Context-Aware Joint Alignment for Cross-Scene Hyperspectral Image Classification

Boshan Shi[1], Jiaxin Chen[1], Yanbo Liu[2], Youqiang Zhang[3], and Guo Cao[1⋆]

[1] Nanjing University of Science and Technology , Nanjing 210094, China
{shiboshan,jiaxinchen,caoguo}@njust.edu.cn
[2] Shandong University of Science and Technology, Qingdao 266590, China
liuyanbo@njust.edu.cn
[3] Nanjing University of Posts and Telecommunications, Nanjing 210023, China
zhangyq@njupt.edu.cn
https://github.com/BoshanCj/CAJA-ECCV26

Abstract. Cross-scene hyperspectral image (HSI) classification faces persistent challenges due to domain shifts in spectral signatures, spatial structures, and semantic compositions. Existing domain generalization methods typically align either features or predictions, but they overlook how local context (e.g., boundary mixing and class co-occurrence patterns) changes across scenes. As a result, one-sided alignment often yields partial adaptation: feature-invariant models may still produce unstable outputs, while output-aligned models can rely on domain-specific shortcuts. We propose **Context-Aware Joint Alignment** (**CAJA**), a pixel-faithful framework that preserves the single-label objective while injecting explicit context into training. CAJA combines **Context-Aware Supervision** (**CAS**) to improve boundary-aware supervision using neighborhood context and homogeneity, and **Context-Conditioned Joint Alignment** (**CCJA**) to enforce context-conditioned alignment in both feature and prediction spaces across domains. By coupling supervision and joint alignment under shared context, CAJA mitigates the partial-alignment limitation of one-sided methods. Extensive experiments on three cross-scene HSI benchmarks show that CAJA consistently improves generalization, especially at class boundaries where context shift is most severe, establishing a simple, principled, and deployment-friendly solution for cross-scene HSI domain generalization.

Keywords: Cross-scene hyperspectral image classification · Domain generalization · Context-aware supervision

1 Introduction

Cross-scene hyperspectral image (HSI) classification is challenged not only by spectral distribution shift but, more critically, by *context shift*: when boundary

⋆ Corresponding author

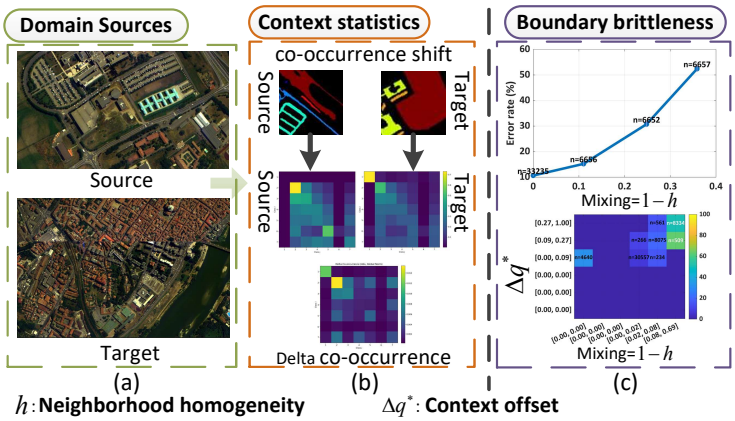

Fig. 1: Co-occurrence shift makes boundaries brittle. (a) The Pseudo-color images from two HSI datasets. (b) The neighborhood co-occurrence of the same class shifts across scenes. (c) Error rates rise sharply at low homogeneity ($h\downarrow$) and increase with co-occurrence divergence $\Delta q^\star$, indicating that boundary/mixed regions are more fragile under context shift. CAJA leverages the same context $C = (q^\star, h)$ for supervision and alignment, mitigating this brittleness.

mixing is prevalent and class co-occurrence patterns vary across scenes, local decision rules learned on the source domain break on the target [1, 28]. In the absence of target supervision, global alignment or one-sided alignment (features-only or predictions-only) cannot restore such local consistency [34, 36], so the same class yields unstable outputs under similar neighborhoods across scenes. Empirically, boundary pixels account for about 9.31% of samples on benchmark Houston, and their cross-scene error rises by 20.28% relative to interior pixels. In Fig. 1, we further demonstrate the fragility of boundaries on Pavia benchmark by analyzing both the degree of boundary mixing and the strength of context shift. This underscores the need to enforce *stability conditioned on context*: given the same neighborhood context, both representations and decisions should remain consistent.

Existing domain generalization (DG) methods typically align either *features* (e.g., statistics matching via MMD/CORAL, adversarial objectives) or *predictions* (e.g., divergence-based consistency) [10, 22, 31]. While helpful, these one-sided strategies induce a *partial-alignment pathology*: feature-invariant representations can still yield unstable decisions under context shift, whereas output-aligned models may rely on domain-specific shortcuts and ignore mismatched representations [33]. In HSI, this pathology is exacerbated at class boundaries and in regions where land-cover composition changes across scenes [39].

We argue that *local context*—captured by neighborhood composition and homogeneity—is the missing conditioning variable in cross-scene HSI. Generalization should be enforced *conditioned on context* rather than globally. Concretely, we aim to learn a *context-conditioned invariant mapping*: given the

same context across scenes, both representation geometry and induced decisions should agree.

We propose **CAJA** (Context-Aware Joint Alignment), a pixel-faithful framework that preserves the single-label objective while injecting explicit context into training and enforcing joint alignment. First, **Context-Aware Supervision (CAS)** converts neighborhood labels into a histogram and a homogeneity score; it applies homogeneity-adaptive label smoothing to temper over-confidence near boundaries and imposes center–context consistency so that predictions remain compatible with the observed local composition. Second, **Context-Conditioned Joint Alignment (CCJA)** is cast as *context-conditioned invariance*: conditioned on a shared context variable C, CCJA learns a domain-invariant mapping by requiring both the *conditional feature moments* and the *conditional output distributions* to coincide across domains, addressing the partial-alignment pathology of one-sided methods [34, 36]. CAJA requires no target-domain data and no change to the pixel-wise inference protocol. We detail the construction of the neighborhood histogram, homogeneity, and context buckets in Sec. 3.

Unlike patch-based HSI classifiers that implicitly assume single-class patches and label the whole patch by the center pixel [3], *CAJA* turns neighborhood labels into a *histogram* $q^\star$ and a *homogeneity* score h and uses them as *continuous supervisory signals* during training, rather than as post-hoc smoothing [45]. This preserves the pixel-wise single-label objective while injecting context: h governs homogeneity-adaptive label smoothing and $q^\star$ anchors center–context consistency. Moreover, *CCJA* performs *context-conditioned joint alignment* by grouping samples into *context buckets* (from $q^\star, h$) and, within each bucket, aligning both representation statistics and predictive distributions. In short, we do not merely combine two losses; we learn a *domain-invariant decision function conditioned on context*. In summary, our contributions are threefold:

- We formulate *context-conditioned invariance* for cross-scene HSI, lifting joint alignment from a juxtaposition of losses to two projections of a single objective (representation and decision) while preserving the pixel-wise single-label task.
- We instantiate this through CAS, which uses neighborhood statistics $(q^\star, h)$ as *continuous supervision* (homogeneity-adaptive smoothing and center–context consistency), and CCJA, which enforces *context-bucketed* agreement of conditional feature moments and conditional output distributions. No target data are used and the inference protocol is unchanged.
- CAJA delivers consistent improvements over strong DG baselines across three cross-scene benchmarks, with markedly improved stability on boundary/mixed regions; ablations and bucket visualizations confirm the complementarity and necessity of CAS and CCJA. Training overhead is modest and there is *no inference-time overhead*.

2 Related Work

2.1 DG-based cross-scene HSI classification

A common line of work trains on a single source domain and enlarges source diversity via adversarial or style perturbations; early efforts adopt AdaIN-style transfer, and later variants use frequency-domain statistics or phase–amplitude perturbations [2,27,29,32,38,40]. Models typically follow the center-pixel single-label protocol while feeding a local spectral–spatial cube to CNNs or transformers [14, 15]. Cross-scene deployment, however, introduces spectral, geometric, and semantic-composition shifts; boundary mixing and scene-dependent class co-occurrence can weaken local decision rules learned on the source domain [8,43]. Many DG approaches align distributions in the representation space (e.g., MMD, CORAL, adversarial/domain-invariant learning) or regularize in the output space (e.g., KL/JS-based consistency) [4,6,11,12,34,36,46]. In HSI with complex co-occurrence structure, one-sided alignment is prone to partial adaptation where representations and decisions mismatch, and global or class-conditional alignment rarely models context shift explicitly. In contrast, we preserve the pixel-wise single-label protocol, convert neighborhood statistics into training-time signals, and enforce joint alignment conditioned on context; details appear in Section 3.

2.2 Context and co-occurrence modeling in HSI

Spatial consistency in remote sensing and HSI is often encouraged via CRFs or graph models, superpixel-based smoothing, or segmentation-style supervision [7, 16, 26, 35]. Other lines treat patches as multi-label samples or pursue mixed-pixel unmixing and reconstruction formulations [25, 30]. These paradigms either rely on dense labels and change the inference protocol, or they operate as post hoc smoothing. They typically do not expose neighborhood statistics as training-time signals. In contrast, we preserve the standard pixel-wise single-label protocol and convert neighborhood label histograms and homogeneity into continuous supervision, thereby capturing boundary uncertainty and scene-dependent co-occurrence during training without target-domain data (see Section 3).

2.3 Joint/conditional alignment

Prior work on joint or conditional alignment mainly targets global or class-conditional statistics [5, 17, 20, 21], leaving context variation under-modeled in cross-scene settings. We instead propose *context-conditioned joint alignment*: given a shared context variable, we match conditional feature moments across domains (via CORAL) and align conditional output distributions with a Jensen–Shannon–type divergence [9,34]. This treats joint alignment as *context-conditioned invariance* rather than a loose sum of regularizers, yielding decisions that remain stable and context-faithful under cross-scene shift (see Section 3).

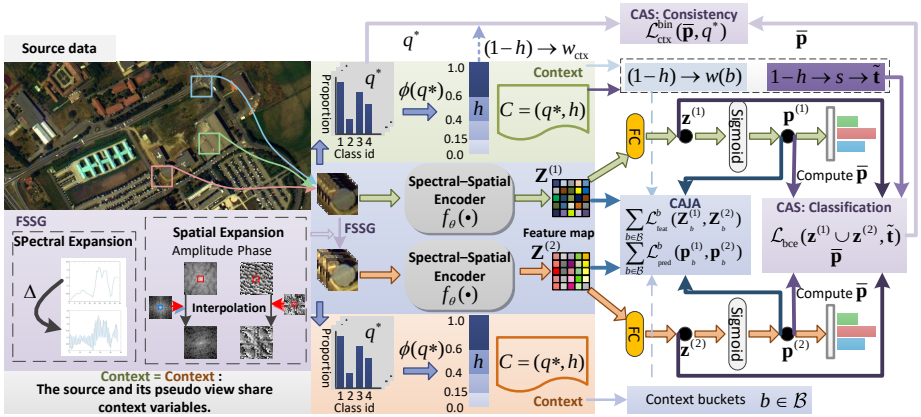

Fig. 2: Overview of CAJA. Under SSDG, we first use FSSG to generate semantics-preserving pseudo-source views that inject style/statistical perturbations without any target-domain cues. From each center-anchored patch we extract neighborhood co-occurrence $q^\star$ and homogeneity h as train-time context $C = (q^\star, h)$. **CAS** injects C into supervision via BCE with adaptive smoothing and a center–context Bernoulli KL, weighted by $w_{\text{ctx}} = \gamma(1-h)$. **CCJA** partitions samples into context buckets and performs joint per-bucket alignment across real/pseudo views (CORAL on features, Bernoulli KL on predictions) with risk weighting $\omega(b)$. The single-label, per-pixel inference protocol remains unchanged.

3 Method

Given the limited availability of labeled HSI datasets, we adopt the *Single-Source Domain Generalization* (SSDG) setting: training uses one labeled source domain and no target data. We formalize SSDG and its cross-scene challenges in Sec. 3.1, and then introduce *Context-Aware Joint Alignment* (CAJA) in Secs. 3.2–3.3. An overview is shown in Fig. 2.

3.1 Single-source domain generalization

Preliminaries. We study cross-scene HSI classification under the *single-source domain generalization* (SSDG) setting [46]: a single labeled source domain (SD) $\mathcal{S} = \{(\mathrm{X}, \mathrm{y})\}$ is available for training, while target domains $\mathcal{T}$ are unseen and unused. Let the source HSI cube be $\mathrm{X} \in \mathbb{R}^{H\times W\times B}$ with pixel labels $\mathrm{y} \in \{1,\dots,K\}^{H\times W}$. A network $f_\theta : \mathbb{R}^{P\times P\times B} \to \mathbb{R}^d$ encodes a $P \times P$ patch into a feature z, and a head $g : \mathbb{R}^d \to \Delta^{K-1}$ produces the center-pixel prediction $\mathrm{p} = g(f_\theta(\cdot))$. Our goal is to learn from *one* source domain a model that generalizes to a fully unseen test domain $\mathcal{T}$, with inference following the standard *pixel-level* (center-pixel) protocol and no segmentation-style supervision.

To enlarge coverage under SSDG, we synthesize $V-1$ *pseudo-source domains* with generator G; together with SD, they form V training-domain views $\{\mathrm{X}^{(v)}\}_{v=1}^{V}$. This augmentation preserves labels and the per-pixel protocol is used

only to broaden the source distribution—no target cues are introduced. We adopt a Fourier-based Spectral–Spatial Generator [32] that shares class semantics in the frequency domain while injecting cube-wise shifts, yielding *class-consistent yet stylistically diverse* pseudo sources. Unlike adversarial DG that couples generation with min–max training, we only use the generator as a distributional augmenter [37]; CAJA can be plugged into adversarial backbones if desired.

Challenges in cross-scene HSI under SSDG. Even with "source↔pseudo-source" alignment, SSDG for cross-scene HSI faces three key challenges: (i) **Context shift**: the co-occurrence pattern and boundary state around a pixel vary across scenes, making decision statistics *conditional on context*; global or class-wise alignment ignores this condition. (ii) **Boundary mixing**: mixed-class patches near edges yield noisier supervision and over-confidence when treated as clean single-class regions, and this uncertainty is itself *context-dependent*. (iii) **Representation–decision mismatch**: aligning only features or only predictions often yields a mismatch between learned representations and the induced decisions in new scenes.

These observations motivate enforcing *invariance conditioned on context*.

3.2 Context-conditioned variables

In cross-scene HSI, a class's statistics depend on its *neighborhood co-occurrence* and *boundary state*. We therefore seek *context-conditioned invariance*: given a local context, both representations and decisions remain stable. Under SSDG, we construct train-time context variables that are *semantically grounded* (from neighborhood labels), *continuous and cheap* (computed online), and *task-aligned* (compatible with pixel-level supervision). Concretely, we use a co-occurrence descriptor $q^\star$, a homogeneity indicator h, and the joint context $C = (q^\star, h)$.

We first expose neighborhood co-occurrence as a distribution $q^\star$ that captures *who* surrounds the center pixel, focusing on *scene composition* rather than *class identity*. Using a $k \times k$ window (default $k = P$), let $\mathcal{N}$ exclude the center and $\mathrm{y}_N = \{\mathrm{y}_j : j \in \mathcal{N}\}$ be the neighbor labels. With $\mathrm{Hist}(\cdot)$ the normalized K-bin histogram and Δ^{K-1} the $(K-1)$-simplex, define

$$q^\star \;=\; \mathrm{Hist}(\mathrm{y}_N) \;\in\; \Delta^{K-1}. \tag{1}$$

This descriptor records which classes surround the center and in what proportions, and provides a label-derived, noise-robust prior for prediction.

We then derive a one-dimensional homogeneity scalar from $q^\star$:

$$h \;=\; \phi(q^\star) \;\in [0,1], \tag{2}$$

where $\mathrm{normEnt}(\cdot)$ is entropy normalized by $\log K$, and $\phi(\cdot) \;= 1 - \mathrm{normEnt}(\cdot)$ (the max-probability variant $\max_j(\cdot)_j$ is analyzed in the Supp. Sec. 4.1). Lower h indicates boundary-like mixing; higher h corresponds to homogeneous interiors. As a monotone proxy of mixing, h is used for confidence calibration and sample weighting [45].

Using only h loses neighborhood composition information (which classes surround the center pixel), while using only $q^\star$ lacks a mixing/uncertainty scale. We therefore combine the semantic fingerprint $q^\star$ with its uncertainty scale h as the joint context $C = (q^\star, h)$.

Supervision and alignment are then enforced *under the same context* C: conditioning on $q^\star$ targets *context shift* (same class, different neighbors), while calibrating by h handles *boundary mixing*. Using the same C to organize both feature statistics and decision distributions mitigates *representation–decision mismatch*. C is used only during training; test-time inference follows the standard center-pixel protocol without computing $(q^\star, h)$.

3.3 Context-aware joint alignment

The neighborhood class distribution around the center pixel provides key evidence for learning locally generalizable parameters. We therefore propose a representation–decision scheme tailored to cross-scene HSI that explicitly enforces joint context-conditioned invariance.

Concretely, we use the neighborhood label histogram and homogeneity as context variables, inject them into supervision (CAS), and use the same signal to form context buckets for joint alignment (CCJA). The single-label, per-pixel inference protocol remains unchanged. Unlike global/class-wise alignment or two-stage distillation, our training constrains representation and decision under the same context, which is seldom systematized for cross-scene HSI.

Context-aware supervision. In CAS we adopt a *BCE-based* objective. *Notation:* Δ^{K-1} is the $(K-1)$-simplex, $\sigma(\cdot)$ the logistic sigmoid, and V the number of training-domain views. Cross-scene HSI exhibits *context shift* and *boundary mixing*, making boundary pixels ambiguous; enforcing a mutually exclusive simplex (softmax–CE) tends to yield over-confident [23], poorly transferring decisions. Modeling each class as an independent Bernoulli (sigmoid + BCE) *relaxes* the simplex at training time, enables per-class uncertainty expression, and supports *context-aware* calibration via class-wise weighting. Empirically (Sec. 4.3), this class-wise Bernoulli modeling reduces boundary over-confidence and improves cross-scene robustness, while test-time prediction remains the standard probability *argmax*.

Formally, with context $C = (q^\star, h)$, the network outputs per-class logits $\mathbf{z} \in \mathbb{R}^K$ and probabilities $\mathbf{p} = \sigma(\mathbf{z}) \in (0,1)^K$ for the center pixel, given ground truth $y \in \{1, \ldots, K\}$ with one-hot $\mathrm{onehot}(y) \in \{0,1\}^K$. We use $h \in [0,1]$ as homogeneity and $q^\star \in \Delta^{K-1}$ as the neighborhood co-occurrence prior.

To address *boundary mixing*, we map h to an *adaptive smoothing* strength s that reduces over-confidence in boundary regions and prevents the source-specific boundary rule from being over-extrapolated: $s = \alpha(1-h) + \beta$, where $\alpha \in [0,1]$ controls sensitivity to homogeneity and $\beta \geq 0$ sets a baseline smoothing level. Applying s to the label produces a soft target

$$\tilde{\mathbf{t}} = (1-s)\,\mathrm{onehot}(y) + \frac{s}{K}\mathbf{1} \ \in [0,1]^K, \tag{3}$$

and the classification loss is

$$\mathcal{L}_{\mathrm{BCE}} = -\sum_{c=1}^{K} \big[\tilde{t}_c \log \sigma(z_c) + (1-\tilde{t}_c)\log\big(1-\sigma(z_c)\big)\big]\,. \tag{4}$$

This adaptive smoothing realizes an *uncertainty-aware margin*: stronger mixing $\Rightarrow$ larger $s \Rightarrow$ less peaky targets and improved cross-scene robustness.

To handle *context shift*, we use $q^\star$ as a local prior and require predictions to be *compatible* with this co-occurrence, i.e., learn $p(y \,|\, \text{context})$ rather than only $p(y \,|\, \text{pixel})$. We enforce center–context consistency via a class-wise symmetric Bernoulli KL between the view-averaged prediction $\bar{\mathbf{p}} = \frac{1}{V}\sum_{v=1}^{V} \sigma(\mathbf{z}^{(v)})$ and $q^\star$:

$$\mathcal{L}_{\mathrm{ctx}}^{\mathrm{bin}} = \frac{1}{2}\sum_{c=1}^{K} \Big[\mathrm{KL}\big(\mathrm{Bern}(\bar{p}_c) \,\|\, \mathrm{Bern}(q_c^\star)\big) + \mathrm{KL}\big(\mathrm{Bern}(q_c^\star) \,\|\, \mathrm{Bern}(\bar{p}_c)\big)\Big], \tag{5}$$

with $\mathrm{KL}(\mathrm{Bern}(u) \,\|\, \mathrm{Bern}(v)) = u\log\frac{u}{v} + (1-u)\log\frac{1-u}{1-v}$. BCE is also consistent with our per-class probabilistic formulation: $\tilde{\mathbf{t}}(h)$ and $q^\star$ are class-wise Bernoulli targets, for which Bernoulli KL is natural (see Supp. Sec. 3). To emphasize fragile regions under domain shift, we weight this consistency by $w_{\mathrm{ctx}}(h) = \gamma(1-h)$, where $\gamma > 0$ sets the overall strength and $1-h$ serves as an uncertainty proxy, focusing training on boundary/mixed pixels.

The CAS objective is thus

$$\mathcal{L}_{\mathrm{CAS,BCE}} = \mathcal{L}_{\mathrm{BCE}} + \lambda_{\mathrm{cas}}\, w_{\mathrm{ctx}}\, \mathcal{L}_{\mathrm{ctx}}^{\mathrm{bin}}, \tag{6}$$

where $\lambda_{\mathrm{cas}} > 0$ balances classification and consistency. At test time, we keep standard center-pixel *argmax*. Unlike patch-level multi-label context modeling, CAS *retains the single-label per-pixel objective* while explicitly handling boundary mixing and co-occurrence shift.

Context-conditioned joint alignment. Global or class-wise alignment (e.g., MMD/CORAL/DANN or output matching) ignores that, in cross-scene HSI, *where a pixel appears* (co-occurrence + boundary state) changes across scenes even for the same class. We therefore discretize homogeneity h into K_h bins and quantize $q^\star$ into K_q codes to form *context buckets* $\mathcal{B}$ that group samples by shared local context rather than by class [19];here $b \in \mathcal{B}$ denotes one bucket, and very small buckets are merged to ensure $|b| \geq M_{\min}$. CAJA is stable across a broad range of (K_h, K_q) and $M_{\min}$, indicating the gains are not due to fragile discretization (Supp. Sec. 4.3).

In each mini-batch, we pair the original source view with one FSSG-generated pseudo-source view. For each context bucket b, samples sharing the same context key are grouped across the two views, and alignment is performed within b when both views contain valid samples.

For two views $v \in \{1,2\}$, let $\mathbf{Z}_b^{(v)} \in \mathbb{R}^{|b|\times d}$ be features in bucket b. We match second-order statistics (CORAL-like) [34]:

$$\mathcal{L}_{\mathrm{feat}}(b) = \big\|\mathrm{Cov}(\mathbf{Z}_b^{(1)}) - \mathrm{Cov}(\mathbf{Z}_b^{(2)})\big\|_F^2, \tag{7}$$

where $\mathrm{Cov}(\cdot)$ is the unbiased estimator with a small ridge $\delta\mathbf{I}$ ($\delta>0$).

Define bucket-wise mean probabilities $\bar{\mathbf{p}}_b^{(v)} = \frac{1}{|b|}\sum_{i\in b}\sigma(\mathbf{z}_i^{(v)}) \in (0,1)^K$, clamped to $[\varepsilon, 1-\varepsilon]$ for stability. Decisions are aligned via class-wise symmetric Bernoulli KL [19]:

$$\mathcal{L}_{\mathrm{pred}}(b) = \tfrac{1}{2}\sum_{c=1}^{K}\Big[\mathrm{KL}\big(\mathrm{Bern}(\bar{p}_{b,c}^{(1)}) \,\|\, \mathrm{Bern}(\bar{p}_{b,c}^{(2)})\big) + \mathrm{KL}\big(\mathrm{Bern}(\bar{p}_{b,c}^{(2)}) \,\|\, \mathrm{Bern}(\bar{p}_{b,c}^{(1)})\big)\Big]. \tag{8}$$

Given the markedly different robustness across contexts, we introduce an uncertainty-based context weighting *before* aggregating bucket-wise alignment terms, enabling risk-sensitive optimization [18]:

$$\omega(b) = \frac{1}{|b|}\sum_{i\in b}\gamma\,(1-h_i), \tag{9}$$

where $\gamma>0$ controls the overall weight scale and $1-h_i$ is a proxy of boundary mixing (lower $h \Rightarrow$ higher ambiguity). Averaging over b yields a stable context-level risk score that prioritizes fragile contexts and down-weights homogeneous ($h\approx 1$) ones. The overall CCJA loss is

$$\mathcal{L}_{\mathrm{CCJA}} = \lambda_{\mathrm{feat}}\sum_{b\in\mathcal{B}}\omega(b)\,\mathcal{L}_{\mathrm{feat}}(b) + \lambda_{\mathrm{pred}}\sum_{b\in\mathcal{B}}\omega(b)\,\mathcal{L}_{\mathrm{pred}}(b), \tag{10}$$

with $\lambda_{\mathrm{feat}}, \lambda_{\mathrm{pred}}>0$. Binding (7) and (8) to the *same* bucket key constrains both sides of the same invariance: within a fixed context, feature statistics are shared and the induced decisions also agree.

Unified objective. We optimize a *single* context-conditioned objective under the same $C=(q^\star,h)$ that jointly minimizes (i) context-aware classification risk at the center pixel and (ii) representation/decision discrepancy across views within the same context buckets:

$$\begin{aligned}\mathcal{L}_{\mathrm{CAJA}} = \frac{1}{V}\sum_{v\in V}\mathcal{L}_{\mathrm{BCE}}\big(\mathbf{z}^{(v)};\,\tilde{\mathbf{t}}(h)\big) + \lambda_{\mathrm{cas}}\,w_{\mathrm{ctx}}(h)\,\mathcal{L}_{\mathrm{ctx}}^{\mathrm{bin}}\big(\bar{\mathbf{p}}, q^\star\big)\\ + \sum_{b\in\mathcal{B}(C)}\omega(b)\Big[\lambda_{\mathrm{feat}}\,\mathcal{L}_{\mathrm{feat}}(b) + \lambda_{\mathrm{pred}}\,\mathcal{L}_{\mathrm{pred}}(b)\Big].\end{aligned} \tag{11}$$

Here, the soft target $\tilde{\mathbf{t}}(h)$ is given in Eq. 3, the BCE term $\mathcal{L}_{\mathrm{BCE}}\big(\mathbf{z};\,\tilde{\mathbf{t}}(h)\big)$ and center–context consistency $\mathcal{L}_{\mathrm{ctx}}^{\mathrm{bin}}\big(\bar{\mathbf{p}}, q^\star\big)$ are defined in Eqs. 4–5. $\mathcal{B}(C)$ denotes context buckets induced by C, and $\omega(b)$ is a risk weight from bucket uncertainty. All terms are conditioned on the same C, enforcing a *context-conditioned coupling* of representation and decision, rather than a mere sum of unrelated regularizers. This view is further motivated by an informal risk-decomposition perspective in Supp. Sec. 1.

4 Experiment

4.1 Experimental setup

Datasets. We evaluate CAJA on three cross-scene HSI benchmarks. **Houston** [41]: Two UH-campus scenes sharing the same spectral range (0.38–1.05 µm) but with different spectral resolutions (144 vs. 48 bands). We use the 48-band subset of **Houston13** (349×1905) and the overlapping region of **Houston18** (209×955), retaining 7 shared classes. The input patch size is 13×13. **Pavia** [42]: Two ROSIS aerial scenes: **Pavia University** (PU, 610×340, 103 bands) and **Pavia Center** (PC, 1096×715, 102 bands), both covering 0.43–0.86 µm. For spectral consistency, the 103rd PU band is removed (102 bands for both). We use 7 shared classes and a patch size of 9×9. **WHU-Hi** [44]: Two high-resolution scenes acquired by Headwall Nano-Hyperspec (400–1000 nm): **HongHu** (940×475, 270 bands, 0.043 m) and **HanChuan** (1217×303, 274 bands, 0.109 m). A 270-band subset is selected from HanChuan to match HongHu, and 3 shared classes are used. The patch size is 9×9.

Training. We follow the SSDG protocol, using source labels only and enforcing *no* spatial overlap with the target domain. We split the source domain into train/val with a **0.8/0.2** ratio. To increase training diversity, we generate $V-1$ semantics-preserving pseudo-source views via FSSG (style/statistics perturbations without target cues). Training is performed on both real and pseudo-source views. We use a lightweight spectral–spatial encoder. Each stage contains a spectral branch (involution + 1×1 conv) and a spatial branch (3×3 conv), which are fused by a learnable softmax gate. A final linear layer outputs class logits, and the pre-classifier features are used for CCJA alignment. Given train-time context $\mathcal{C} = (q^\star, h)$, the loss is $\mathcal{L} = \mathcal{L}_{\text{bce}} + \lambda_{\text{cas}}\, w_{\text{ctx}}\, \mathcal{L}_{\text{ctx}}^{\text{bin}} + \lambda_{\text{coral}} \sum_{b\in\mathcal{B}} \mathcal{L}_{\text{feat}}(b) + \lambda_{\text{kl}} \sum_{b\in\mathcal{B}} \mathcal{L}_{\text{pred}}(b)$. Buckets $\mathcal{B}$ are formed by binning h into N_h bins and coding $q^\star$ into N_q codewords; small buckets are merged until $|b| \geq M_{\min}$. We optionally use risk weighting $\omega(b) = \frac{1}{|b|}\sum_{i\in b} \gamma(1-h_i)$. *Stability:* clamp probabilities to $[10^{-6}, 1-10^{-6}]$, add δI (δ=10^{-4}) to covariances. AdamW (lr 3×10^{-4}, weight decay 5×10^{-4}), cosine decay with 5-epoch warm-up; batch 256; 400 epochs; mixed precision. Warm-up uses BCE+smoothing only, then enables CAS consistency and CCJA. Defaults: α=0.2, β=0, γ=1.0, λ_{cas}=0.20, λ_{coral}=0.01, λ_{kl}=0.30, N_h=3, N_q=2, $M_{\min}$=64, V=2.

Testing. We follow the *single-label, per-pixel* protocol, and no context variables or buckets are computed at test time; no target adaptation or post-processing is used. To analyze brittleness under context shift, we also report *boundary* vs. *interior* performance; ignore label (if any) is masked throughout.

Evaluation metrics. We report per-class accuracy (CA), overall accuracy (OA), Cohen's Kappa (KC), and mean Intersection over Union (mIoU). All results are reported as mean±std over 10 runs with different seeds. In addition, Negative Log-Likelihood (NLL) and Expected Calibration Error (ECE) are used to evaluate the model stability in ablation study [13, 24].

Baselines. We compare with strong DG methods tailored or extended to HSI: SDENet [40], S^2AMSNet [2], FDGNet [29], FSENet [38], and FSSG [32].

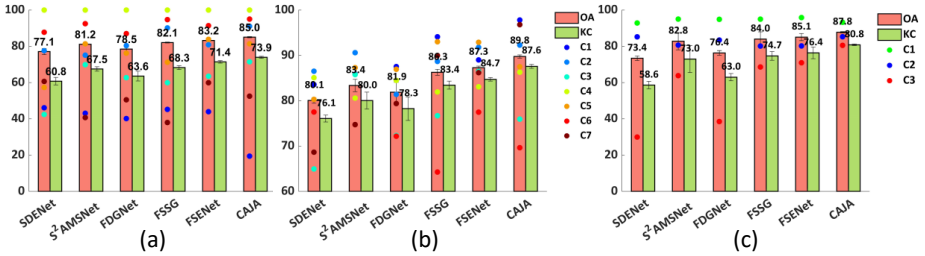

Fig. 3: Quantitative Classification results of CAJA and competing methods on three benchmark hyperspectral datasets. Subfigure (a) shows the Houston18 scene, (b) the PC scene, and (c) the HanChuan scene.

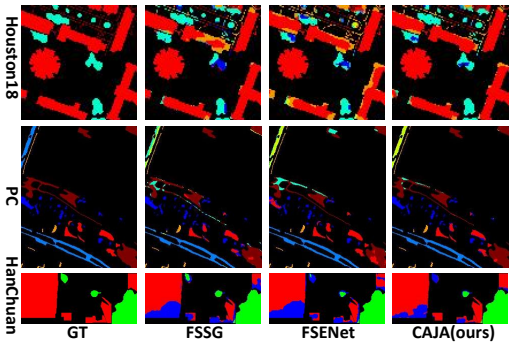

Fig. 4: Visual comparisons on representative regions: CAJA vs. FSSG vs. FSENet. CAJA exhibits fewer boundary artifacts and more coherent region predictions.

All baselines follow their official settings on identical source/target splits and are evaluated with the same metrics and seeds for fairness.

4.2 Performance on cross-scene HSI classification

Quantitative Analysis. As shown in Fig. 3, across all three datasets CAJA attains the highest **OA** (red) and consistently higher **KC** (green), indicating better pixel accuracy and stronger agreement under class imbalance. **Intra-class fairness:** Per-class accuracy points (7 for Houston/Pavia, 3 for WHU-Hi) cluster closer to the bar tops for CAJA, showing that gains extend beyond majority classes, in line with homogeneity-driven adaptive smoothing and center–context Bernoulli KL. We further quantify this effect (Supp. Sec. 7.2) using context-bucket intra-class fairness metrics (**mWorst**↑, **mICD**↓), where CAJA improves worst-group performance while reducing within-class disparity. **Stability:** Smaller error bars on OA/KC reflect lower variance and stronger generalization, consistent with joint alignment of features and predictions within shared context buckets that alleviates partial adaptation. **OA vs. KC:** Cases where

Table 1: Attribution ablation of source-side augmentation and CAJA on Houston.

Method	BCE	CAS	AdaIN	AdaIN+CAJA	FSSG	FSSG+CAJA
OA(%)	78.85	82.61	79.42	84.29	79.98	**85.05**

Table 2: OA (%)/KC ($\kappa\times100$)/mIoU (%) of CAS, CCJA, and CAJA on three datasets.

Dataset	CAS			CCJA			CAJA		
	OA	KC	mIoU	OA	KC	mIoU	OA	KC	mIoU
Houston	83.45	71.40	60.12	84.26	72.42	61.99	**85.05**	**73.94**	**63.07**
Pavia	87.21	84.56	75.38	88.07	85.62	77.04	**89.75**	**87.57**	**79.31**
WHU-Hi	85.85	77.62	72.84	87.18	79.80	74.9	**87.77**	**80.84**	**76.14**
Mean	85.50	77.86	69.45	86.50	79.28	71.31	**87.52**	**80.78**	**72.84**

competitors have similar OA but clearly lower KC suggest majority-class–biased gains, while CAJA's concurrent improvements in KC and CA highlight more balanced, class-level robustness. Compared to FSENet and FSSG, CAJA further reduces errors under co-occurrence shifts and boundary mixing by explicitly enforcing context-conditioned supervision and alignment.

Qualitative Analysis. As shown in Fig. 4, CAJA (rightmost column) produces cleaner regions, sharper boundaries, and visibly fewer speckle and bleeding artifacts. This matches our design: **CAS** reduces boundary over-confidence, while **CCJA** jointly aligns features and predictions under a shared context. On Houston18, CAJA better preserves thin structures with fewer frayed edges; on Pavia Center, large paved areas remain more contiguous and less affected by shadow-induced label swaps; on HanChuan, small roof patches are retained with fewer omissions. Overall, CAJA consistently sharpens boundaries, suppresses mixed-context leakage, and reduces island errors, in agreement with its quantitative gains in OA, Kappa, and per-class accuracy.

4.3 Ablation experiment

Unless otherwise noted, improvements in this section are statistically significant under paired t-tests over 10 runs ($p<0.001$; see Supp. Sec. 7.3 for details).

Attribution of augmentation and CAJA. Since CAJA uses FSSG to generate pseudo-source views, we isolate the effect of augmentation from context-aware modeling on Houston. As shown in Table 1, FSSG alone only improves OA by +1.13% over BCE, while CAS without any augmenter already brings +3.76%. When CAJA is combined with AdaIN or FSSG, it further improves over the corresponding augmentation-only baselines by +4.87% and +5.07%, respectively. This confirms that the main gain comes from CAJA's context-aware supervision and context-conditioned joint alignment, rather than from source-side augmentation alone.

Ablation on CAJA components. As shown in Table 2, across Houston, Pavia, and WHU-Hi, **CAJA** achieves the best OA/KC/mIoU, confirming com-

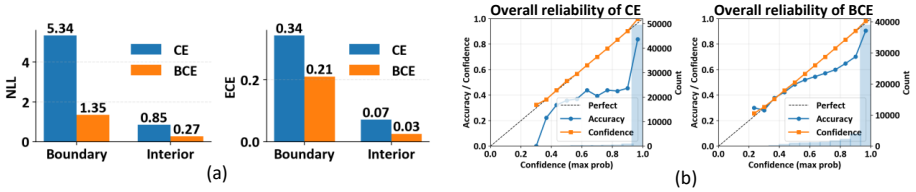

Fig. 5: Calibration diagnosis of BCE vs. CE under the single-label protocol. (a) NLL(left)/ECE(right) on boundary and interior pixels. (b) Reliability diagrams with confidence histograms for the whole image.

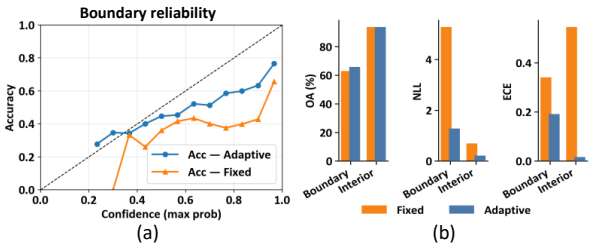

Fig. 6: Adaptive versus fixed label smoothing. (a) Reliability curves on boundary pixels for adaptive smoothing versus fixed smoothing. (b) Comparison of ECE, NLL, and OA on boundary and interior subsets under adaptive and fixed smoothing.

plementary gains from context-aware supervision (CAS) and context-conditioned alignment (CCJA). Averaged across datasets, CAJA improves **OA**/**KC**/**mIoU** from 85.50%/77.86%/69.45% to 87.52%/80.78%/72.84%. Larger gains on **KC**/mIoU than on OA suggest benefits beyond majority classes, improving class-level agreement and per-class overlap. CAS alone enhances boundary robustness (homogeneity-adaptive smoothing + center–context KL), whereas CCJA alone enforces invariance by aligning features and predictions within context buckets. Their joint use—coupling representation and decision under the same context—yields the most stable improvements across heterogeneous scene shifts. Supplementary ablations further support this shared-context design: the full context $C = (q^\star, h)$ outperforms using only h or only $q^\star$.

BCE *vs.* CE. The bars in Fig. 5 (a) show that replacing softmax+CE with sigmoid+BCE clearly improves probabilistic quality. On boundary pixels, BCE markedly lowers NLL and ECE, indicating reduced over-confidence where labels are ambiguous; in interior regions, BCE still yields lower NLL/ECE than CE, so calibration improves without harming well-determined pixels. The reliability curves in Fig. 5 (b) echo this: CE is systematically over-confident, with a curve below the diagonal and confidence mass pushed into the highest bin, whereas BCE moves the curve closer to the diagonal with a less extreme histogram. These results support that class-wise Bernoulli modeling relaxes the simplex constraint, better captures uncertainty in mixed/boundary contexts, and provides a more

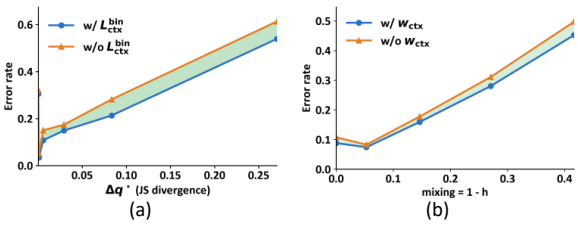

Fig. 7: Ablation of center–context consistency and context weighting within CAS. (a) Error rate vs. $\Delta q^\star$ with and without $\mathcal{L}_{\text{ctx}}^{\text{bin}}$. (b) Error rate vs. $1-h$ with and without $w_{\text{ctx}}(h)$

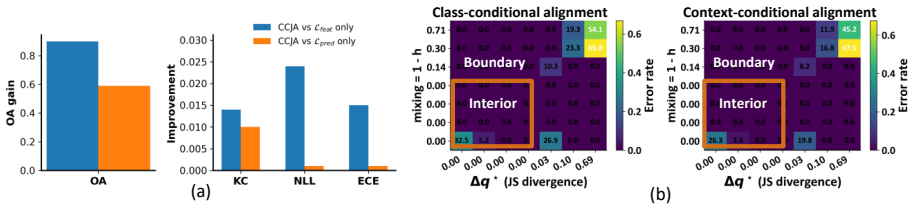

Fig. 8: Effect of context-conditioned joint alignment (CCJA). (a) CCJA vs. the feature-only ($\mathcal{L}_{\text{feat}}$) and prediction-only ($\mathcal{L}_{\text{pred}}$) variants. (b) Class conditional alignment vs. context-conditioned alignment

reliable basis for our subsequent context-aware calibration and alignment. This empirical trend is also consistent with the probabilistic view in Supp. Sec. 3.

Ablation of CAS. Encoding homogeneity h into the smoothing strength improves prediction reliability, especially in ambiguous regions. In Fig. 6(b), adaptive smoothing reduces NLL/ECE on **boundary pixels** versus fixed smoothing, while also improving boundary OA, indicating better calibration without sacrificing accuracy. In **interior regions**, OA and calibration remain comparable or slightly better, and the reliability plot in Fig. 6(a) shows a curve closer to the diagonal with a less extreme confidence histogram. These results support homogeneity-aware smoothing: modulating s with h improves boundary calibration and robustness while preserving interior performance.

In Fig. 7(a), adding the center–context consistency term $\mathcal{L}_{\text{ctx}}^{\text{bin}}$ consistently lowers error, with the largest gains at high $\Delta q^\star$, indicating that matching center predictions to neighborhood statistics is most effective under strong context shift. In Fig. 7(b), the homogeneity-based weight $w_{\text{ctx}}(h)$ mainly reduces errors in low-h (strongly mixed) regions while leaving high-h interiors nearly unchanged, showing that context weighting focuses optimization on high-risk mixed neighborhoods without degrading clean regions. Together, $\mathcal{L}_{\text{ctx}}^{\text{bin}}$ and $w_{\text{ctx}}(h)$ play complementary roles within CAS.

Ablation of CCJA. Fig 8 (a) shows that CCJA consistently improves OA and KC over both $\mathcal{L}_{\text{feat}}$-only and $\mathcal{L}_{\text{pred}}$-only baselines, while also reducing NLL and ECE, with the largest gains on boundary pixels. This indicates that aligning

features or predictions alone is not sufficient and that coupling them under the same context is crucial for stable calibration and accuracy. In Fig 8 (b), class-conditional alignment leaves high error rates in buckets with low homogeneity (large $1 - h$) and large co-occurrence shift ($\Delta q^{\star}$), whereas context-conditional alignment systematically suppresses these high-risk regions. Further bucket-wise analyses further confirm that the gains concentrate in hard, context-shifted regions (Supp. Sec. 5). Together, these results support CCJA as a more effective mechanism for handling context shift and boundary mixing than purely feature-based or prediction-based alignment.

5 Conclusion

We presented CAJA, a context-aware joint alignment framework for single-source cross-scene HSI classification. Instead of treating pixels in isolation, CAJA elevates local context—captured by neighborhood co-occurrence $q^{\star}$ and homogeneity h—to a conditioning variable. On the supervision side, CAS combines Bernoulli modeling, homogeneity-adaptive smoothing, and a center-context consistency term to reduce boundary over-confidence and encode uncertainty in mixed regions. For alignment, CCJA performs context-conditioned alignment of features and predictions within shared context buckets of real and FSSG-generated pseudo-source views, reducing representation–decision mismatch under context shift. On three cross-scene HSI benchmarks, CAJA consistently surpasses strong SSDG baselines in accuracy, boundary robustness, and calibration, and ablations show its three context-aware components are complementary and most effective when combined. In future work, we plan to extend CAJA to multi-source and semi-supervised settings, and explore context-aware invariance in remote sensing and scene understanding.

Acknowledgements

This work is supported by the Natural Science Foundation of Jiangsu Province, China under Grant No.BK20231456; and the National Natural Science Foundation of China under Grant 62201282.

References

1. Camps-Valls, G., Tuia, D., Bruzzone, L., Benediktsson, J.A.: Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. IEEE signal processing magazine **31**(1), 45–54 (2013)
2. Chen, X., Gao, L., Zhang, M., Chen, C., Yan, S.: Spectral–spatial adversarial multidomain synthesis network for cross-scene hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–16 (2024)
3. Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y.: Deep learning-based classification of hyperspectral data. IEEE Journal of Selected topics in applied earth observations and remote sensing **7**(6), 2094–2107 (2014)

4. Chu, C.T., Rohmatillah, M., Lee, C.H., Chien, J.T.: Augmentation strategy optimization for language understanding. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 7952–7956. IEEE (2022)
5. Tachet des Combes, R., Zhao, H., Wang, Y.X., Gordon, G.J.: Domain adaptation with conditional distribution matching and generalized label shift. Advances in Neural Information Processing Systems **33**, 19276–19289 (2020)
6. Cui, J., Zhu, B., Xu, Q., Tian, Z., Qi, X., Yu, B., Zhang, H., Hong, R.: Generalized kullback-leibler divergence loss. arXiv preprint arXiv:2503.08038 (2025)
7. Cui, K., Li, R., Polk, S.L., Lin, Y., Zhang, H., Murphy, J.M., Plemmons, R.J., Chan, R.H.: Superpixel-based and spatially regularized diffusion learning for unsupervised hyperspectral image clustering. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–18 (2024)
8. Dong, L., Geng, J., Jiang, W.: Spectral–spatial enhancement and causal constraint for hyperspectral image cross-scene classification. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–13 (2024)
9. Englesson, E., Azizpour, H.: Generalized jensen-shannon divergence loss for learning with noisy labels. Advances in Neural Information Processing Systems **34**, 30284–30297 (2021)
10. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. pp. 1180–1189. PMLR (2015)
11. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of machine learning research **17**(59), 1–35 (2016)
12. Gulrajani, I., Lopez-Paz, D.: In search of lost domain generalization. arXiv preprint arXiv:2007.01434 (2020)
13. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: International conference on machine learning. pp. 1321–1330. PMLR (2017)
14. He, Y., Tu, B., Liu, B., Li, J., Plaza, A.: Hsi-mformer: Integrating mamba and transformer experts for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing (2025)
15. Huang, S., Xiao, W., Chen, H., Bejo, S.K., Zhang, H.: Hyperspectral image classification based on a locally enhanced transformer network. IEEE Transactions on Geoscience and Remote Sensing (2025)
16. Jia, S., Jiang, S., Zhang, S., Xu, M., Jia, X.: Graph-in-graph convolutional network for hyperspectral image classification. IEEE Transactions on Neural Networks and Learning Systems **35**(1), 1157–1171 (2022)
17. Li, J., Li, G., Shi, Y., Yu, Y.: Cross-domain adaptive clustering for semi-supervised domain adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2505–2514 (2021)
18. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
19. Liu, X., Guo, Z., Xing, F., You, J., Kuo, C.C.J., El Fakhri, G., Woo, J.: Adversarial unsupervised domain adaptation with conditional and label shift: Infer, align and iterate. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10367–10376 (2021)
20. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: International conference on machine learning. pp. 97–105. PMLR (2015)

21. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. Advances in neural information processing systems **31** (2018)
22. Mancini, M., Bulo, S.R., Caputo, B., Ricci, E.: Robust place categorization with deep domain generalization. IEEE Robotics and Automation Letters **3**(3), 2093–2100 (2018)
23. Müller, R., Kornblith, S., Hinton, G.E.: When does label smoothing help? Advances in neural information processing systems **32** (2019)
24. Niculescu-Mizil, A., Caruana, R.: Predicting good probabilities with supervised learning. In: Proceedings of the 22nd international conference on Machine learning. pp. 625–632 (2005)
25. Parente, M., Iordache, M.D.: Sparse unmixing of hyperspectral data: The legacy of sunsal. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. pp. 21–24. IEEE (2021)
26. Pastorino, M., Moser, G., Serpico, S.B., Zerubia, J.: Crfnet: A deep convolutional network to learn the potentials of a crf for the semantic segmentation of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing (2024)
27. Peng, D., Wu, J., Han, T., Li, Y., Wen, Y., Yang, G., Qu, L.: Disentanglement-inspired single-source domain-generalization network for cross-scene hyperspectral image classification. Knowledge-Based Systems **303**, 112413 (2024)
28. Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., et al.: Recent advances in techniques for hyperspectral image processing. Remote sensing of environment **113**, S110–S122 (2009)
29. Qin, B., Feng, S., Zhao, C., Xi, B., Li, W., Tao, R.: Fdgnet: Frequency disentanglement and data geometry for domain generalization in cross-scene hyperspectral image classification. IEEE Transactions on Neural Networks and Learning Systems (2024)
30. Rasti, B., Zouaoui, A., Mairal, J., Chanussot, J.: Image processing and machine learning for hyperspectral unmixing: An overview and the hysupp python package. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–31 (2024)
31. Saito, K., Watanabe, K., Ushiku, Y., Harada, T.: Maximum classifier discrepancy for unsupervised domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3723–3732 (2018)
32. Shi, B., Cao, G., Zhang, Y., Liu, Y., Yang, K.: Fourier-based spectral–spatial generator for cross-scene hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing (2025)
33. Shu, R., Bui, H.H., Narui, H., Ermon, S.: A dirt-t approach to unsupervised domain adaptation. arXiv preprint arXiv:1802.08735 (2018)
34. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: European conference on computer vision. pp. 443–450. Springer (2016)
35. Sun, H., Zheng, X., Lu, X.: A supervised segmentation network for hyperspectral image classification. IEEE Transactions on Image Processing **30**, 2810–2825 (2021)
36. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: Maximizing for domain invariance. arxiv 2014. arXiv preprint arXiv:1412.3474 (2019)
37. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14383–14392 (2021)
38. Yang, K., Shi, B., Cao, G., Zhang, Y.: Fourier phase preservation and spectral random augmentation for cross-scene hyperspectral image classification with ensemble networks. IEEE Transactions on Geoscience and Remote Sensing (2025)

39. Zhang, L., Zhang, L., Tao, D., Huang, X.: On combining multiple features for hyperspectral remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing **50**(3), 879–893 (2011)
40. Zhang, Y., Li, W., Sun, W., Tao, R., Du, Q.: Single-source domain expansion network for cross-scene hyperspectral image classification. IEEE Transactions on Image Processing **32**, 1498–1512 (2023)
41. Zhang, Y., Li, W., Zhang, M., Qu, Y., Tao, R., Qi, H.: Topological structure and semantic information transfer network for cross-scene hyperspectral image classification. IEEE Transactions on Neural Networks and Learning Systems **34**(6), 2817–2830 (2021)
42. Zhang, Y., Li, W., Zhang, M., Wang, S., Tao, R., Du, Q.: Graph information aggregation cross-domain few-shot learning for hyperspectral image classification. IEEE Transactions on Neural Networks and Learning Systems **35**(2), 1912–1925 (2022)
43. Zhang, Y., Zhang, M., Li, W., Wang, S., Tao, R.: Language-aware domain generalization network for cross-scene hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–12 (2023)
44. Zhong, Y., Hu, X., Luo, C., Wang, X., Zhao, J., Zhang, L.: Whu-hi: Uav-borne hyperspectral with high spatial resolution (h2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with crf. Remote Sensing of Environment **250**, 112012 (2020)
45. Zhong, Z., Li, J., Luo, Z., Chapman, M.: Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework. IEEE transactions on geoscience and remote sensing **56**(2), 847–858 (2017)
46. Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C.: Domain generalization: A survey. IEEE transactions on pattern analysis and machine intelligence **45**(4), 4396–4415 (2022)