# WHAT SHOULD BE SHARED ACROSS SCENES? SELF-SUPERVISED SCENE-STYLE-CONDITIONED CORRESPONDENCE LEARNING FOR HYPERSPECTRAL IMAGE DOMAIN ADAPTATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Cross-scene hyperspectral image (HSI) domain adaptation aims to transfer knowledge from a labeled source scene to an unlabeled target scene under substantial spectral and spatial differences. Existing methods mainly rely on feature alignment, pseudo-labeling, or cross-scene similarity to establish transferable relations between the two scenes. However, such relations are typically inferred directly from observed features or predictions, without separating scene-dependent spectral variation from class-related structure. To address this issue, we propose scene-style-conditioned correspondence learning (SCCL), a cross-scene adaptation framework that shares only relations that remain meaningful under scene-style variation. We first model scene-dependent feature statistics with a bounded stochastic scene-style posterior and construct scene-style transport under a spectral-admissibility constraint. We then examine whether source–target semantic relations persist across the resulting admissible views, yielding a correspondence-survival probability. The surviving relations are organized by pairwise correspondence admissibility and a globally normalized correspondence plan, which directly induces soft target semantics for final learning. Experiments on Houston, Pavia, and Shanghai–Hangzhou cover six directed cross-scene tasks, demonstrating the effectiveness and superiority of the proposed SCCL method. The code is available at `https://github.com/Sky-byte-box/SCCL_code`.

## 1 INTRODUCTION

Cross-scene hyperspectral image (HSI) domain adaptation aims to transfer a classifier learned from a labeled source scene to an unlabeled target scene that contains the same land-cover categories but differs in spectral responses and spatial appearance. These differences commonly arise from changes in acquisition time, geographic environment, atmospheric condition, or sensor configuration, causing a model trained on one scene to generalize poorly to another. Unlike traditional HSI classification, cross-scene adaptation must preserve class-related spectral-spatial information while remaining effective under scene-dependent distribution shifts, without using target annotations. This setting provides a practical way to reuse labeled HSI knowledge across scenes and has motivated a growing body of cross-scene adaptation studies (Zhang et al., 2021; Tang et al., 2022; Zhang et al., 2023).

Most cross-scene HSI adaptation methods address domain shift by aligning source and target representations while preserving class discrimination. For example, discriminative cooperative alignment (DCA) reduces geometric and statistical discrepancies through coupled projections (Zhang et al., 2021), while unsupervised joint adversarial domain adaptation (UJADA) introduces adversarial learning for labeled-source and unlabeled-target transfer (Tang et al., 2022). More recent methods extend this idea through multilevel or classwise alignment. Specifically, multilevel unsupervised domain adaptation (MLUDA) performs image-, feature-, and logic-level alignment (Cai et al., 2024a), whereas classwise prototype-guided alignment network (CPGAN) uses high-confidence target pseudo-labels and class prototypes to reduce category-dependent discrepancy (Xie et al., 2024). Adaptive smooth adversarial learning further combines adversarial adaptation with class-discrimination refinement for cross-domain HSI classification (Liu et al., 2025). BiDA links separate source and target adaptation spaces through cross-domain attention (Zhang et al., 2025), while DAMamba and TAGIN introduce



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

multilevel prototype alignment and gated feature interaction, respectively (Duan et al., 2026; Yang et al., 2026). These approaches make source–target alignment a central mechanism for transferring spectral–spatial knowledge across scenes.

To exploit the unlabeled target scene more directly, another line of work introduces self-supervised or self-training signals into cross-scene HSI adaptation. For example, supervised contrastive learning-based unsupervised domain adaptation (SCL-UDA) uses target pseudo-labels to support class-aware contrastive representation learning (Li et al., 2023). Masked self-distillation domain adaptation constructs a target-side pretext task by matching predictions from masked and complete HSI patches (Fang et al., 2024). Meanwhile, cycle self-training (Cycle-ST) iteratively improves target pseudo-label use while combining domain- and class-level adversarial alignment (Li et al., 2024). In addition, consistency-aware customized learning (CACL) uses agreement between prototype-based labels and predicted pseudo-labels to guide sample-dependent contrastive learning (Ding et al., 2025), while uncertainty-informed prototype contrastive learning selects target pseudo-labels according to uncertainty before prototype-based alignment (Xu et al., 2025). BMDMnet combines teacher-based target predictions with low-entropy source–target sample mixing (Dang et al., 2026). They show that unlabeled target samples can provide useful semantic cues beyond direct distribution matching.

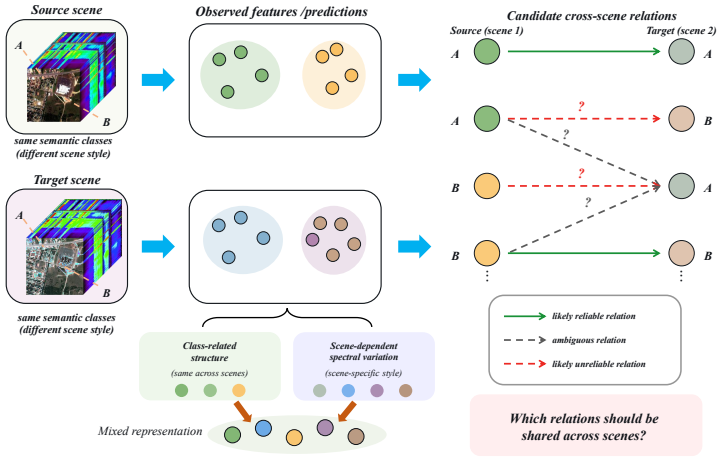



Figure 1: Cross-scene relations inferred from observed features or predictions may mix class-related structure with scene-dependent spectral variation, making reliable correspondences difficult to distinguish from ambiguous or unreliable ones.

It should be emphasized that these feature alignment and target-side learning provide useful cues for adaptation, but relations inferred from aligned features or target-side predictions may reflect both class-related structure and scene-dependent spectral variation. As illustrated in Fig. 1, this ambiguity raises the central question of this work: *what should be shared across scenes?*

To address the above issue, we propose scene-style-conditioned correspondence learning (SCCL) for cross-scene HSI domain adaptation. An overview of the proposed SCCL is shown in Fig. 2. More concretely, we first separate channel-wise scene statistics from normalized feature content and model the former with a bounded scene-style posterior. We next generate stochastic source–target style paths from this posterior and use spectral relations to retain only spectrally admissible views. Third, we examine whether classifier predictions and source-prototype assignments persist across these views, and combine their survival with target-memory agreement and the latent style gap to determine admissible source–target relations. Finally, the admitted relations are organized by pairwise correspondence admissibility and globally normalized into a correspondence plan, from which target semantic assignments and the empirical training distribution are derived for final learning.

The main contributions of this work are fourfold. *First*, we formulate cross-scene HSI adaptation as scene-style-conditioned correspondence learning to transfer semantic relations that remain admissible under scene variation. *Second*, we develop bounded scene-style modeling with spectrally constrained stochastic style transport to explore scene variation while preserving class-related spectral structure. *Third*, we establish style-survival correspondence with target-memory agreement and global normalization to select admissible cross-scene relations and derive target semantic supervision. *Fourth*, extensive experiments on six transfer tasks validate the effectiveness and superiority of our SCCL.

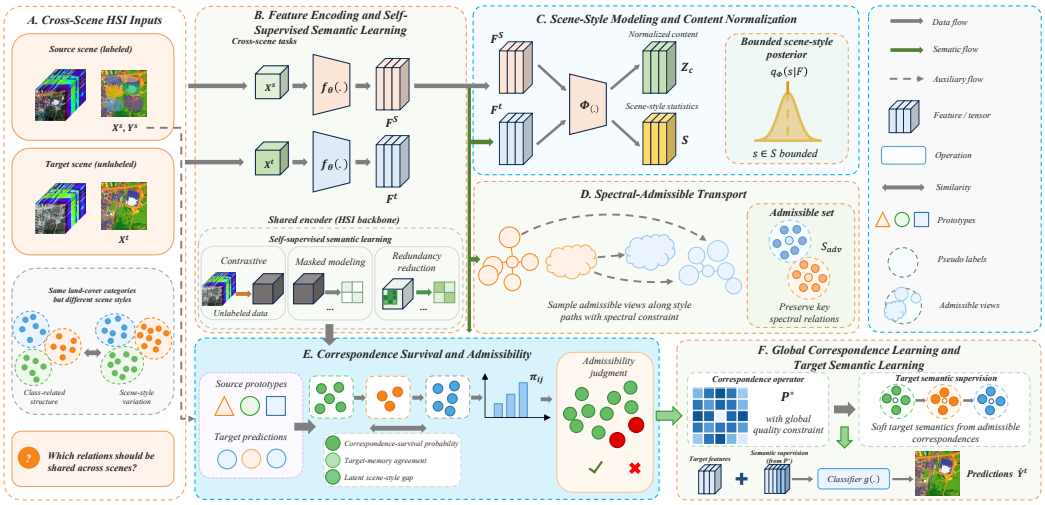



Figure 2: Overview of SCCL for hyperspectral image domain adaptation.

## 2 Related Work

### 2.1 Self-Supervised Semantic Learning

Self-supervised learning derives supervision from the internal structure of unlabeled data and has become an important way to learn semantic representations without manual annotation. Contrastive methods such as SimCLR and MoCo relate different views of the same sample to organize the representation space (Chen et al., 2020; He et al., 2020). Self-distillation and masked modeling obtain supervision from prediction agreement or reconstruction across transformed views (Caron et al., 2021; He et al., 2022), while redundancy-reduction methods encourage informative representations without requiring explicit negative pairs (Zbontar et al., 2021; Bardes et al., 2022). These approaches show that unlabeled samples can provide useful semantic cues through relations among views, predictions, or latent features. In cross-scene adaptation, however, such cues are obtained from data whose appearance may change with the scene. Consequently, view agreement alone does not determine whether a semantic relation inferred between scenes should be transferred, particularly when scene-dependent variation is mixed with class-related information.

### 2.2 Scene-Style Modeling

Feature statistics have been widely used to characterize domain-dependent style in visual representation learning. MixStyle enlarges domain diversity by mixing instance-level feature statistics (Zhou et al., 2021), while Stochastic StyleMatch introduces stochastic style generation to expose models to broader appearance variation (Zhou et al., 2023). Recent studies further separate content and style through contrastive or masked representation learning (Cai et al., 2024b; Zhang et al., 2024). These studies provide a useful basis for modeling scene-dependent variation separately from semantic content. For hyperspectral data, however, spectral responses are closely tied to land-cover discrimination, making style manipulation more constrained than in ordinary natural-image settings. Fourier-based spectral–spatial expansion has also been studied for cross-scene HSI domain generalization (Shi et al., 2025).

## 3 Method

### 3.1 Problem Formulation and Overview

Cross-scene adaptation considers two scenes that share the same land-cover categories but differ in spectral response and spatial appearance. Instead of forcing their feature distributions to coincide, we seek source–target correspondences that remain meaningful under scene-dependent variation. We denote the labeled source scene by $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{N_s}$ and the unlabeled target scene by

$\mathcal{D}_t = \{\mathbf{x}_j^t\}_{j=1}^{N_t}$, where $y_i^s \in \{1, \ldots, C\}$ and $C$ is the number of shared classes. The input $\mathbf{x}$ denotes the dataset-specific spectral or spectral-context descriptor described in Section 4.1. Both scenes are mapped into a common representation space by a shared encoder, which produces $\mathbf{F} = f_\theta(\mathbf{x})$ and $\mathbf{z} = h(\mathbf{F})/\left\|h(\mathbf{F})\right\|_2$, where $h$ is the projection head and its trainable parameters are included in $\theta$. The class posterior is $\mathbf{p} = \mathrm{softmax}(g_\omega(\mathbf{z}))$, where $g_\omega$ is the classifier. As summarized in Fig. 2, the method proceeds through shared feature encoding, bounded scene-style modeling, spectral-admissible transport, correspondence survival and admissibility, and globally normalized target semantic learning. The contrastive, masked-modeling, and redundancy-reduction elements in Fig. 2 schematically summarize self-supervised semantic cues at the shared representation stage rather than three additional optimization branches in the formulation below. For notation, the normalized-content block $Z_c$ corresponds to $\mathbf{C}(\mathbf{F})$, the admissible set $S_{adv}$ denotes the transported views satisfying the spectral constraint, and the correspondence operator $\mathbf{P}^\star$ is the global correspondence plan defined in Eq. (6). The "global quality constraint" in the figure refers to its marginal and support constraints.

### 3.2 Bounded Scene-Style Posterior and Content Normalization

Cross-scene variation in HSI often changes channel-wise feature responses while preserving the underlying land-cover semantics. Dynamic perturbation and destylization have also been explored for cross-scene HSI generalization (Hong et al., 2025). We treat these response statistics as scene style and remove them from the normalized content used by the subsequent correspondence modules. For an intermediate feature response $\mathbf{F}$ with $K$ channels, we compute its channel-wise mean $\boldsymbol{\mu}(\mathbf{F})$ and standard deviation $\boldsymbol{\sigma}(\mathbf{F})$ over the corresponding feature elements, where $\epsilon > 0$ is used for numerical protection in the standard-deviation computation. To keep the sampled scene variation within a controlled range, the style encoder maps $\mathbf{F}$ to a bounded mean and variance. Specifically, $\mathbf{m}_\phi(\mathbf{F}) = r_s \mathbf{b}_\phi(\mathbf{F})/(\left\|\mathbf{b}_\phi(\mathbf{F})\right\|_2 + \epsilon)$ and $\mathbf{v}_\phi(\mathbf{F}) = v_{\min} + (v_{\max} - v_{\min})\,\mathrm{sigm}(\mathbf{a}_\phi(\mathbf{F}))$, where $\mathbf{a}_\phi(\mathbf{F}), \mathbf{b}_\phi(\mathbf{F}) \in \mathbb{R}^{d_s}$ are the two encoder outputs. The bounded scene-style posterior is defined as:

$$q_\phi(\mathbf{s} \mid \mathbf{F}) = \mathcal{N}(\mathbf{m}_\phi(\mathbf{F}), \mathrm{diag}(\mathbf{v}_\phi(\mathbf{F}))), \tag{1}$$

where $d_s$ is the style-code dimension, $r_s$ bounds the style-mean radius, $0 < v_{\min} < v_{\max}$ bounds posterior uncertainty, and $\mathrm{sigm}(\cdot)$ is the element-wise sigmoid function. Samples from Eq. (1) are obtained by $\mathbf{s} = \mathbf{m}_\phi(\mathbf{F}) + \sqrt{\mathbf{v}_\phi(\mathbf{F})} \odot \boldsymbol{\varepsilon}_s$, where $\boldsymbol{\varepsilon}_s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d_s})$ is the reparameterization noise. The channel-wise scene statistics are then removed through $\mathbf{C}(\mathbf{F}) = (\mathbf{F} - \boldsymbol{\mu}(\mathbf{F}))/\boldsymbol{\sigma}(\mathbf{F})$, where $\boldsymbol{\mu}(\mathbf{F})$ and $\boldsymbol{\sigma}(\mathbf{F})$ denote the channel-wise statistics defined above. The normalized content $\mathbf{C}(\mathbf{F})$ is subsequently passed to the scene-style transport module.

### 3.3 Spectral-Admissible Stochastic Scene-Style Transport

To expose normalized content to target-like variation without disrupting the spectral organization that supports its semantics, we construct a stochastic source-to-target style path constrained by the spectral relations of the current feature. Spectral random augmentation has also been used to diversify cross-scene HSI representations (Yang et al., 2025). Our transported views are explicitly checked against the current feature's spectral relations. For a source-target pair, let $\mathbf{s}^s$ and $\mathbf{s}^t$ be samples from Eq. (1). With $\lambda \sim \mathrm{Beta}(\alpha, \alpha)$, the intermediate style is defined as follows:

$$\mathbf{s}(\beta) = (1-\beta)\mathbf{s}^s + \beta \mathbf{s}^t + \rho \boldsymbol{\xi}, \qquad 0 \le \beta \le \lambda, \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d_s}), \tag{2}$$

where $\beta$ locates the current point on the path, $\alpha > 0$ controls the concentration of the sampled interpolation budget, and $\rho \ge 0$ controls local stochastic variation. Each sampled style is decoded by $d_\psi$ to form the candidate $T_\beta(\mathbf{F}) = \tilde{\boldsymbol{\sigma}}(\beta) \odot \mathbf{C}(\mathbf{F}) + \tilde{\boldsymbol{\mu}}(\beta)$, where $[\tilde{\boldsymbol{\mu}}(\beta), \log \tilde{\boldsymbol{\sigma}}(\beta)]$ is the decoder output and the log-scale is exponentiated before the affine transformation. A candidate is accepted only if its spectral organization remains close to that of the original feature. We partition the $K$ channels into $G$ groups and form a cosine-relation matrix with entries $R_{gh}(\mathbf{F}) = \mathbf{r}_g(\mathbf{F})^\top \mathbf{r}_h(\mathbf{F})/(\left\|\mathbf{r}_g(\mathbf{F})\right\|_2 \left\|\mathbf{r}_h(\mathbf{F})\right\|_2 + \epsilon)$, where $\mathbf{r}_g(\mathbf{F})$ is the spatially pooled response of group $g$. The spectral-relation change is measured by $D_R(\mathbf{F}, \mathbf{F}') = G^{-2} \left\|\mathbf{R}(\mathbf{F}) - \mathbf{R}(\mathbf{F}')\right\|_F^2$, where $\left\|\cdot\right\|_F$ denotes the Frobenius norm. The path is then truncated at the farthest point that remains within the spectral-admissibility radius, *i.e.,*

$$\beta^\star(\mathbf{F}) = \sup\{\beta \in [0, \lambda] \mid D_R(\mathbf{F}, T_\beta(\mathbf{F})) \le \delta_R\}, \tag{3}$$

where $\delta_R > 0$ controls the admitted change. The accepted feature is $\tilde{\mathbf{F}} = T_{\beta^\star(\mathbf{F})}(\mathbf{F})$. The same construction is used for the reverse target-to-source path, and only admissible features enter the correspondence stage.

### 3.4 STYLE-SURVIVAL CORRESPONDENCE DISCOVERY

Spectral admissibility controls which style changes are allowed but does not determine their class semantics. We therefore use normalized source prototypes as class anchors. Source-guided prototypes have also been used to refine target pseudo-labels in HSI adaptation (Xie et al., 2026). Here, the anchors support a test of whether assignments survive admissible scene-style changes. For each class $k \in \{1, \dots, C\}$, we compute $\mathbf{c}_k^s$ from the mean normalized source embeddings with $y_i^s = k$ and set $\bar{\mathbf{c}}_k^s = \mathbf{c}_k^s / \|\mathbf{c}_k^s\|_2$, where $\bar{\mathbf{c}}_k^s$ is the normalized source class anchor. Here, $\mathbb{I}[\cdot]$ denotes the indicator function used below. For target sample $x_j^t$, we draw $M$ admissible target-to-source views from Eq. (3) and use the original feature as view $b = 0$. For each view, the classifier prediction and nearest-prototype assignment are $\hat{y}_j^{(b)} = \arg\max_k p_{jk}^{(b)}$ and $a_j^{(b)} = \arg\max_k \mathbf{z}_j^{(b)\top} \bar{\mathbf{c}}_k^s$, respectively, where $b \in \{0, \dots, M\}$, $\mathbf{z}_j^{(0)} = \mathbf{z}_j^t$, and $\mathbf{p}_j^{(b)} = \mathrm{softmax}(g_\omega(\mathbf{z}_j^{(b)}))$. We then measure whether the same class assignment survives the admissible style changes through:

$$\pi_{jk} = \frac{1}{M+1} \sum_{b=0}^{M} \mathbb{I}[\hat{y}_j^{(b)} = k] \mathbb{I}[a_j^{(b)} = k], \tag{4}$$

where a larger $\pi_{jk}$ indicates stronger agreement between prediction and prototype assignment across the sampled views. This evidence is checked against a running target prototype memory so that a surviving relation is also consistent with the target structure. Let $\mathcal{K}_t$ denote the classes with valid target memory and $\bar{\mathbf{c}}_k^t = \mathbf{c}_k^t / (\|\mathbf{c}_k^t\|_2 + \epsilon)$. The memory gate is $v_{jk} = \mathbb{I}[k = \arg\max_{\ell \in \mathcal{K}_t} \mathbf{z}_j^{t\top} \bar{\mathbf{c}}_\ell^t]$ for $k \in \mathcal{K}_t$ and $v_{jk} = 1$ otherwise, where a class enters $\mathcal{K}_t$ after receiving nonzero correspondence mass. The admitted relation is further modulated by $\chi_{ij} = 1 + \kappa \tanh(d_{ij}^{\mathrm{sty}} / (\bar{d}_{\mathrm{sty}} + \epsilon))$, where $d_{ij}^{\mathrm{sty}} = \left\| \mathbf{m}_\phi(\mathbf{F}_i^s) - \mathbf{m}_\phi(\mathbf{F}_j^t) \right\|_2$ is the scene-style gap, $\bar{d}_{\mathrm{sty}}$ is its mean over active source-target pairs in the current minibatch, and $\kappa \geq 0$ controls the modulation strength.

### 3.5 CORRESPONDENCE ADMISSIBILITY AND GLOBAL PLAN

The survival score measures semantic persistence across admissible views, whereas the memory gate checks consistency with the current target structure. These two criteria provide complementary evidence but do not yet determine the strength of an individual source–target relation. We therefore combine them with the style modulation term to define the pairwise correspondence admissibility. For a source–target pair $(i, j)$, the survival term denoted by $\pi_{ij}$ in Fig. 2 corresponds to $\pi_{j,y_i^s}$ in the formulation below, since the relevant class is fixed by the source label $y_i^s$. The resulting admissibility is given by:

$$\mathcal{S}_{ij} = \mathbb{I}[\pi_{j,y_i^s} \geq \gamma] \, v_{j,y_i^s} \, \pi_{j,y_i^s} \, \chi_{ij}, \tag{5}$$

where $\gamma \in [0, 1]$ is the minimum survival probability. Thus, $\mathcal{S}_{ij} > 0$ only when the semantic relation survives the admissible views and agrees with target memory. For each minibatch, let $\mathcal{I}$ contain source samples with at least one admitted relation and $\mathcal{J}$ contain the corresponding target samples. Their local affinity is $A_{ij} = \mathcal{S}_{ij} \exp(\mathbf{z}_i^{s\top} \mathbf{z}_j^t / \tau)$, where $i \in \mathcal{I}$, $j \in \mathcal{J}$, and $\tau > 0$ is the affinity temperature. To normalize these affinities globally, we define $a_i = \sum_{j \in \mathcal{J}} \mathcal{S}_{ij} / \sum_{u \in \mathcal{I}, v \in \mathcal{J}} \mathcal{S}_{uv}$ and $b_j = \sum_{i \in \mathcal{I}} \mathcal{S}_{ij} / \sum_{u \in \mathcal{I}, v \in \mathcal{J}} \mathcal{S}_{uv}$, where $a_i$ and $b_j$ denote the source and target correspondence masses, respectively. The global correspondence plan is then obtained through:

$$\begin{aligned} \mathbf{P}^\star = \arg\min_{\mathbf{P} \geq 0} & \sum_{i \in \mathcal{I},\, j \in \mathcal{J}} P_{ij} \log \frac{P_{ij}}{A_{ij} + \epsilon} \\ & \text{s.t. } \mathbf{P}\mathbf{1} = \mathbf{a}, \quad \mathbf{P}^\top \mathbf{1} = \mathbf{b}, \quad P_{ij} = 0 \text{ if } \mathcal{S}_{ij} = 0, \end{aligned} \tag{6}$$

where $\mathbf{1}$ is an all-ones vector of compatible dimension. Eq. (6) can be solved by Sinkhorn matrix scaling (Cuturi, 2013). The resulting correspondence mass induces a soft target assignment below:

$$q_j(k) = \frac{\sum_{i \in \mathcal{I}} \mathbb{I}[y_i^s = k] P_{ij}^\star}{\sum_{i \in \mathcal{I}} P_{ij}^\star + \epsilon}. \tag{7}$$

The target prototype is estimated as $\hat{\mathbf{c}}_k^t = \sum_{j \in \mathcal{J}} q_j(k) \mathbf{z}_j^t / (\sum_{j \in \mathcal{J}} q_j(k) + \epsilon)$ and updated by $\mathbf{c}_k^t \leftarrow m\mathbf{c}_k^t + (1-m)\hat{\mathbf{c}}_k^t$, where $\hat{\mathbf{c}}_k^t$ is the current class-$k$ estimate and $m \in [0, 1)$ is the memory momentum.

### 3.6 CORRESPONDENCE-INDUCED SEMANTIC LEARNING

The soft assignments from Eq. (7) provide target supervision and are combined with labeled source samples and their spectrally admissible transported counterparts. The resulting training representation is summarized by the empirical semantic measure as follows:

$$\widehat{\mathbb{Q}}_{\mathbf{P}^\star} = \frac{1}{Z}\left[\sum_i \delta_{(\mathbf{z}_i^s, \mathbf{e}_{y_i^s})} + \sum_i \delta_{(\tilde{\mathbf{z}}_i^s, \mathbf{e}_{y_i^s})} + \sum_{j\in\mathcal{J}} \delta_{(\mathbf{z}_j^t, \mathbf{q}_j)}\right], \tag{8}$$

where $\mathbf{e}_{y_i^s}$ is the one-hot source label, $\tilde{\mathbf{z}}_i^s = h(\tilde{\mathbf{F}}_i^s)/\left\|h(\tilde{\mathbf{F}}_i^s)\right\|_2$ is the embedding of an admissible transported source feature, $\mathbf{q}_j = [q_j(1), \ldots, q_j(C)]^\top$, $Z = 2N_s + |\mathcal{J}|$, and $\delta_{(\mathbf{z},\mathbf{q})}$ denotes the Dirac measure at $(\mathbf{z}, \mathbf{q})$. The network is optimized by cross-entropy over samples represented by $\widehat{\mathbb{Q}}_{\mathbf{P}^\star}$, where $\theta$ covers the shared encoder and projection head, $\omega$ the classifier, $\phi$ the style encoder, and $\psi$ the style decoder. During warm-up, only the two labeled source terms are used. After source prototypes are initialized, the target term is refreshed from the current correspondence plan at each adaptation iteration. Algorithm 1 summarizes the procedure.

**Algorithm 1** Training procedure of SCCL

**Require:** Labeled source $\mathcal{D}_s$, unlabeled target $\mathcal{D}_t$, encoder $f_\theta$, classifier $g_\omega$, style encoder $q_\phi$, decoder $d_\psi$

1: Warm up on source samples and spectrally admissible source-to-target views
2: Initialize source prototypes from labeled embeddings and target memories with zeros
3: **for** each adaptation iteration **do**
4: Sample source and target minibatches and estimate scene-style posteriors by Eq. (1)
5: Generate spectrally admissible transported features by Eqs. (2)–(3)
6: Compute source prototypes and evaluate target style-survival probabilities by Eq. (4)
7: Evaluate target-memory consistency and scene-style modulation
8: Build $\mathcal{S}$ and the global correspondence plan $\mathbf{P}^\star$ by Eqs. (5)–(6)
9: Derive target semantic assignments and update target memories by Eq. (7)
10: Form $\widehat{\mathbb{Q}}_{\mathbf{P}^\star}$ by Eq. (8) and update the network by cross-entropy
11: **end for**

**Ensure:** Adapted encoder and classifier for target-scene inference

## 4 EXPERIMENTS

### 4.1 DATASETS AND TRANSFER TASKS

We use three public cross-scene HSI benchmarks. Houston 2013 (H13) and Houston 2018 (H18) are both $210 \times 954$ with 48 aligned bands and seven shared categories, with 2,530 and 53,200 labeled pixels, respectively. Pavia University (PU, $610 \times 340$) and Pavia Center (PC, $1096 \times 715$) contain 102 bands and seven shared categories, with 39,332 and 39,355 labeled pixels. Shanghai (SH, $1600 \times 230$) and Hangzhou (HZ, $590 \times 230$) contain 198 bands and three shared categories, with 368,000 and 135,700 labeled pixels. These pairs define six directed adaptation tasks: H13→H18, H18→H13, PU→PC, PC→PU, SH→HZ, and HZ→SH. Source labels are used for training and source-side validation, whereas target labels are withheld during adaptation and reserved for final evaluation. Each source uses up to 180 labeled samples per class for training and 40 for validation. Classes with fewer than 220 labeled pixels, including one H18 category, use an approximately 80/20 stratified split. This split is shared by all compared methods. Each scene is band-wise standardized from its own pixel statistics. Houston and Pavia concatenate the center spectrum with local per-band mean and standard deviation from $7 \times 7$ and $11 \times 11$ neighborhoods, respectively. SH and HZ use only the center spectrum, following their public setting.

We report overall accuracy (OA) as the primary quantitative metric. Main comparisons and component ablations are repeated five times under identical data-split and training protocols, and mean±standard deviation is reported.

### 4.2 Implementation Details

A shared two-block encoder with a 128-dimensional hidden layer, gaussian error linear unit (GELU) activation, LayerNorm, dropout, and a 64-dimensional embedding is adopted as the common feature extractor for all methods, together with the task-specific heads required by each approach. All methods follow the same source-side training and adaptation schedule where applicable. The warm-up stage uses AdamW with a learning rate of $10^{-3}$ and a weight decay of $10^{-4}$ for 24 epochs, while adaptation uses AdamW with a learning rate of $5 \times 10^{-4}$ for 12 epochs and a batch size of 128. To control computational cost on large target scenes, at most 12,000 unlabeled target pixels are sampled for each adaptation run, while evaluation uses the complete target ground truth. For SCCL, we set $M = 4$, $\delta_R = 0.05$, $\gamma = 0.75$, $\tau = 0.2$, $m = 0.95$, and $\kappa = 0.5$, with all hyperparameters fixed before target-label evaluation.

### 4.3 Compared Methods and Evaluation Protocol

We compare Source only with domain-adversarial neural network (DANN) (Ganin et al., 2016), deep adaptation network (DAN) (Long et al., 2015), joint adaptation network (JAN) (Long et al., 2017), conditional domain adversarial network (CDAN) (Long et al., 2018), and maximum classifier discrepancy (MCD) (Saito et al., 2018). HSI-oriented baselines include DCA (Zhang et al., 2021), UJADA (Tang et al., 2022), single-source domain expansion network (SDEnet) (Zhang et al., 2023), SCL-UDA (Li et al., 2023), MLUDA (Cai et al., 2024a), CPGAN (Xie et al., 2024), Cycle-ST (Li et al., 2024), and CACL (Ding et al., 2025). Public implementations are used when available to verify data conventions and optimization details, while methods without executable code are reproduced from their original objectives in the same pipeline. All results follow the common backbone and training protocol above, avoiding a mixture of published and rerun numbers.

Table 1: OA (%) on six directed cross-scene HSI adaptation tasks. Values are mean±standard deviation over five runs. Best and second-best means in each column are shown in bold and underline, respectively.

| Method | H13→H18 | H18→H13 | PU→PC | PC→PU | SH→HZ | HZ→SH | Avg. |
|---|---|---|---|---|---|---|---|
| Source only | <u>69.45±1.19</u> | 74.81±3.06 | 70.06±3.21 | 53.30±2.63 | 64.53±2.13 | 90.54±0.57 | 70.45 |
| DANN | 57.58±2.25 | 76.95±1.64 | 62.10±11.34 | 52.41±4.18 | 63.05±4.01 | 89.54±1.66 | 66.94 |
| DAN | 60.79±1.40 | 79.04±1.76 | 73.09±2.27 | **54.53±2.63** | 64.29±1.25 | 89.59±1.02 | 70.22 |
| JAN | 59.41±1.10 | 79.81±1.64 | 75.26±1.34 | 54.39±2.34 | 65.07±2.02 | 89.06±1.01 | 70.50 |
| CDAN | 50.60±8.32 | 74.46±1.95 | 66.25±2.81 | 47.40±5.57 | 62.12±4.86 | 90.74±0.76 | 65.26 |
| MCD | 68.36±1.47 | 75.72±2.71 | 67.75±3.09 | 51.97±3.75 | 64.44±1.57 | 90.88±0.82 | 69.85 |
| DCA | 59.34±1.07 | 79.75±1.45 | 74.55±1.87 | 53.44±3.44 | 64.70±2.41 | 90.30±0.70 | 70.35 |
| UJADA | 53.62±1.49 | 74.11±2.64 | 67.49±7.76 | 48.53±7.55 | 63.49±4.54 | 90.58±1.14 | 66.30 |
| SDEnet | 69.07±1.42 | 75.13±2.30 | 69.48±3.52 | 51.09±3.16 | 64.05±1.35 | 90.94±1.02 | 69.96 |
| SCL-UDA | 59.62±2.59 | 82.25±3.32 | 75.48±1.25 | 52.59±2.31 | 63.35±2.68 | 90.82±0.83 | 70.69 |
| MLUDA | 57.14±2.23 | <u>83.62±1.40</u> | <u>77.12±2.12</u> | 52.29±1.65 | <u>65.75±2.35</u> | <u>91.56±0.70</u> | <u>71.25</u> |
| CPGAN | 63.78±1.85 | 77.72±1.05 | 76.06±2.25 | <u>54.49±2.55</u> | 62.87±2.67 | 91.02±0.56 | 70.99 |
| Cycle-ST | 68.95±1.60 | 75.89±1.54 | 75.51±4.13 | 52.01±2.63 | 62.79±1.47 | 90.59±1.17 | 70.96 |
| CACL | **71.15±0.78** | 74.97±3.97 | 73.10±3.13 | 52.80±2.22 | 64.16±2.62 | 90.75±0.33 | 71.15 |
| SCCL | 64.97±3.00 | **85.74±3.12** | **80.87±2.57** | 53.72±4.57 | **67.35±5.27** | **92.17±0.49** | **74.14** |

For the quantitative comparison, Table 1 reports the OA results on the six directed cross-scene tasks. SCCL achieves the highest OA on H18→H13, PU→PC, SH→HZ, and HZ→SH, while CACL and DAN perform better on H13→H18 and PC→PU, respectively. Overall, SCCL obtains the best six-task average OA of 74.14%, compared with 71.25% for MLUDA. The differences between opposite transfer directions further show that adaptation difficulty depends on the source–target configuration, particularly for the Pavia and Shanghai–Hangzhou pairs. Fig. 3 presents the target classification maps on PU→PC for visual comparison. Compared with Source only and the competing methods, SCCL preserves more large homogeneous regions that are consistent with the ground truth and reduces visible class confusion in several local areas. Some errors remain near class boundaries and isolated regions, which is consistent with the quantitative results rather than suggesting perfect classification.

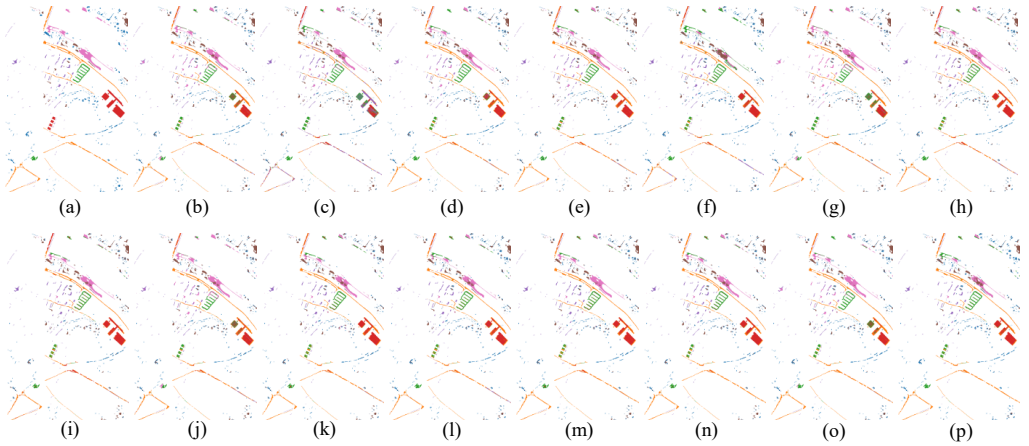



Figure 3: Target classification maps on PU→PC for a representative common split. All panels use the same seven-class color mapping.

### 4.4 Ablation Study

To further examine the contribution of each component in SCCL, we conduct ablation experiments on H13→H18, PU→PC, and SH→HZ under the same source split and five-run protocol. Specifically, the spectral-admissibility bound is removed, multi-view survival is replaced by a single-view check, the target-memory gate is disabled, the style modulation factor is fixed to one, and global correspondence normalization is replaced by local pairwise normalization. As shown in Table 2, all ablated variants yield lower average OA than the full model. The largest reduction occurs with local pairwise normalization, followed by single-view survival, indicating the importance of global correspondence organization and multi-view semantic persistence. Disabling the target-memory gate also leads to a clear decrease, while removing spectral admissibility or style modulation causes smaller but consistent performance drops. These results show that the five components provide complementary contributions to the final cross-scene adaptation performance.

Table 2: Component ablation results in OA (%) on the three forward tasks. The final column reports the mean over the three task means.

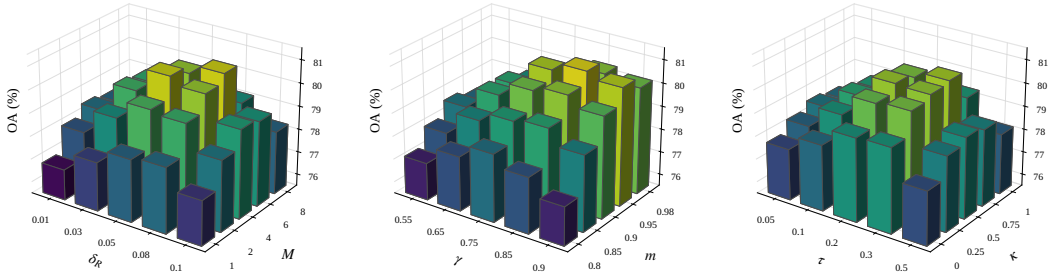
| Variant | H13→H18 | PU→PC | SH→HZ | Avg. |
|---|---|---|---|---|
| w/o target-memory gate | 60.92±3.41 | 76.74±3.02 | 62.91±5.66 | 66.86 |
| single-view survival | 58.83±3.71 | 74.98±3.26 | 60.84±6.02 | 64.88 |
| w/o spectral admissibility | 61.74±3.33 | 77.63±2.94 | 63.98±5.53 | 67.78 |
| w/o style modulation | 62.28±3.21 | 78.18±2.86 | 64.62±5.42 | 68.36 |
| local pairwise normalization | 57.96±3.86 | 73.92±3.39 | 59.73±6.18 | 63.87 |
| **Full SCCL** | **64.97**±3.00 | **80.87**±2.57 | **67.35**±5.27 | **71.06** |

### 4.5 Parameter Analysis

To further examine the influence of the main parameters in SCCL, we conduct pairwise sensitivity analysis on PU→PC using a fixed common source split. Two related parameters are varied in each panel of Fig. 4, while the remaining settings are kept at their default values. Specifically, $\delta_R \in \{0.01, 0.03, 0.05, 0.08, 0.10\}$ and $M \in \{1, 2, 4, 6, 8\}$ control spectral admissibility and the number of transported views, $\gamma \in \{0.55, 0.65, 0.75, 0.85, 0.90\}$ and $m \in \{0.80, 0.85, 0.90, 0.95, 0.98\}$ govern survival selection and target-memory updating, and $\tau \in \{0.05, 0.10, 0.20, 0.30, 0.50\}$ and $\kappa \in \{0, 0.25, 0.50, 0.75, 1.00\}$ determine affinity sharpness and scene-style modulation.

As can be observed in Fig. 4, the three parameter pairs exhibit clear but moderate variations in OA. For $(\delta_R, M)$, performance improves as the admissible range and number of views increase from small values, remains high around $\delta_R = 0.05$–$0.08$ and $M = 4$–$6$, and then decreases when either parameter becomes larger. A similar trend appears for $(\gamma, m)$, where higher values are concentrated

around $\gamma = 0.75$–$0.85$ and $m = 0.90$–$0.98$. The $(\tau, \kappa)$ surface is comparatively flatter, with favorable results around $\tau = 0.20$–$0.30$ and $\kappa = 0.50$–$0.75$. These observations support the default settings used in the experiments and show that neighboring parameter choices yield comparable performance rather than an isolated optimum.

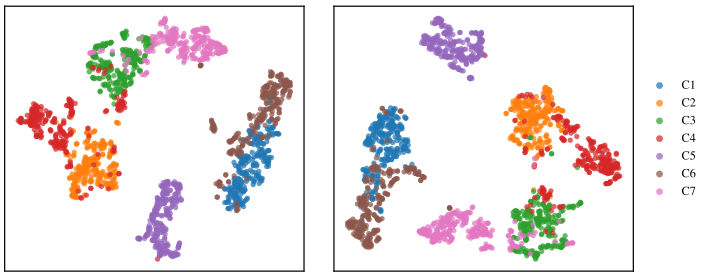



(a) $\delta_R$ and $M$ (b) $\gamma$ and $m$ (c) $\tau$ and $\kappa$

Figure 4: Pairwise parameter sensitivity on PU→PC using a fixed source split. Each three-dimensional bar plot jointly varies two parameters while fixing all remaining settings: (a) spectral-admissibility radius $\delta_R$ and number of admissible views $M$, (b) survival threshold $\gamma$ and memory momentum $m$, and (c) affinity temperature $\tau$ and scene-style modulation strength $\kappa$.

### 4.6 Visualization and Qualitative Analysis

To further examine the semantic organization learned by SCCL, we visualize the target embeddings on PU→PC using t-distributed stochastic neighbor embedding (t-SNE) in Fig. 5. The same labeled target samples are used only to color the two-dimensional points after training and are not involved in adaptation. As can be observed in Fig. 5(a), the source-only model contains several dispersed groups and noticeable overlap between neighboring classes. In contrast, Fig. 5(b) shows that SCCL forms more compact clusters for most categories and reduces inter-class mixing in several ambiguous regions, while a small amount of overlap remains for spectrally similar classes. These observations indicate that the proposed correspondence learning improves target-side semantic discrimination without producing artificially perfect separation.




(a) Source only (b) SCCL

Figure 5: The t-SNE visualization of PU→PC target embeddings for (a) Source only and (b) SCCL. Target labels are used only to color points after training.

## 5 Conclusion

In this paper, we propose the SCCL method for cross-scene HSI domain adaptation. The proposed method models scene variation through a bounded scene-style posterior and spectrally constrained stochastic transport, and identifies transferable source-target relations using multi-view survival, target-memory agreement, and global correspondence normalization. The resulting correspondences provide target semantic supervision while reducing the influence of scene-specific appearance changes. SCCL is evaluated on three public benchmark pairs with six directed transfer tasks, and the experimental results demonstrate its effectiveness and superiority over the compared methods. In future work, we will investigate more efficient correspondence construction and extend the framework to larger scene shifts and more diverse cross-scene settings.

## REFERENCES


Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In *Proceedings of the International Conference on Learning Representations*, 2022.

Mingshuo Cai, Bobo Xi, Jiaojiao Li, Shou Feng, Yunsong Li, Zan Li, and Jocelyn Chanussot. Mind the gap: Multilevel unsupervised domain adaptation for cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024a.

Yichao Cai, Yuhang Liu, Zhen Zhang, and Javen Qinfeng Shi. Clap: Isolating content from style through contrastive learning with augmented prompts. In *Proceedings of the European Conference on Computer Vision*, 2024b.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9650–9660, 2021.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning*, pp. 1597–1607, 2020.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Proceedings of the Conference on Neural Information Processing Systems*, volume 26, pp. 2292–2300, 2013.

Junzhe Dang, Chengwang Guo, Mengmeng Zhang, Yuxiang Zhang, Wen Jia, and Wei Li. Cross-scene hyperspectral image classification via bidirectional mamba and domain mixing network. *IEEE Transactions on Neural Networks and Learning Systems*, 37(8):3618–3631, 2026.

Kexing Ding, Ting Lu, Wei Fu, and Leyuan Fang. Cross-scene hyperspectral image classification with consistency-aware customized learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(1):418–430, 2025.

Puhong Duan, Shiyu Jin, Xiaotian Lu, Lianhui Liang, Xudong Kang, and Antonio Plaza. Domain-adaptive mamba for cross-scene hyperspectral image classification. *IEEE Transactions on Image Processing*, 35:1206–1217, 2026.

Zhuoqun Fang, Wenqiang He, Zhaokui Li, Qian Du, and Qiusheng Chen. Masked self-distillation domain adaptation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–20, 2024.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9729–9738, 2020.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022.

Yuhang Hong, Zhixi Feng, Shuyuan Yang, and Zhihao Chang. Cross-scene hyperspectral image classification network with dynamic perturbation and self-knowledge distillation. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025.

Zhaokui Li, Qiang Xu, Li Ma, Zhuoqun Fang, Yan Wang, Wenqiang He, and Qian Du. Supervised contrastive learning-based unsupervised domain adaptation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–17, 2023.

Zhongwei Li, Yajie Yang, Leiquan Wang, Mingming Xu, Ziqi Xin, Jie Wei, and Yuewen Wang. Cycle self-training with joint adversarial for cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024.

Rong Liu, Mengqing Zhou, Jiaqi Yang, and Junjue Wang. Adaptive smooth adversarial learning for cross-domain hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–17, 2025.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I. Jordan. Learning transferable features with deep adaptation networks. In *Proceedings of the International Conference on Machine Learning*, pp. 97–105, 2015.

Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I. Jordan. Deep transfer learning with joint adaptation networks. In *Proceedings of the International Conference on Machine Learning*, pp. 2208–2217, 2017.

Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Conditional adversarial domain adaptation. In *Proceedings of the Conference on Neural Information Processing Systems*, 2018.

Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3723–3732, 2018.

Boshan Shi, Guo Cao, Youqiang Zhang, Yanbo Liu, and Kai Yang. Fourier-based spectral-spatial generator for cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–17, 2025.

Xuebin Tang, Chunchao Li, and Yuanxi Peng. Unsupervised joint adversarial domain adaptation for cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

Yuhao Xie, Ronghua Shang, Jinhong Ren, Jingyu Zhong, Jie Feng, and Songhua Xu. Source-driven global–local prototype contrastive learning for hyperspectral domain adaptation. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–15, 2026.

Zhuojun Xie, Puhong Duan, Xudong Kang, Wang Liu, and Shutao Li. Classwise prototype-guided alignment network for cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–11, 2024.

Kai Xu, Zhou Zhu, Bocai Wu, and Chengcheng Fan. Uncertainty-informed prototype contrastive learning for cross-scene hyperspectral image classification. *Knowledge-Based Systems*, 311: 113165, 2025.

Cuiyu Yang, Lin Gao, Yongxin Feng, and Junchang Xin. Topology-aligned gated interaction network for cross-domain hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–21, 2026.

Kai Yang, Boshan Shi, Guo Cao, and Youqiang Zhang. Fourier phase preservation and spectral random augmentation for cross-scene hyperspectral image classification with ensemble networks. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–17, 2025.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *Proceedings of the International Conference on Machine Learning*, pp. 12310–12320, 2021.

An Zhang, Han Wang, Xiang Wang, and Tat-Seng Chua. Disentangling masked autoencoders for unsupervised domain generalization. In *Proceedings of the European Conference on Computer Vision*, 2024.

Yuxiang Zhang, Wei Li, Ran Tao, Jiangtao Peng, Qian Du, and Zongwei Cai. Cross-scene hyperspectral image classification with discriminative cooperative alignment. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11):9646–9660, 2021.

Yuxiang Zhang, Wei Li, Weidong Sun, Ran Tao, and Qian Du. Single-source domain expansion network for cross-scene hyperspectral image classification. *IEEE Transactions on Image Processing*, 32:1498–1512, 2023.

Yuxiang Zhang, Wei Li, Wen Jia, Mengmeng Zhang, Ran Tao, and Shunlin Liang. Cross-domain hyperspectral image classification based on bi-directional domain adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(12):12038–12051, 2025.

Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *Proceedings of the International Conference on Learning Representations*, 2021.

Kaiyang Zhou, Chen Change Loy, and Ziwei Liu. Semi-supervised domain generalization with stochastic stylematch. *International Journal of Computer Vision*, 131(9):2377–2387, 2023.