LEARNING WHAT AND WHEN TO TRANSFER FOR CROSS-SCENE HYPERSPECTRAL CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-scene clustering of hyperspectral imagery (HSI) asks an unlabeled source scene to help organize an unlabeled target scene into consistent land-cover groups despite scene-dependent spectral shifts. Existing methods typically align scene representations as a whole, although transferable land-cover semantics and scene-specific spectral variation may remain entangled. This makes it difficult to determine both what structure should be transferred and when target evidence should enter the transfer process. We propose disentangled transfer with recurrent admission (DTRA), which treats these two decisions as coupled parts of one recurrent transfer process. To learn *what to transfer*, it factorizes each representation into semantic and scene-specific branches, tests whether semantic assignments are preserved under cross-scene variation exchange, and summarizes the preserved semantic structure by balanced prototypes. These prototypes are then matched across scenes to initialize the target-oriented transfer state. To learn *when to transfer*, assignment confidence is combined with the exchange-induced assignment discrepancy to rank target regions and admit them progressively. The terminal target-oriented prototype state is carried into the next training round, so an admission decision also changes the prototype geometry used to judge subsequent transfer content. We further provide an assignment-preservation analysis that connects semantic deviation under cross-scene variation exchange to the assignment discrepancy reused by recurrent admission, and formulate training as recurrent prototype-state refinement under a unified clustering objective. Experiments on six directional tasks from Houston, Pavia, and HyRANK have conducted to validate the effectiveness and superiority of our DTRA method. The code is released at `https://github.com/Sky-byte-box/DTRA_code`.

1 INTRODUCTION

As a form of spectral imaging data, hyperspectral imagery (HSI) records dense spectral measurements over numerous contiguous bands and provides rich material-sensitive information for land-cover analysis (Liao & Wang, 2026; He et al., 2026). These data are often acquired under different illumination, atmospheric, seasonal, sensor, and geometric conditions, under which the same land-cover category may exhibit different spectral responses across scenes (Shang et al., 2026; Wang et al., 2026). Such scene-dependent shifts can obscure category correspondence and distort clustering when acquisition-related variation is mistaken for semantic structure. Among HSI-related unsupervised learning problems, cross-scene hyperspectral clustering aims to derive a common land-cover partition from multiple unlabeled scenes despite these spectral discrepancies. Since neither scene provides annotations, transferable semantics must be inferred directly from the observations while limiting the influence of scene-specific variation.

Cross-scene HSI classification has addressed spectral distribution shifts by transferring discriminative representations between scenes. Recent methods reduce source–target discrepancy through pixel-aware decoupled alignment and bidirectional adaptation (Li et al., 2025; Zhang et al., 2025), while domain-adaptive sequence modeling captures fine-grained cross-scene dependencies (Duan et al., 2026). Domain-generalization approaches enlarge source variability through Fourier and spectral augmentation (Yang et al., 2025). Other studies exploit spectral-coordinate modeling or transformer representations (Li et al., 2024; Ye et al., 2024), and prototype-guided methods establish category correspondence in the learned feature space (Xie et al., 2024; Ding et al., 2025a). These approaches

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.

primarily address classification with labeled source samples. In cross-scene clustering, neither scene supplies annotations, so both the category partition and cross-scene cluster correspondence must instead be inferred from unlabeled observations.

Generic deep clustering methods couple representation learning with latent assignments (Xie et al., 2016; Caron et al., 2020). Unsupervised HSI clustering provides a natural basis for learning partitions without annotations. Graph- and anchor-based methods organize samples by exploiting spectral–spatial relationships (Wang et al., 2019; Chen et al., 2023; Zhang et al., 2024), whereas deep clustering methods improve feature discrimination through neighborhood contrast, graph filtering, and transformer representations (Cai et al., 2022; Luo et al., 2024; Ding et al., 2025c). Pixel–superpixel contrast with pseudo-label correction and superpixel contrastive clustering with semantic-invariant augmentations provide further examples of unlabeled HSI partition learning (Guan et al., 2024; Qi et al., 2024). Although effective for within-scene clustering, these approaches generally operate under a single spectral distribution and therefore do not explicitly resolve cluster correspondence when scene responses change. More recently, anchor-guided discriminative subspace alignment and clustering (ADSAC) extends clustering to a fully unlabeled cross-scene setting by deriving source clusters, aligning source and target subspaces, and inferring target assignments (Zhang et al., 2026). However, its transferred structure is still defined through subspace alignment rather than an explicit separation of shared semantics and scene-dependent spectral variation.

We should note that the methods discussed above have improved cross-scene feature alignment and unlabeled partition learning. However, a lower source–target discrepancy does not necessarily show that the aligned structure reflects shared land-cover semantics rather than scene-dependent spectral variation. In other words, a target assignment can appear confident while still depending on illumination or sensor response. If such evidence enters prototype refinement too early, incorrect cross-scene correspondence may be reinforced in later rounds. Fig. 1 illustrates how this ambiguity can lead to a scene-biased target partition under direct alignment. The central questions are *what* semantic structure should be transferred and *when* target evidence should influence it.

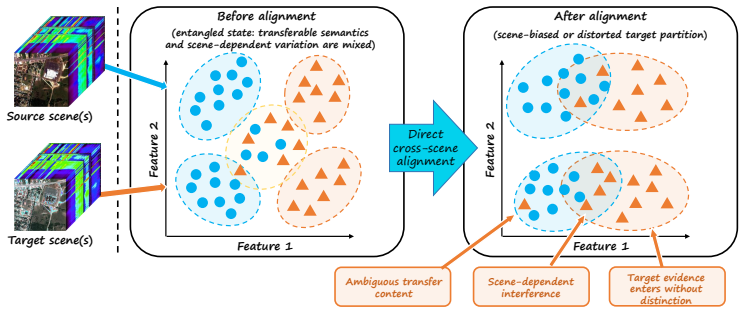


Figure 1: Motivation. When transferable semantics and scene-dependent variation remain entangled, direct cross-scene alignment can propagate ambiguous content and produce a scene-biased target partition, leaving both transfer content and target participation unresolved.

To address these issues, we propose DTRA, a recurrent framework for cross-scene hyperspectral clustering, as shown in Fig. 2. It separates semantic and scene-specific representations and tests whether semantic assignments are preserved under cross-scene variation exchange. Balanced prototypes summarize the learned semantic structure, and entropy-regularized transport aligns them to initialize a target-oriented state. Assignment uncertainty and exchange-induced discrepancy then order target samples for progressive admission into prototype refinement. The selectively retained state is passed to the next training round, allowing admitted target evidence to affect subsequent estimates of transferable structure. Experiments on six directional tasks from three benchmark HSI datasets assess clustering performance and the contribution of each component.

The main contributions of this work are summarized as follows. (1) We develop a transfer-content learning strategy that disentangles semantic and scene-specific factors and identifies transferable cross-scene structure through assignment preservation under variation exchange. (2) We introduce a recurrent admission mechanism that progressively incorporates reliable target evidence and feeds the selectively retained prototype state back to subsequent transfer-content estimation. (3) We provide a theoretical characterization of assignment preservation under cross-scene variation exchange, which supports the transferability criterion adopted in DTRA.

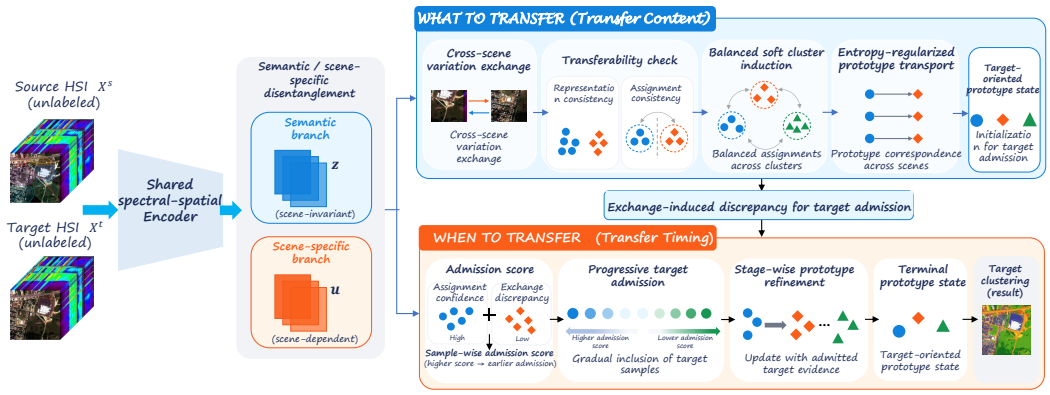


Figure 2: Overall framework of DTRA for cross-scene hyperspectral clustering.

2 Related Work

2.1 Disentangled representation learning

Disentangled representation learning seeks to separate task-relevant factors from domain-dependent variation. Domain-generalization studies have examined how to isolate factors that support transfer across changing environments (Zhang et al., 2022), while domain separation networks partition features into shared and private components and use them to reconstruct observations (Bousmalis et al., 2016). This distinction is pertinent to cross-scene HSI, where illumination, sensor response, and acquisition conditions can alter spectral measurements without changing land-cover identity. Yet separating latent factors does not by itself establish whether representations learned from two unlabeled scenes retain a common category structure. Prototype representations can summarize latent groups, but their correspondence across scenes remains to be determined.

2.2 Selective adaptation

Selective adaptation controls when target observations influence an evolving model. Confidence- and uncertainty-aware methods suppress unreliable updates, while anytime adaptation incorporates target information progressively instead of waiting for a complete adaptation cycle (Wang et al., 2021; Niu et al., 2022; 2023; Charkraborty et al., 2023). Progressive feature alignment has also used an easy-to-hard transfer strategy in unsupervised domain adaptation (Chen et al., 2019). Such strategies reduce the effect of uncertain target samples, but their admission criteria are usually defined independently of the evidence used to determine transfer content. In cross-scene clustering, a confident assignment may still depend strongly on scene-specific variation. The relation between target admission and transferable cluster structure therefore remains insufficiently addressed.

3 Method

3.1 Problem Formulation and Overview

Given two unlabeled HSI scenes, cross-scene clustering seeks a common partition that remains semantically consistent across scene-dependent spectral changes. Let $X^s = \{x_i^s\}_{i=1}^{N_s} \subset \mathbb{R}^D$ and $X^t = \{x_j^t\}_{j=1}^{N_t} \subset \mathbb{R}^D$ denote the source and target samples or superpixel representatives, where N_s and N_t are the corresponding sample numbers and D is the input feature dimensionality. Let C be the number of shared latent land-cover categories, and define $N_d = N_s$ for $d = s$ and $N_d = N_t$ for $d = t$. The desired mapping is $(\hat{Y}^s, \hat{Y}^t) = \mathcal{F}(X^s, X^t)$, where $\hat{Y}^d \in \{1, \dots, C\}^{N_d}$ for $d \in \{s, t\}$. To separate shared semantic information from scene-dependent variation, each observation is encoded into two latent components as follows:

$$(z_i^d, u_i^d) = E_\theta(x_i^d), \qquad z_i^d \in \mathbb{R}^{d_z},\ u_i^d \in \mathbb{R}^{d_u}, \tag{1}$$

where E_θ is the shared encoder, z_i^d and u_i^d denote the semantic and scene-specific representations, and d_z and d_u are their latent dimensions. A decoder D_ϕ reconstructs the observation as $\hat{x}_i^d =$

$D_\phi(z_i^d, u_i^d)$. DTRA treats transfer content and target participation as coupled decisions. Cross-scene variation exchange tests whether semantic assignments are preserved, while progressive admission determines which target samples can refine the transported prototypes. The resulting target-oriented prototype state is reused in the next training round, allowing admission decisions to affect subsequent assignments and prototype correspondence. Fig. 2 summarizes this recurrent process.

3.2 Learning What to Transfer

Separating scene-specific variation. The separation of shared and domain-specific components has precedent in domain adaptation (Bousmalis et al., 2016). Here, for each scene, the encoder–decoder pair preserves the observation while reducing local dependence between the two representations. Reconstruction and decorrelation are imposed as follows:

$$D_\phi(z_i^d, u_i^d) \simeq x_i^d, \qquad \frac{\bar{Z}_d^\top \bar{U}_d}{n_d - 1} \simeq 0, \quad d \in \{s, t\}, \tag{2}$$

where n_d is the number of samples from scene d in the current minibatch, $Z_d \in \mathbb{R}^{n_d \times d_z}$ and $U_d \in \mathbb{R}^{n_d \times d_u}$ collect the two representations, and $\bar{Z}_d$ and $\bar{U}_d$ are their centered forms. A scene discriminator g_ω is additionally applied to the latent codes. The semantic representation is driven toward an uninformative scene posterior, $g_\omega(z_i^d) \simeq [\frac{1}{2}, \frac{1}{2}]$, whereas the scene-specific representation is encouraged to predict its scene through $[g_\omega(u_i^d)]_d \to 1$. Here, $\simeq$ denotes a relation enforced approximately during training.

Testing transferability across scenes. A semantic representation should retain its cluster identity when only scene-specific variation changes. The exchanged codes are decoded and re-encoded by the operations given below:

$$\begin{aligned} \tilde{x}_i^{s\leftarrow t} &= D_\phi(z_i^s, u_{\pi_t(i)}^t), & \tilde{x}_j^{t\leftarrow s} &= D_\phi(z_j^t, u_{\pi_s(j)}^s), \\ (\tilde{z}_i^{s\leftarrow t}, \tilde{u}_i^{s\leftarrow t}) &= E_\theta(\tilde{x}_i^{s\leftarrow t}), & (\tilde{z}_j^{t\leftarrow s}, \tilde{u}_j^{t\leftarrow s}) &= E_\theta(\tilde{x}_j^{t\leftarrow s}), \end{aligned} \tag{3}$$

where π_t and π_s randomly permute target and source minibatch indices, respectively, and the arrows indicate the direction of scene-specific code exchange. Let $P = [p_1, \ldots, p_C]^\top$ denote the semantic prototypes. Their normalized similarities are converted into assignment probabilities as follows:

$$q(c \mid z) = \frac{\exp(\langle \bar{z}, \bar{p}_c \rangle / \tau)}{\sum_{k=1}^{C} \exp(\langle \bar{z}, \bar{p}_k \rangle / \tau)}, \tag{4}$$

where $\bar{z}$ and $\bar{p}_c$ are ℓ_2-normalized and $\tau > 0$ is the assignment temperature. For a sample from scene d, we record the assignment discrepancy $\delta_i^d = \mathrm{JS}(q(\cdot \mid z_i^d), q(\cdot \mid \tilde{z}_i^{d\leftarrow\bar{d}}))$, where $\mathrm{JS}(\cdot,\cdot)$ denotes the Jensen–Shannon divergence, $\bar{s} = t$, and $\bar{t} = s$. A small δ_i^d indicates that the semantic assignment is preserved after the scene-specific code is changed. This discrepancy is reused for target admission in Section 3.3.

Because the same exchange-induced discrepancy is used to judge transfer content and to regulate target admission, it should reflect semantic preservation under scene-specific changes rather than serve as an isolated empirical score. The following proposition formalizes this task-specific connection.

Proposition 1 (Assignment preservation under cross-scene variation exchange)**.** *Let $A \in \mathbb{R}^{C \times d_z}$ have rows $\bar{p}_c^\top$, and suppose cross-scene variation exchange yields $\|\bar{\tilde{z}} - \bar{z}\|_2 \le \epsilon$ for some $\epsilon \ge 0$. For the assignment map in Eq. (4), the resulting deviation bounds are given below:*

$$\|q(\cdot \mid z) - q(\cdot \mid \tilde{z})\|_2 \le \frac{\|A\|_2}{2\tau}\epsilon, \qquad \mathrm{JS}(q(\cdot \mid z), q(\cdot \mid \tilde{z})) \le \frac{Ce^{2/\tau}\|A\|_2^2}{16\tau^2}\epsilon^2, \tag{5}$$

where $\|A\|_2$ denotes the spectral norm and $\mathrm{JS}(\cdot,\cdot)$ uses natural logarithms.

Proof. Writing Eq. (4) as $q = \mathrm{softmax}(A\bar{z}/\tau)$, the softmax Jacobian has spectral norm at most $1/2$. The mean-value theorem gives the first bound in Eq. (5). Because normalized similarities lie in $[-1, 1]$, every assignment probability is at least $m_\tau = e^{-2/\tau}/C$. Let $a = q(\cdot \mid z)$, $b = q(\cdot \mid \tilde{z})$, and $m = (a + b)/2$. The same lower bound applies to each component of m. Using the standard bound $\mathrm{KL}(a\|m) \le \sum_c (a_c - m_c)^2/m_c$, where KL denotes the Kullback–Leibler divergence, gives $\mathrm{JS}(a, b) \le \|a - b\|_2^2/(4m_\tau)$. Substituting the first bound yields the second. □

Proposition 1 gives δ_i^d a direct role in cross-scene clustering: a small exchange response indicates that a semantic assignment is insensitive to scene-specific variation, supporting its use to identify transferable structure and, in Section 3.3, to determine when target evidence can participate.

Forming transferable semantic structure. The semantic structure is summarized by balanced prototypes rather than transferred as a complete feature representation. For scene d, let $P_d = [p_1^d, \dots, p_C^d]^\top$ and $S_{ic}^d = \langle \bar{z}_i^d, \bar{p}_c^d \rangle / \tau$. To limit concentration on dominant clusters, we adopt a balanced assignment principle used in self-labelling without annotations (Asano et al., 2020) and solve the following problem:

$$\begin{aligned} Q_d^\star &= \arg\max_{Q \in \mathcal{U}_d} \langle Q, S_d \rangle_F + \varepsilon H(Q), \\ \mathcal{U}_d &= \left\{ Q \geq 0 \;\middle|\; Q\mathbf{1}_C = \frac{1}{N_d}\mathbf{1}_{N_d},\ Q^\top \mathbf{1}_{N_d} = \frac{1}{C}\mathbf{1}_C \right\}, \end{aligned} \tag{6}$$

where $Q \in \mathbb{R}_+^{N_d \times C}$ is the soft assignment matrix, $\langle \cdot, \cdot \rangle_F$ is the Frobenius inner product, $H(Q) = -\sum_{i,c} Q_{ic} \log Q_{ic}$, $\varepsilon > 0$ controls entropy regularization, and $\mathbf{1}_m$ is the m-dimensional all-ones vector. The prototype of cluster c is computed as follows:

$$p_c^d = \frac{\sum_{i=1}^{N_d} Q_{ic}^{d\star} z_i^d}{\sum_{i=1}^{N_d} Q_{ic}^{d\star} + \epsilon_0}, \tag{7}$$

where $\epsilon_0 > 0$ avoids a zero denominator. The assignments are optimized through the clustering objective below:

$$\mathcal{R}_{\mathrm{clu}}(\Psi, Q) = -\sum_{d \in \{s,t\}} \sum_{i=1}^{N_d} \sum_{c=1}^{C} Q_{ic}^{d\star} \log q(c \mid z_i^d), \tag{8}$$

where $\Psi = \{\theta, \phi, \omega\}$ and $Q = \{Q_s^\star, Q_t^\star\}$. Cross-scene correspondence is established between the resulting prototypes. For $a, b \in \{1, \dots, C\}$, let $M = [M_{ab}] \in \mathbb{R}^{C \times C}$ with $M_{ab} = 1 - \langle \bar{p}_a^s, \bar{p}_b^t \rangle$. Following entropy-regularized transport (Cuturi, 2013), the prototype correspondence and barycentric mapping are given below:

$$\Pi^\star = \arg\min_{\Pi \in \mathcal{U}_\Pi} \langle \Pi, M \rangle_F - \epsilon_\pi H(\Pi), \qquad \tilde{p}_b^{s \to t} = \frac{\sum_{a=1}^{C} \Pi_{ab}^\star p_a^s}{\sum_{a=1}^{C} \Pi_{ab}^\star + \epsilon_0}, \tag{9}$$

where $\Pi \in \mathbb{R}_+^{C \times C}$ is the transport plan, $\mathcal{U}_\Pi = \{\Pi \geq 0 \mid \Pi \mathbf{1}_C = \frac{1}{C}\mathbf{1}_C,\ \Pi^\top \mathbf{1}_C = \frac{1}{C}\mathbf{1}_C\}$, $H(\Pi) = -\sum_{a,b} \Pi_{ab} \log \Pi_{ab}$, and $\epsilon_\pi > 0$ controls transport entropy. We denote the transported prototype set by $\tilde{P}^{s \to t} = [\tilde{p}_1^{s \to t}, \dots, \tilde{p}_C^{s \to t}]^\top$. It initializes the target-oriented state refined in Section 3.3.

3.3 Learning When to Transfer

Target prototypes are refined first by samples whose assignments are confident and insensitive to scene-specific variation. For target sample j, let $q_{jc} = q(c \mid z_j^t)$, $h_j = -\sum_c q_{jc} \log q_{jc}$, and $\delta_j = \mathrm{JS}(q(\cdot \mid z_j^t), q(\cdot \mid \tilde{z}_j^{t \leftarrow s}))$. These quantities are combined into the admission score as follows:

$$r_j = \exp\left[-\alpha \frac{h_j}{\log C} - \beta \frac{\delta_j}{\delta_j + 1} \right] \in (0, 1], \tag{10}$$

where $\alpha, \beta > 0$ control assignment uncertainty and exchange discrepancy. These scores order the target samples for progressive admission. At stage $k \in \{1, \dots, K\}$, the fraction γ_k with the highest scores is selected by the criterion below:

$$\mathcal{A}_k = \left\{ j : r_j \geq \mathrm{Quantile}\Big(\{r_l\}_{l=1}^{N_t}, 1 - \gamma_k \Big) \right\}, \quad 0 < \gamma_1 < \cdots < \gamma_K = 1, \tag{11}$$

where K is the number of stages, $\mathcal{A}_k$ is the admitted target subset, and $\mathrm{Quantile}(\cdot, \cdot)$ denotes the empirical quantile operator. Once $\mathcal{A}_k$ is determined, its target evidence is aggregated into the stage-specific prototype estimate given below:

$$\mu_{c,k}^t = \frac{\sum_{j \in \mathcal{A}_k} r_j Q_{jc}^{t\star} z_j^t}{\sum_{j \in \mathcal{A}_k} r_j Q_{jc}^{t\star} + \epsilon_0}, \tag{12}$$

where $\mu_{c,k}^t$ is the target estimate for cluster c, weighted by the admission scores. Rather than replacing the transported prototype directly, this estimate is incorporated through the controlled update:

$$P_{c,k} = (1-\eta_k)P_{c,k-1} + \eta_k \mu_{c,k}^t, \qquad P_{c,0} = \tilde{p}_c^{s\to t}, \tag{13}$$

where $P_{c,k}$ denotes the target-oriented prototype of cluster c after stage k, and $\eta_k \in (0,1]$ controls the contribution of newly admitted evidence. After the final stage, $\hat{y}_j^t = \arg\max_c \langle \bar{z}_j^t, \bar{P}_{c,K} \rangle$, with $\bar{P}_{c,K} = P_{c,K}/\|P_{c,K}\|_2$. The terminal state $P_{c,K}$ is carried to the next round.

3.4 Coupling What and When

The two decisions are coupled through the prototype state passed between training rounds. Let $\mathcal{P}^{(e)}$ denote the target-oriented state at round e, initialized from Eq. (9). It determines the target prototypes used in Eq. (6), while Eqs. (10)–(13) produce $\mathcal{P}^{(e+1)} = \{P_{c,K}^{(e)}\}_{c=1}^C$. The admission result from one round therefore changes the semantic reference used to estimate transfer content in the next. Omitting round superscripts in the preceding derivations for readability, we denote this transition by Γ_Ψ, where $\Psi = \{\theta, \phi, \omega\}$ and $Q^{(e)} = \{Q_s^{\star(e)}, Q_t^{\star(e)}\}$. Training is summarized as follows:

$$\left(Q^{(e)}, \mathcal{P}^{(e+1)}\right) = \Gamma_\Psi\left(X^s, X^t; \mathcal{P}^{(e)}\right), \qquad \Psi^\star = \arg\min_\Psi \ \mathcal{R}_{\mathrm{clu}}\left(\Psi, Q^{(e)}\right). \tag{14}$$

The state transition supplies the recurrent feedback, while $\mathcal{R}_{\mathrm{clu}}$ updates the learnable parameters under the factorization relations in Eq. (2). To prevent an inferior recurrent state from propagating, the candidate state is retained only when the unlabeled target-assignment confidence improves by at least a fixed margin ρ. Otherwise, the preceding state is reused. This screening uses no ground-truth labels. Algorithm 1 summarizes the training and stage-wise inference procedure.

Algorithm 1 Training and stage-wise inference of DTRA

Require: Unlabeled scenes X^s, X^t, cluster count C, epochs T, admission fractions $\{\gamma_k\}_{k=1}^K$, screening margin ρ

Initialize E_θ, D_ϕ, g_ω, P_s, P_t, $\Psi = \{\theta, \phi, \omega\}$, and prototype state $\mathcal{P}^{(1)}$
for $e = 1$ to T **do**
  Initialize the target prototypes with $\mathcal{P}^{(e)}$ and encode the sampled minibatches using Eq. (1)
  Enforce the factorization relations in Eq. (2), exchange scene-specific codes, and re-encode samples by Eq. (3)
  Compute assignments by Eq. (4) and the sample-wise assignment discrepancy δ_i^d
  Obtain $Q_s^\star, Q_t^\star$, update P_s, P_t, and evaluate $\mathcal{R}_{\mathrm{clu}}$ by Eqs. (6)–(8)
  Compute $\Pi^\star$ and $\tilde{P}^{s\to t}$ by Eq. (9), then $\{r_j\}_{j=1}^{N_t}$ by Eq. (10)
  for $k = 1$ to K **do**
    Form $\mathcal{A}_k$, estimate $\mu_{c,k}^t$, and update $P_{c,k}$ by Eqs. (11)–(13)
  end for
  Retain the candidate $\{P_{c,K}\}_{c=1}^C$ as $\mathcal{P}^{(e+1)}$ only when the unlabeled screening condition is met. Otherwise, reuse $\mathcal{P}^{(e)}$
  Update Ψ with $\mathcal{R}_{\mathrm{clu}}$ according to Eq. (14)
end for
return final source and target clusters

4 Experiments

4.1 Experimental setup

Datasets. We evaluate six directional cross-scene tasks on three benchmark datasets: Houston, Pavia, and HyRANK. The aligned Houston2013 (H13) and Houston2018 (H18) overlap crops each contain 210×954 pixels, 48 retained spectral bands, and seven shared classes. Pavia University (PU) and Pavia Center (PC) contain 610×340 and 1096×715 pixels, respectively, with 102 spectral bands and seven shared classes. The HyRANK scenes Dioni (Di) and Loukia (Lo) contain 250×1376 and 249×945 pixels, respectively, with 176 spectral bands. The 12 classes shared by the two scenes are retained. The resulting tasks are H13→H18, H18→H13, PU→PC, PC→PU, Di→Lo, and Lo→Di. Ground-truth labels are excluded from representation learning, clustering, prototype matching, admission, and model selection. They are loaded only after prediction for Hungarian-matched ACC and display-time color matching. All reported comparison results are averaged over

five independent runs under fixed unlabeled selection rules, and ground-truth accuracy is never used by the recurrent state-screening rule.

Implementation Details. We adopt the same label-free pipeline for all six cross-scene tasks, including recurrent transfer, target-oriented prototype construction, progressive admission, and final clustering. The default values of the four parameters analyzed in Section 4.4 are $\tau = 0.20$, $\epsilon_\pi = 0.05$, $K = 4$, and $\beta = 1.0$. The uncertainty weight is fixed at $\alpha = 1.0$, and the recurrent state-screening margin is fixed at $\rho = 0.25$ throughout all experiments. The balanced-assignment entropy coefficient ε, admission fractions $\{\gamma_k\}_{k=1}^K$, and prototype-update coefficients $\{\eta_k\}_{k=1}^K$ are shared across all tasks. Only the initialization structure is adapted to the characteristics of individual scene pairs. H13→H18 uses eight cluster-specific sub-prototypes with soft-3 aggregation, while PC→PU uses a 12-dimensional representation with three-way over-clustered initialization. H18→H13, PU→PC, Di→Lo, and Lo→Di use the standard initialization without additional task-specific adjustments.

4.2 Comparison with SOTA Approaches

Competing Approaches. To evaluate the performance of DTRA, nine representative HSI clustering approaches are compared, including SGCNR (Wang et al., 2019), S3AGC (Chen et al., 2023), BGPC (Zhang et al., 2024), SAPC (Jiang et al., 2025), NCSC (Cai et al., 2022), SDST (Luo et al., 2024), AHSGC (Ding et al., 2025b), SLCGC (Ding et al., 2025c), and ADSAC (Zhang et al., 2026). SGCNR performs scalable graph clustering through nonnegative relaxation, whereas S3AGC builds a spectral–spatial superpixel anchor graph to organize the HSI samples. BGPC introduces local-region guidance into bipartite graph projected clustering, and SAPC learns structured anchors for large-scale projected clustering. NCSC combines superpixel contraction with neighborhood contrastive subspace learning, while SDST employs a self-supervised double-structure transformer to capture complementary clustering cues. AHSGC learns an adaptive homophily graph with graph filtering, and SLCGC uses low-pass graph filtering and self-supervised contrastive learning in a lightweight clustering framework. ADSAC is a recent cross-scene method that derives source clusters and establishes source–target correspondence through anchor-guided discriminative subspace alignment. The parameter settings of the competing approaches follow their original papers or released implementations whenever available.

Table 1: ACC (%) on six directional tasks under the controlled experimental protocol. Values are mean±standard deviation. Bold denotes the highest mean in each column.

Method	H13→H18	H18→H13	PU→PC	PC→PU	Di→Lo	Lo→Di
SGCNR	37.98±4.00	51.58±6.38	60.06±1.98	48.61±3.42	33.80±3.15	33.12±3.06
S3AGC	57.92±0.00	18.85±0.00	54.54±0.01	56.25±0.06	39.27±1.13	38.03±2.50
BGPC	40.08±4.62	48.77±4.76	60.09±1.53	49.85±2.82	32.74±1.46	35.48±1.99
SAPC	55.22±0.00	18.85±0.00	68.05±0.00	56.57±0.00	41.38±0.07	41.10±3.21
NCSC	40.55±2.77	40.07±0.56	70.75±0.03	66.08±0.01	46.40±0.35	52.36±0.97
SDST	41.90±1.65	40.52±1.96	70.72±0.00	66.27±0.11	47.01±0.82	53.68±0.11
AHSGC	40.48±2.33	39.66±0.48	70.90±0.02	66.10±0.02	46.43±0.83	53.03±0.16
SLCGC	42.34±3.33	38.96±0.12	71.09±0.00	62.46±4.69	45.96±1.65	44.50±1.34
ADSAC	45.80±5.37	34.46±0.69	61.48±3.31	43.75±0.52	47.38±3.18	53.84±1.06
DTRA	**68.96±1.12**	**52.21±7.04**	**72.37±0.03**	**66.76±4.32**	**51.84±1.64**	**62.40±1.89**

Comparison Results. Table 1 presents the quantitative ACC results of all compared approaches on the six directional cross-scene tasks. As shown in Table 1, DTRA obtains the highest mean ACC on all six evaluated tasks. The improvements are particularly clear on H13→H18 and Lo→Di, where the ACC values reach 68.96% and 62.40%, exceeding the corresponding second-best results by 11.04 and 8.56 percentage points, respectively. On PU→PC, DTRA achieves 72.37%, compared with 71.09% for SLCGC, 70.90% for AHSGC, 70.75% for NCSC, and 70.72% for SDST. The margins on H18→H13 and PC→PU are comparatively smaller, indicating that the degree of improvement varies with the transfer direction. In addition, Fig. 3 shows the clustering maps of all compared methods on the PU→PC task. Compared with the ground truth in Fig. 3(b), SGCNR, S3AGC, BGPC, SAPC, and ADSAC exhibit more scattered assignments or larger local deviations in several spatial regions, whereas NCSC, SDST, AHSGC, and SLCGC recover most of the principal structures, although local

fragmentation and boundary mismatches can still be observed. The result of DTRA in Fig. 3(l) preserves the dominant class regions and reduces part of these local inconsistencies, yielding a spatial partition that is generally closer to the ground truth. Nevertheless, the visual differences from the stronger baselines remain modest in some local regions, which agrees with the smaller quantitative margin observed for PU→PC in Table 1. Overall, the quantitative and qualitative results consistently show that DTRA improves the cross-scene clustering performance across the six transfer directions, while the gain becomes smaller when the competing methods already recover a strong target partition.

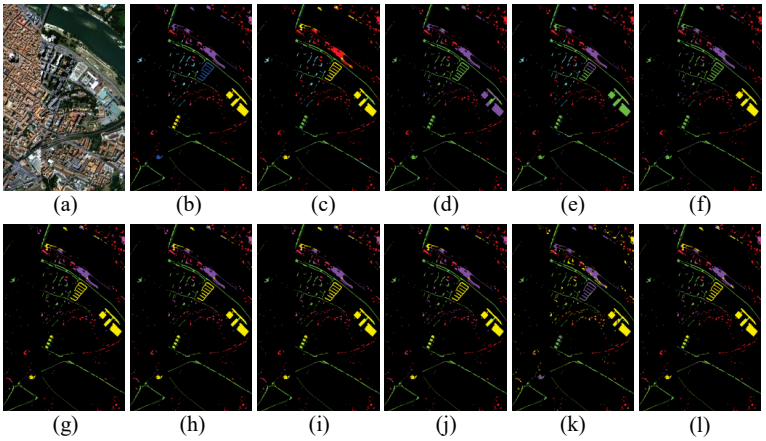


Figure 3: Clustering maps of different methods for the PU→PC task. (a) Pseudo-RGB image; (b) ground truth; (c) SGCNR; (d) S3AGC; (e) BGPC; (f) SAPC; (g) NCSC; (h) SDST; (i) AHSGC; (j) SLCGC; (k) ADSAC; and (l) the proposed DTRA.

4.3 Ablation analysis

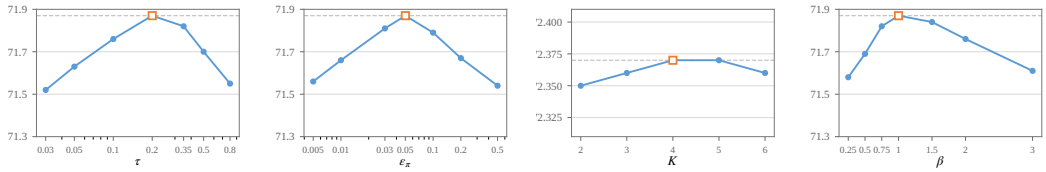
We conduct ablation experiments on PU→PC to assess the contribution of each component of DTRA. As shown in Table 2, the full model obtains 72.37% ACC. We first evaluate the admission mechanism. Removing admission ordering or conservative state screening decreases ACC to 61.34% and 61.57%, respectively, indicating that both the admission sequence and candidate-state screening rule are important for recurrent target refinement. We then examine recurrent feedback and factor separation, whose removal gives 62.35% and 62.36% ACC, respectively. Without variation exchange, ACC decreases to 64.36%, while removing prototype OT yields 67.97%. The results show that transfer-content estimation, prototype correspondence, and progressive target admission work jointly to support cross-scene clustering.

Table 2: Component ablation on PU→PC in ACC (%). Values are three-run means under the same fixed unlabeled initialization state.

Variant	ACC
Full DTRA	**72.37**
− factor separation	62.36
− variation exchange	64.36
− prototype OT	67.97
− admission ordering	61.34
− recurrent feedback	62.35
− conservative state screening	61.57

4.4 Parameter analysis

In this subsection, we examine how four core parameters affect the clustering performance of DTRA on PU→PC. We first analyze the assignment temperature τ, as shown in Fig. 4(a). The ACC increases as τ approaches 0.20 and then decreases gradually, indicating that an intermediate temperature provides a suitable assignment sharpness. In Fig. 4(b), we examine the prototype-transport entropy coefficient ϵ_π. The performance improves toward $\epsilon_\pi = 0.05$ and declines as the coefficient moves farther from this value, suggesting that moderate entropy regularization benefits prototype correspondence. In Fig. 4(c), we analyze the number of admission stages K. The ACC improves from $K = 2$ to $K = 4$ and then becomes nearly saturated, so we adopt $K = 4$. Finally, Fig. 4(d) evaluates the exchange-discrepancy weight β in the admission score. The ACC improves as β increases to 1.0 and then decreases gradually. A small β weakens the influence of the cross-scene



(a) τ (b) ϵ_π (c) K (d) β

Figure 4: Sensitivity analysis of four core parameters on PU→PC. The panels report ACC under varying τ, ϵ_π, K, and β. Open squares mark the selected value in each parameter sweep.

exchange response, whereas an excessively large value suppresses target samples too strongly. These results support the reference settings $\tau = 0.20$, $\epsilon_\pi = 0.05$, $K = 4$, and $\beta = 1.0$.

4.5 Empirical visualization

In this subsection, we conduct empirical visualization to examine how prototype OT affects cross-scene prototype correspondence on PU→PC. We first visualize the raw prototype affinity in Fig. 5(a), where several source prototypes exhibit competing similarities with the target modes. In Fig. 5(b), we analyze the entropy-regularized transport plan. Compared with the raw affinity, the transport mass is concentrated on a smaller set of correspondences, indicating that prototype OT suppresses ambiguous matches while preserving the dominant relations. In Fig. 5(c), we further compare the matched-pair cosine discrepancy before and after barycentric transport. The mean discrepancy decreases from 0.583 to 0.520, and source prototype 2 shows the largest change from 0.821 to 0.422, whereas most other pairs vary only modestly. These results show that prototype OT mainly corrects uncertain cross-scene correspondences without uniformly altering already compatible prototype relations. No ground-truth labels are used to construct the transport plan or the displayed pair matching.

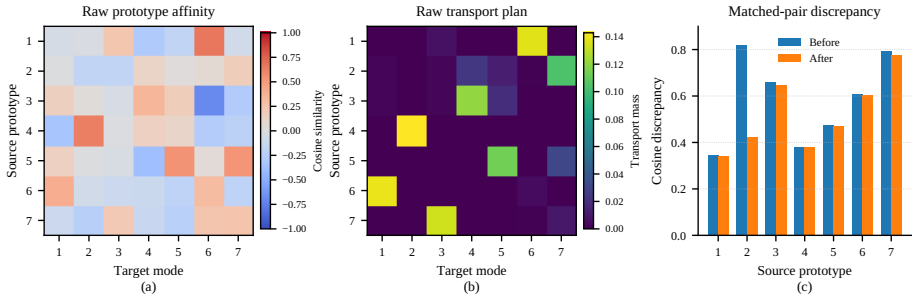


Figure 5: Mechanism-level visualization of prototype OT on PU→PC. (a) Raw cosine affinity in the original target-mode order. (b) Raw entropy-regularized transport plan with $\epsilon_\pi = 0.05$ and the same ordering. (c) Matched-pair cosine discrepancy before and after barycentric transport.

5 Conclusion

This article presents DTRA for cross-scene HSI clustering, which aims to transfer semantic cluster structure between unlabeled scenes while limiting scene-specific spectral variation. More specifically, DTRA separates semantic and scene-specific factors and evaluates transferable semantics through cross-scene variation exchange. Based on the preserved semantic structure, balanced prototypes are constructed and aligned by entropy-regularized transport to establish source–target correspondence. To control when target information participates, assignment confidence and exchange discrepancy are combined to order target samples for progressive admission. The resulting target-oriented prototype state is recurrently fed back to the next training round, allowing transfer content and transfer timing to be coupled through prototype refinement. We further provide an assignment-preservation analysis to justify the exchange discrepancy shared by transfer-content estimation and target admission. Extensive experimental results on six directional tasks from Houston, Pavia, and HyRANK demonstrate that DTRA achieves the highest mean ACC among the compared methods and validate the effectiveness of its main components.

REFERENCES

Yuki M. Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. In *Proceedings of the International Conference on Learning Representations*, 2020.

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Proceedings of the Conference on Neural Information Processing Systems*, volume 29, 2016.

Yaoming Cai, Zhili Zhang, Pedram Ghamisi, Yao Ding, Xiaohua Liu, Zhihua Cai, and Richard Gloaguen. Superpixel contracted neighborhood contrastive subspace clustering network for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Proceedings of the Conference on Neural Information Processing Systems*, volume 33, pp. 9912–9924, 2020.

Omprakash Charkraborty, Aadarsh Sahoo, Rameswar Panda, and Abir Das. Anyda: Anytime domain adaptation. In *Proceedings of the International Conference on Learning Representations*, 2023.

Chaoqi Chen, Weiping Xie, Wenbing Huang, Yu Rong, Xinghao Ding, Yue Huang, Tingyang Xu, and Junzhou Huang. Progressive feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

Xin Chen, Yongshan Zhang, Xue Feng, Xinwei Jiang, and Zhihua Cai. Spectral-spatial superpixel anchor graph-based clustering for hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023.

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

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, 2025a.

Yao Ding, Weijie Kang, Aitao Yang, Zhili Zhang, Junyang Zhao, Jie Feng, Danfeng Hong, and Qinhe Zheng. Adaptive homophily clustering: Structure homophily graph learning with adaptive filter for hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–13, 2025b.

Yao Ding, Zhili Zhang, Aitao Yang, Yaoming Cai, Xiongwu Xiao, Danfeng Hong, and Junsong Yuan. Slcgc: A lightweight self-supervised low-pass contrastive graph clustering network for hyperspectral images. *IEEE Transactions on Multimedia*, 27:8251–8262, 2025c.

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.

Renxiang Guan, Zihao Li, Xianju Li, and Chang Tang. Pixel-superpixel contrastive learning and pseudo-label correction for hyperspectral image clustering. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024.

Zijun He, Ping Wang, Xiaodong Wang, Chang Chen, and Xin Yuan. Joint spectral image reconstruction and semantic segmentation with cooperative unfolding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6910–6919, 2026.

Guozhu Jiang, Yongshan Zhang, Xinxin Wang, Xinwei Jiang, and Lefei Zhang. Structured anchor learning for large-scale hyperspectral image projected clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(3):2328–2340, 2025.

Jiaojiao Li, Zhiyuan Zhang, Rui Song, Yunsong Li, and Qian Du. Scformer: Spectral coordinate transformer for cross-domain few-shot hyperspectral image classification. *IEEE Transactions on Image Processing*, 33:840–855, 2024.

Zhaokui Li, Mingtai Qi, Yan Wang, Xuewei Gong, Cuiwei Liu, and Jinjun Wang. An open-set domain adaptation framework for hyperspectral image classification with pixel-aware weighting and decoupled alignment. *IEEE Geoscience and Remote Sensing Letters*, 22:1–5, 2025.

Jianshang Liao and Liguo Wang. Hierarchicalmamba: A multiscale state space model with dual transitions for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–25, 2026.

Fulin Luo, Yi Liu, Yule Duan, Tan Guo, Lefei Zhang, and Bo Du. Sdst: Self-supervised double-structure transformer for hyperspectral images clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 16888–16905, 2022.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *Proceedings of the International Conference on Learning Representations*, 2023.

Jianhan Qi, Yuheng Jia, Hui Liu, and Junhui Hou. Superpixel graph contrastive clustering with semantic-invariant augmentations for hyperspectral images. arXiv preprint arXiv:2403.01799, 2024.

Ronghua Shang, Chi Wang, Jingyi Ding, and Jie Feng. Multimodal cross-scene hyperspectral image classification guided by dual-agent priori knowledge. *IEEE Transactions on Geoscience and Remote Sensing*, 64, 2026.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *Proceedings of the International Conference on Learning Representations*, 2021.

Haoyu Wang, Tianyang Duan, and Xiaomin Liu. Sibdnet: Spectrum information bottleneck decoupling network for domain generalization in cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–16, 2026.

R. Wang, F. Nie, Z. Wang, F. He, and X. Li. Scalable graph-based clustering with nonnegative relaxation for large hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):7352–7364, 2019.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pp. 478–487, 2016.

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 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.

Minchao Ye, Jiawei Ling, Wanli Huo, Zhaojuan Zhang, Fengchao Xiong, and Yuntao Qian. Discriminative vision transformer for heterogeneous cross-domain hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.

Hanlin Zhang, Yi-Fan Zhang, Weiyang Liu, Adrian Weller, Bernhard Schölkopf, and Eric P. Xing. Towards principled disentanglement for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8024–8034, 2022.

Yongshan Zhang, Guozhu Jiang, Zhihua Cai, and Yicong Zhou. Bipartite graph-based projected clustering with local region guidance for hyperspectral imagery. *IEEE Transactions on Multimedia*, 26:9551–9563, 2024.

Yongshan Zhang, Zixuan Zhang, Xinxin Wang, Lefei Zhang, and Zhihua Cai. Anchor-guided discriminative subspace alignment and clustering for cross-scene hyperspectral imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12961–12969, 2026.

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.