.N., Saligrama, V.: Federated learning based on dynamic regularization. arXiv preprint arXiv:2111.04263 (2021) 4, 10
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PmLR (2020) 3
3. Ghosh, A., Chung, J., Yin, D., Ramchandran, K.: An efficient framework for clustered federated learning. Advances in neural information processing systems **33**, 19586–19597 (2020) 4
4. Gong, Z., Wei, Z., Wang, D., Hu, X., Ma, X., Chen, H., Jia, Y., Deng, Y., Ji, Z., Zhu, X., et al.: Crossearth: Geospatial vision foundation model for domain generalizable remote sensing semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025) 4
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 10
6. Huang, W., Ye, M., Shi, Z., Li, H., Du, B.: Rethinking federated learning with domain shift: A prototype view. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16312–16322. IEEE (2023) 4, 11
7. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE international conference on computer vision. pp. 1501–1510 (2017) 4
8. Jiang, Z., Li, Y., Yang, C., Gao, P., Wang, Y., Tai, Y., Wang, C.: Prototypical contrast adaptation for domain adaptive semantic segmentation. In: European conference on computer vision. pp. 36–54. Springer (2022) 4
9. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems **33**, 18661–18673 (2020) 4
10. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10713–10722 (2021) 10
11. Li, T., Hu, S., Beirami, A., Smith, V.: Ditto: Fair and robust federated learning through personalization. In: International conference on machine learning. pp. 6357–6368. PMLR (2021) 4
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) 4, 10
13. Li, X., Jiang, M., Zhang, X., Kamp, M., Dou, Q.: Fedbn: Federated learning on non-iid features via local batch normalization. arXiv preprint arXiv:2102.07623 (2021) 4
14. Liu, Q., Chen, C., Qin, J., Dou, Q., Heng, P.A.: Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1013–1023 (2021) 4
15. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008) 12
16. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P.: High-resolution aerial image labeling with convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing **55**(12), 7092–7103 (2017) 2

17. 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) 4, 5, 10
18. Miao, J., Yang, Z., Fan, L., Yang, Y.: Fedseg: Class-heterogeneous federated learning for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8042–8052 (2023) 4, 10, 11
19. 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) 4
20. Reddi, S., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečnỳ, J., Kumar, S., McMahan, H.B.: Adaptive federated optimization. arXiv preprint arXiv:2003.00295 (2020) 10
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015) 2
22. Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U.: The isprs benchmark on urban object classification and 3d building reconstruction (2012) 4, 10
23. Sattler, F., Müller, K.R., Samek, W.: Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. IEEE transactions on neural networks and learning systems **32**(8), 3710–3722 (2020) 4
24. Shan, Y., Zhuge, Y., Lu, H.: HDVS: Semi-supervised semantic segmentation via heterogeneous dual-branch voting supervision. Visual Intelligence **4**(1), 4 (2026). https://doi.org/10.1007/S44267-026-00107-3, article number: 4 4
25. Tan, Y., Long, G., Liu, L., Zhou, T., Lu, Q., Jiang, J., Zhang, C.: Fedproto: Federated prototype learning across heterogeneous clients. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 8432–8440 (2022) 2, 4, 10
26. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7472–7481 (2018) 4
27. Tuia, D., Persello, C., Bruzzone, L.: Domain adaptation for the classification of remote sensing data: An overview of recent advances. IEEE geoscience and remote sensing magazine **4**(2), 41–57 (2016) 6
28. Wang, A., Yang, L., Wu, H., Iwahori, Y.: Heterogeneous defect prediction based on federated prototype learning. IEEE Access **11**, 98618–98632 (2023) 2
29. Wang, J., Zheng, Z., Ma, A., Lu, X., Zhong, Y.: Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation. arXiv preprint arXiv:2110.08733 (2021) 4, 10
30. Wang, X., Zhang, R., Shen, C., Kong, T., Li, L.: Dense contrastive learning for self-supervised visual pre-training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3024–3033 (2021) 4
31. Wang, Z., Wang, Z., Fan, X., Wang, C.: Federated learning with domain shift eraser. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 4978–4987 (2025) 4
32. Wenke, H., Mang, Y., Bo, D.: Adaptive heterogeneous federated learning. Journal of Image and Graphics **29**(7), 1849–1860 (2024). https://doi.org/10.11834/jig.230239 4
33. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems **34**, 12077–12090 (2021) 10

34. Zhang, J., Liu, Y., Hua, Y., Cao, J.: Fedtgp: Trainable global prototypes with adaptive-margin-enhanced contrastive learning for data and model heterogeneity in federated learning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 16768–16776 (2024) 2
35. Zhang, X., Zhang, B., Yu, W., Kang, X.: Federated deep learning with prototype matching for object extraction from very-high-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–16 (2023) 4
36. Zhou, Y., Qu, X., You, C., Zhou, J., Tang, J., Zheng, X., Cai, C., Wu, Y.: Fedsa: A unified representation learning via semantic anchors for prototype-based federated learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 23009–23017 (2025) 2
37. Zhu, M., He, Y., He, Q.: A review of researches on deep learning in remote sensing application. International journal of geosciences **10**(1), 1–11 (2019) 7
38. Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F.: Deep learning in remote sensing: A comprehensive review and list of resources. IEEE geoscience and remote sensing magazine **5**(4), 8–36 (2017) 2
39. Zou, X., Li, Y., Zhang, S., Li, K., Wang, S., Tao, P., Xing, J., Lang, C.: Dynamic dictionary learning for remote sensing image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 22457–22466 (2025) 4