

Spectral Property-Driven Data Augmentation for Hyperspectral Single-Source Domain Generalization

Taiqin Chen^{1,2}, Yifeng Wang³, Xiaochen Feng¹, Zhilin Zhu,^{1,2} Hao Sha¹, Yingjian Li,² Yongbing Zhang^{1*}

¹School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China

²Pengcheng Laboratory, Shenzhen, China

³Department of Computer Science and Technology, Tsinghua University, Beijing, China
chentq@stu.hit.edu.cn, ybzhang08@hit.edu.cn

Abstract

While hyperspectral images (HSI) benefit from numerous spectral channels that provide rich information for classification, the increased dimensionality and sensor variability make them more sensitive to distributional discrepancies across domains, which in turn can affect classification performance. To tackle this issue, hyperspectral single-source domain generalization (SDG) typically employs data augmentation to simulate potential domain shifts and enhance model robustness under the condition of single-source domain training data availability. However, blind augmentation may produce samples misaligned with real-world scenarios, while excessive emphasis on realism can suppress diversity, highlighting a tradeoff between realism and diversity that limits generalization to target domains. To address this challenge, we propose a spectral property-driven data augmentation (SPDDA) that explicitly accounts for the inherent properties of HSI, namely the device-dependent variation in the number of spectral channels and the mixing of adjacent channels. Specifically, SPDDA employs a spectral diversity module that resamples data from the source domain along the spectral dimension to generate samples with varying spectral channels, and constructs a channel-wise adaptive spectral mixer by modeling inter-channel similarity, thereby avoiding fixed augmentation patterns. To further enhance the realism of the augmented samples, we propose a spatial-spectral co-optimization mechanism, which jointly optimizes a spatial fidelity constraint and a spectral continuity self-constraint. Moreover, the weight of the spectral self-constraint is adaptively adjusted based on the spatial counterpart, thus preventing over-smoothing in the spectral dimension and preserving spatial structure. Extensive experiments conducted on three remote sensing benchmarks demonstrate that SPDDA outperforms state-of-the-art methods.

Code — <https://github.com/hnsytq/SPDDA>

Introduction

Hyperspectral image (HSI) is capable of accurately depicting the intrinsic properties of matter, which is facilitated by its simultaneous representation of both spectral and spatial information of the imaging scene. Compared with RGB

images, HSI shows better recognition performance with its richer spectral depicting ability, and thus has a broad application prospect in the fields of remote sensing, medical diagnosis, etc (Kumar et al. 2024).

As a fundamental task in HSI-based analysis, hyperspectral image classification (HSIC) for remote sensing (Li et al. 2019) achieves pixel-level land-cover identification in regions of interest. With the rapid development of deep learning (He et al. 2024; Yuan et al. 2018), the performance of HSIC has been significantly improved. Deep learning-based methods typically leverage a large annotated dataset to learn mappings from HSI to categorical labels (Wang et al. 2022), under the assumption that test samples are independent and identically distributed with the training data, where test samples are referred to as the target domain (TD) and the training data is referred to as the source domain (SD). However, in practice, substantial discrepancies often exist between the training and testing data distributions due to multiple factors such as illumination and weather (Yang et al. 2021). This domain disparity results in a dramatic degradation of classification performance (Zhou et al. 2022).

To tackle this issue, domain generalization (DG) is proposed to learn domain-invariant features with robustness and transferability in the training phase (Blanchard, Lee, and Scott 2011), thereby enhancing classification performance on the unseen TD. The classical setting of DG assumes that multi-source domain data with annotations can be accessed during classifier training (Li et al. 2017). However, obtaining labeled data from multiple domains is often impractical due to data privacy and expensive annotation costs in real-world remote sensing scenarios. Therefore, a more challenging setting that is single source domain generalization (SDG) (Zhang et al. 2023b) is more suitable for HSIC, where only samples from a single SD is available.

Since single-source samples are insufficient to support the learning of domain-invariant features, existing methods typically employ data augmentation to simulate distributional shift between different domains in SDG. In brief, these methods typically apply either standard augmentations (Zheng, Huai, and Zhang 2024) or trainable generative models (Wang et al. 2021) to introduce random perturbations into SD, thereby generating diverse synthetic samples, referred to as the extended domain (ED). However, HSI depicts real-world scenes, which is contrasty to vari-

*Corresponding author: Yongbing Zhang
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ous stylistic variants of RGB images such as cartoons and skeletons. The mentioned blind data augmentation may generate unrealistic samples, thus biasing the classifier away from its intended prediction mechanism. To tackle this challenge, pioneering studies (Qin et al. 2024; Chen et al. 2024) have introduced semantic constraints to ensure the reliability of ED during augmentation process. Unfortunately, such constraints inevitably result in similarity between ED and SD, thereby limiting the augmentation diversity. Overall, existing hyperspectral SDG methods suffer from a diversity-realism tradeoff, which constrains the performance of classifiers in unknown TD.

To tackle this challenge, we explicitly account for inherent properties of HSI during augmentation procedure. Specifically, HSI exhibits substantial variability across different scenes and imaging devices (Hyttinen et al. 2020). The spectral sensing range and the number of spectral channels of HSI are varied, which depend on the employed imaging devices or specific application requirements, referred to as spectral heterogeneity (Zhang et al. 2023a). Moreover, the acquired HSI may be aliased due to overlapping sensor spectral responses or atmospheric scattering interference (Jiang et al. 2023; Hong et al. 2018), causing adjacent channels in HSI to be mixed or single pixel mixing the spectral responses of multiple matters. The above properties lead to potentially huge distribution shift even for HSIs describing the same scene. In other words, introducing random perturbations into SD grounded in these inherent properties can avoid imposing strict constraints between ED and SD, thereby preserving diversity and realism simultaneously.

Based on the above analysis, we propose a spectral property-driven data augmentation (SPDDA) to mitigate the tradeoff between diversity and realism in hyperspectral SDG. Specifically, we introduce a spectral diversity module (SDM) for constructing a channel mask and a channel-wise adaptive spectral mixer (CASM). The channel mask is utilized to mask channels with less semantic information, thereby generating HSIs with arbitrary number of channels, corresponding to spectral heterogeneity. CASM then adaptively mixes channels of the generated HSI, whose mixing kernel size and weights are tailored for each channel based on inter-channel similarity. Furthermore, we develop a spatial-spectral co-optimization mechanism (SSCOM) to ensure reliability of the generation, which is comprised of a spatial fidelity constraint and a spectral continuity self-constraint. To alleviate overly smoothening in the spectral dimension, the weight of the spectral self-constraint is dynamically adjusted based on the spatial constraint. Briefly, our contributions are as follow:

- We propose SPDDA, a novel augmentation framework that balances the tradeoff between diversity and realism by leveraging the inherent properties of HSI.
- We design SDM, which masks a variable number of channels based on semantic information and constructs the CASM to avoid fixed perturbation patterns, thus enhancing the diversity of augmented samples. The procedure is supervised by the SSCOM, a dynamic joint optimization mechanism that constrains across both spatial

and spectral dimensions.

- We conduct extensive experiments on three widely-used hyperspectral benchmarks, demonstrating that the proposed method achieves state-of-the-art (SOTA) performance compared to existing SDG methods.

Related Work

Domain Generalization

Domain generalization aims to mitigate the performance degradation caused by the domain shifts, which can be categorized into multi-source domain generalization and single-source domain generalization (SDG). The former intends to learn domain-invariant representations to enhance classifier performance in the unseen target domain (TD) by utilizing labeled samples from multiple source domain during training (Zhao et al. 2025). In contrast, SDG merely accesses labeled samples from a single-source domain. Hence, existing methods typically inject perturbation into samples to simulate the domain shifts and obtain diverse extended data. For example, PDEN (Li et al. 2021) proposes a progressive domain expansion network to simulate photometric and geometric transforms in TD. AdvST (Zheng, Huai, and Zhang 2024) introduces an adversarial framework that learns parameters of composed standard data augmentations as semantics transformations to generate challenging and diverse samples. UniFreqSDG (Liu et al. 2024a) learns adaptive low-frequency perturbations across multiple feature levels and combines a domain-divergence inducement loss to enhance analysis performance in TD. PEER (Cho, Hwang, and Lee 2025) utilizes a proxy model in place of the classifier to learn from augmented data and maximizes their mutual information to mitigate effects of distortion. However, directly applying these methods to hyperspectral SDG often fails to effectively exploit spectral characteristics.

Hyperspectral Single Source Domain Generation

SDENet (Zhang et al. 2023b) is the first to apply SDG to hyperspectral image classification to tackle the performance decline caused by domain shifts in cross-scene tasks. SDENet proposes a hyperspectral-tailored generator that comprises a spatial-spectral model and a morph model to obtain augmented samples. SFT (Huang et al. 2025) proposes a Fourier transformation that learns dynamic attention maps in the frequency domain of source-domain samples to enhance diversity of the augmented samples. However, these methods overemphasize diversity, which may result in generated data deviating from the practical distribution and introducing bias into classifier optimization. To tack this issue, FDGNet (Qin et al. 2024) incorporates data geometry into the generation process to preserve authenticity, enforcing augmented samples to retain low-dimensional structures similar to those of the source domain. S2AMSNet (Chen et al. 2024) introduces a mutual information regulation to prevent random perturbation from corrupting the semantic information of samples. However, such a constraint inevitably compromises diversity.

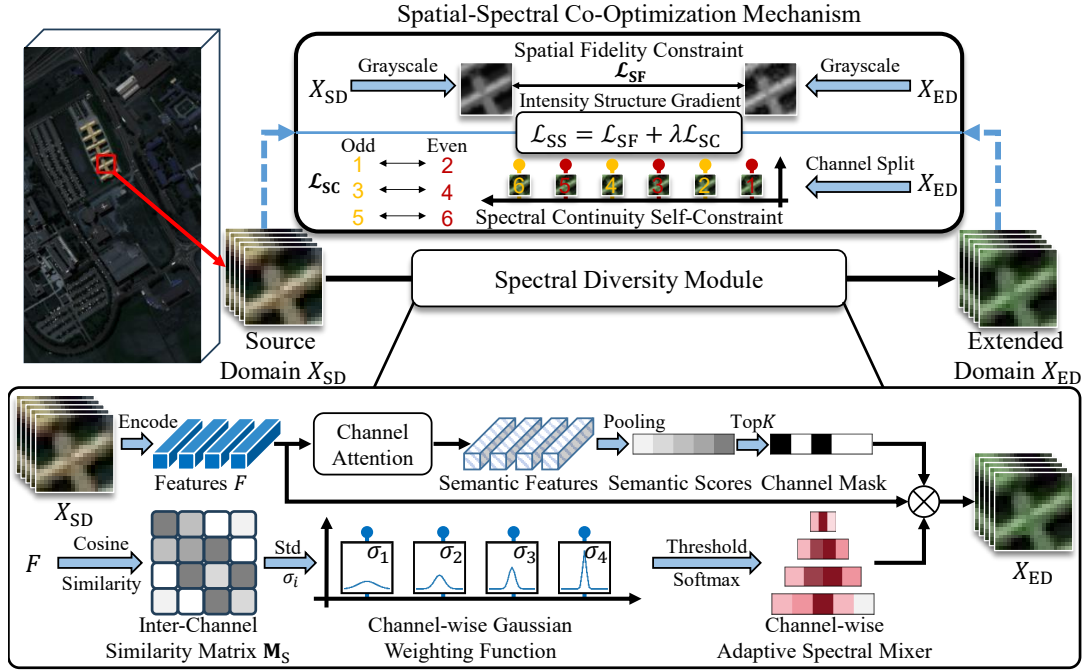


Figure 1: Overview of spectral property-driven data augmentation (SPDDA). Briefly, a spectral diversity module (SDM) is utilized to inject perturbations into source-domain samples $X_{SD} \in \mathbb{R}^{H \times W \times C}$ to obtain extended-domain samples $X_{ED} \in \mathbb{R}^{H \times W \times K}$, where K is a random integer. SDM constructs a channel mask and a channel-wise adaptive spectral mixer to masks an arbitrary number of channels and mixes the remaining channels, which adheres to the practical phenomenon in hyperspectral image. The augmentation procedure is supervised by a spatial-spectral co-optimization mechanism, which is comprised of a spatial fidelity constraint and a spectral continuity self-constraint.

Methodology

Framework Overview

Although existing hyperspectral image (HSI) classification methods have achieved excellent performance in remote sensing, they typically assume that the processed samples are independent and identically distributed. Thus, the cross-domain shifts that exist in the practical scenarios would result in a drastic performance degradation. Under the condition of single-source domain samples availability, pioneering works address this issue by generating augmented samples to simulate domain shifts between different domains. However, there is a tradeoff between realism and diversity during the generation process, which constraints their performance. To tackle this challenge, we propose a spectral property-driven data augmentation (SPDDA), which obtains the augmented samples by accounting for inherent properties of HSI. The framework of SPDDA is illustrated in Fig. 1.

SPDDA develops a spectral diversity module to inject perturbations to the source-domain samples $X_{SD} \in \mathbb{R}^{H \times W \times C}$. Firstly, X_{SD} is encoded by multiple ResBlock (He et al. 2016) and obtain features $F \in \mathbb{R}^{H \times W \times C}$, where H and W denote the spatial size, C denotes the number of spectral channels. Then a channel mask and a channel-wise adaptive spectral mixer are constructed based on the semantic information and inter-channel similarity of F , respectively. Finally, we integrate channel mask, adaptive spectral mixer

and F by channel-wise multiplication to acquire the augmented samples $X_{ED} \in \mathbb{R}^{H \times W \times K}$, where K is a random integer. To supervise the sample generation process, a spatial-spectral co-optimization mechanism is proposed to enforce X_{ED} to adhere to spatial fidelity constraint \mathcal{L}_{SF} and spectral continuity self-constraint \mathcal{L}_{SC} simultaneously. In addition, the weight of spectral constraint is adjusted adaptively based on $\|\mathcal{L}_{SC}\|_1$, thereby achieving a dynamic balance and cooperative optimization between two constraints.

Spectral Diversity Module

A spectral diversity module (SDM) is proposed to simulate device-dependent variations in the number of spectral channels and the mixing of adjacent channels, thereby injecting perturbations for X_{SD} and generating diverse extended samples X_{ED} . The structure of SDM is depicted in Fig. 1. In brief, SDM constructs a channel mask and a channel-wise adaptive spectral mixer (CASM), where the former is utilized to mask channels with low semantic scores to obtain HSI with an arbitrary number of channel. CASM accounts for the linear spectral mixing model and constructs a channel-wise Gaussian weighting function based on inter-channel similarity to obtain mixing kernels with varying weights and sizes. Details are as follows.

SDM firstly encodes X_{SD} to obtain features F through multiple ResBlock. Then a channel-wise self attention mechanism (Dosovitskiy et al. 2020) is utilized to extract

semantic information from F , and obtain semantic features F_s . In brief, a layer normalization is first employed to normalize F , a convolution group is further utilized to generate query token \mathbf{Q} . The key token \mathbf{K} and the value token \mathbf{V} are obtained by the same manner. Then, attention matrix $\mathbf{A} \in \mathbb{R}^{C \times C}$ is acquired through a dot product between \mathbf{Q} and \mathbf{K} . Finally, we integrate \mathbf{A} and \mathbf{V} by matrix production to focus on more valuable channels, and a feed-forward network (Bian et al. 2024) is further employed to enhance representative ability to obtain F_s .

Furthermore, F_s is compressed by the pooling layers to serve as a semantic score for each channel. The channels with low scores are masked to obtain the channel mask. The calculation is formulated as:

$$S = \text{Avg}(F_s) + \text{Max}(F_s), \quad (1)$$

where $S \in \mathbb{R}^{1 \times C}$ denotes the semantic score, $\text{Avg}(\cdot)$ denotes the average pooling, and $\text{Max}(\cdot)$ denotes the maximum pooling. Then, the channel mask is acquired by activating the first K channels and masking the remaining depending on S , where K is a random integer.

Moreover, we design the CSAM to enable diverse augmentation patterns. Specifically, we employ cosine similarity (Guan et al. 2025) to characterize correlation between channels of F , and obtain the counterpart matrix $\mathbf{M}_S \in \mathbb{R}^{C \times C}$. Then, we compute the row-wise standard deviation σ_i of \mathbf{M}_S to characterize the distribution of similarities between each channel and the others, where $i \in [1, \dots, C]$ denotes the i -th channel. σ_i is further used to act as the scale parameter to construct channel-wise Gaussian weighting function, achieving channel-tailored adaptive adjustment while keeping continuous and differentiable. The above procedure is formulated as follows.

$$w_j^i = \frac{e^{-\frac{l(j)^2}{2\sigma_i^2}}}{\sqrt{2\pi}\sigma_i}, \quad (2)$$

where w_j^i denotes the j -th weight in the i -th mixing kernel, $j \in [-p, p]$, p is a human-set parameter to indicate the maximum mixer size, and $l(j) = m + \epsilon(j + p)$. A thresholding operation is then applied to w_j^i , which preserves the elements that satisfy $w_j^i \leq \sigma_i$, and then w^i is normalized via softmax function to complete the construction of CSAM.

After that, feature F is masked by the channel mask to obtain HSI X_{mask} with arbitrary number of channels. Then CSAM introduces perturbations in the spectral dimension via channel-wise multiplication, which is formulated as:

$$X_{\text{ED}}^i = \sum_{j=-p}^{j=p} X_{\text{mask}}^{i+j} \times w_j^i, \quad (3)$$

where X_{ED}^i denotes the i -th channel of X_{ED} , and X_{mask} is padded to ensure the correctness of computation.

Spatial-Spectral Co-Optimization Mechanism

Furthermore, we design a spatial-spectral co-optimization mechanism (SSCOM) to ensure that the spatial information described by X_{ED} is consistent with X_{SD} , while preventing excessive aliasing from disrupting the continuity of

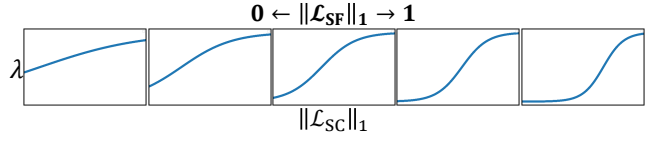


Figure 2: Illustration of the variation in the spectral continuity self-constraint weight λ .

the spectral dimension. SSCOM comprises a spatial fidelity constraint \mathcal{L}_{SF} and a spectral continuity self-constraint \mathcal{L}_{SC} , which is formulated as:

$$\mathcal{L}_{\text{SS}} = \mathcal{L}_{\text{SF}} + \lambda \mathcal{L}_{\text{SC}}, \quad (4)$$

where \mathcal{L}_{SS} denotes the optimized objective of SSCOM, λ denotes an adaptive coefficient, aiming to dynamically adjust the weight of \mathcal{L}_{SC} in accordance with the optimization progress of \mathcal{L}_{SF} .

For the spatial fidelity constraint, we grayscale X_{ED} and X_{SD} , and denote counterparts as I_{ED} and I_{SD} , respectively. Then gray images from different domains are forced to share similar spatial information such as structure and texture, which ensures the spectral diversity of the augmented samples is not constrained by the source domain. Accordingly, we formulate \mathcal{L}_{SF} as:

$$\mathcal{L}_{\text{SF}} = \mathcal{L}_{\text{MSE}}(I_{\text{SD}}, I_{\text{ED}}) + \mathcal{L}_{\text{SSIM}}(I_{\text{SD}}, I_{\text{ED}}) + \mathcal{L}_{\text{GRA}}(I_{\text{SD}}, I_{\text{ED}}), \quad (5)$$

where $\mathcal{L}_{\text{MSE}}(\cdot, \cdot)$ denotes the mean square error (MSE) loss, $\mathcal{L}_{\text{SSIM}}(\cdot, \cdot)$ denotes the structural similarity index measure (SSIM) loss, and $\mathcal{L}_{\text{GRA}}(\cdot, \cdot)$ denotes the gradient loss.

In addition, we introduce the spectral continuity self-constraint by taking into account the inherent property of HSI, i.e., neighboring channels characterize similar content. Briefly, we split X_{ED} by odd-even channel, which are denoted as X_{ED}^{E} and X_{ED}^{O} . Then they are constrained by MSE and SSIM, which is formulated as:

$$\mathcal{L}_{\text{SC}} = \mathcal{L}_{\text{MSE}}(X_{\text{ED}}^{\text{E}}, X_{\text{ED}}^{\text{O}}) + \mathcal{L}_{\text{SSIM}}(X_{\text{ED}}^{\text{E}}, X_{\text{ED}}^{\text{O}}). \quad (6)$$

Although \mathcal{L}_{SC} facilitates maintaining spectral continuity, it lacks effective supervisory information. When the network places excessive focus on this term during training, it may lead to augmented samples that are overly smoothed in the spectral dimension, neglecting spatial structure preservation and thereby reducing both diversity and realism. To tackle this issue, we implement a nonlinear adjustment of the weight for \mathcal{L}_{SC} based on the hyperbolic tangent function and employ $\|\mathcal{L}_{\text{SF}}\|_1$ to dynamically tune the translation and curvature change of the function. Hence, the decreasing rate of λ accelerates and enables reaching a lower value when $\|\mathcal{L}_{\text{SF}}\|_1$ takes a higher value, whose variation is illustrated in Fig. 2. The mentioned procedure is formulated as:

$$\begin{cases} \lambda = 1 + \frac{e^t - e^{-t}}{e^t + e^{-t}}, \\ t = \frac{\|\mathcal{L}_{\text{SC}}\|_1 - f_{\text{T}}(\|\mathcal{L}_{\text{SF}}\|_1)}{f_{\text{C}}(\|\mathcal{L}_{\text{SF}}\|_1)}, \end{cases} \quad (7)$$

where $f_{\text{T}}(\cdot) = \frac{s\|\mathcal{L}_{\text{SF}}\|_1}{2}$ denotes the translating function, $f_{\text{C}}(\cdot) = \frac{1}{s\|\mathcal{L}_{\text{SF}}\|_1}$ denotes the curvature function, s denotes a scaling parameter.

Methods	Houston			Pavia			HyRank		
	OA \uparrow	F1 \uparrow	Kappa \uparrow	OA \uparrow	F1 \uparrow	Kappa \uparrow	OA \uparrow	F1 \uparrow	Kappa \uparrow
SDENet	67.02	0.6768	0.4702	<u>79.05</u>	<u>0.7860</u>	<u>0.7471</u>	58.40	<u>0.5574</u>	0.4916
FDGNet	75.44	0.7329	0.5360	78.28	0.7736	0.7371	56.81	0.5035	0.4554
S2AMSNet	71.13	0.7291	0.5541	73.42	0.7303	0.6792	59.46	0.5561	0.4940
S2ECNet	72.84	0.7355	0.5641	78.05	0.7828	0.7358	<u>60.21</u>	0.5617	0.5047
ProRandConv	<u>75.71</u>	<u>0.7635</u>	<u>0.5911</u>	78.27	0.7763	0.7375	57.76	0.5295	0.4753
ABA	52.68	0.5634	0.3321	57.87	0.5698	0.5070	48.31	0.3625	0.3235
StyDeSty	70.99	0.7155	0.5260	70.33	0.7075	0.6432	55.25	0.4559	0.4255
Ours	78.60	0.7826	0.6252	83.03	0.8348	0.7963	60.70	0.5316	<u>0.5024</u>

Table 1: Quantitative Results on Cross-scene Hyperspectral Image Classification for Remote Sensing. The optimal results are bold, and the sub-optimal results are underlined.

Loss Functions

Except for SSCOM \mathcal{L}_{SS} , a cross-entropy loss \mathcal{L}_{CE} is also employed to supervise the training procedure of SPDDA, whose optimized objective is formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{SS} + \mathcal{L}_{CE}(P, y), \quad (8)$$

where P denotes the prediction outputted by the classifier which takes X_{ED} as inputs, y denotes the category label.

Experiments and Results

Experimental Settings

Datasets We employ three cross-scene hyperspectral image classification datasets (Zhang et al. 2021, 2022) to evaluate the proposed method, including Houston, Pavia, and HyRank. Both datasets contain two scenes, which were acquired by capturing different areas in the same time or capturing the same areas in different years using distinct imaging devices. For Houston, Houston13 and Houston18 are used as the source domain (SD) and target domain (TD), respectively. In the case of Pavia, PaviaU serves as the SD and PaviaC as the TD. For HyRank, Dioni and Loukia are designated as the SD and TD, respectively. Their number of categories is 7, 7, and 12, respectively. In addition, each dataset is partitioned into 13×13 patches, where the label of the central pixel is assigned as the category of the corresponding patch.

Implementation Details In all experiments, we employ LeNet (LeCun et al. 2002) as the classifier, where the first convolution is replaced by a channel-adaptive convolution (Chen et al. 2023) to suit inputs with varying numbers of channels. During training, the chosen optimizer is Adam with the learning rate 1×10^{-4} . The hyperparameters in Eq. 2 is set as: $m = -5$, $\epsilon = 0.5$, and $p = 10$. The hyperparameter in Eq. 7 is set as: $s = 15$. Following existing methods (Zhang et al. 2023b), we linearly integrate SD and ED to generate intermediate-domain samples and train the classifier using both cross-entropy loss and contrastive loss (Khosla et al. 2020). The classifier of each comparative method is kept the same as the proposed method. All experiments are conducted on the Pytorch platform, and the computing device is equipped with an NVIDIA RTX 4090D GPU. Each method is trained for 400 epochs.

Evaluation Metrics Several evaluation metrics are employed to quantify results of each method, including the overall accuracy (OA), the weighted F1 score (F1), and the kappa coefficient (Kappa).

Comparative Results

To demonstrate the superiority of our method, we employ multiple state-of-the-art (SOTA) methods in hyperspectral single-source domain generalization (SDG) as comparative methods, comprising of SDENet (Zhang et al. 2023b), FDGNet (Qin et al. 2024), S2ECNet (Dong, Geng, and Jiang 2024), and S2AMSNet (Chen et al. 2024). In addition, several SOTA methods for RGB images also taken as comparison, including ProRandConv (Choi et al. 2023), ABA (Cheng, Gokhale, and Yang 2023), and StyDeSty (Liu et al. 2024b). To ensure a fair comparison, all the above methods are implemented using their official open-source codes, with configurations strictly following those reported in their respective papers.

Quantitative Analysis Experimental datasets are configured as Houston, Pavia, and HyRank. Quantitative results are presented in Tab. 1. In terms of OA, the proposed method achieves optimal results compared to the comparisons, indicating the significant advantages of our method in hyperspectral SDG. FDGNet introduces geometry constraints to ensure that the augmented samples keep realism, resulting in a unbiased distribution compared with other comparative methods. However, such constraints often suppress the diversity of augmented samples, thereby limiting the generalization ability of the classifier. A similar issue is even more pronounced in S2AMSNet. In addition, the RGB-based method ProRandConv employs random convolution layers to perturb the local semantic information of source-domain samples, thereby obtaining diverse augmented samples and introducing extra spatial information like structure for the classifier. However, blind convolution may disrupt spectral characteristics and thus impair classification performance. In contrast to existing methods, the proposed method generates augmented samples by accounting for the inherent spectral properties, i.e., simulating real-world variations in spectral channel counts and aliasing between adjacent channels. Due to this generative manner, the proposed method is

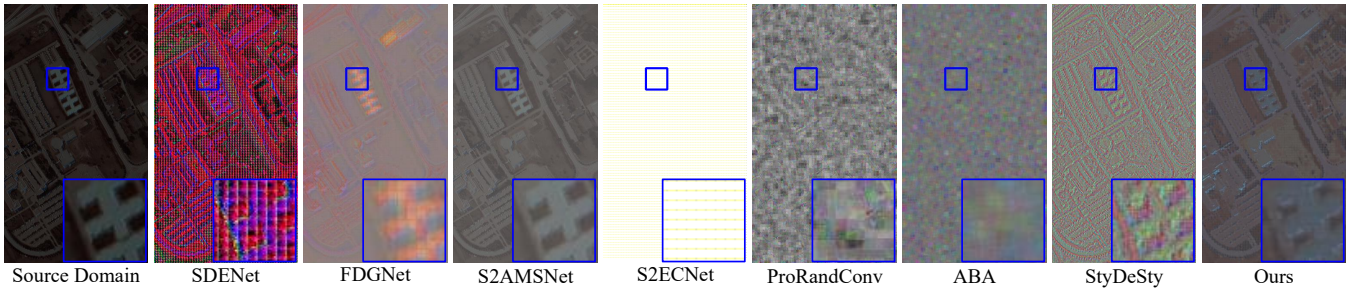


Figure 3: Illustration of the pseudo-color images for the extended-domain scene for each method. Regions marked by the blue rectangle are enlarged and displayed in the lower right corner of each image.

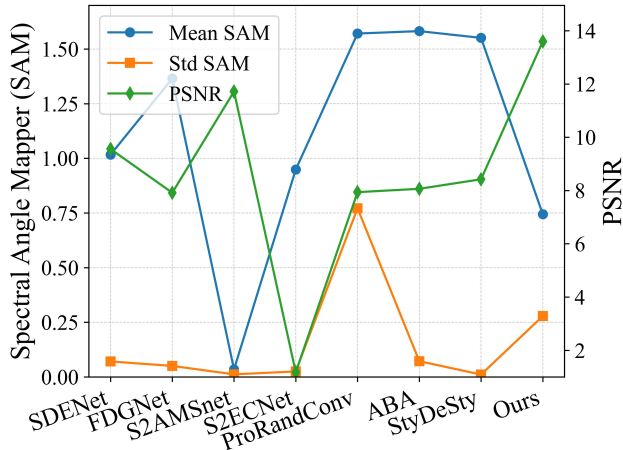


Figure 4: Comparison of realism and diversity. PSNR measures spatial realism, while the mean and standard deviation (Std) of SAM represent spectral realism and diversity, respectively. A lower mean SAM indicates higher spectral realism, and a greater Std means higher diversity.

able to preserve both the realism and diversity of the augmented samples, thereby achieving superior performance.

Qualitative Analysis We further conduct a qualitative evaluation of the proposed method and the comparison methods on the Pavia dataset. Firstly, all generated samples are sequentially integrated to construct the entire extended-domain scene, and pseudo-color images are produced for qualitative visualization, which is illustrated in Fig. 3. To facilitate a more effective comparison, we enlarge regions highlighted by the blue rectangle in each visualization. In addition, we evaluate the peak signal to noise ratio (PSNR) of the pseudo-color images to quantify the spatial realism for each method, taking the source-domain counterpart as the reference. The spectral angle mapper (SAM) (De Carvalho and Meneses 2000) is also computed, where SAM assesses the similarity between source-domain samples and generated data based on the angular difference in spectral dimension. The mean value and the standard deviation of SAM are employed to quantify the realism and diversity, respectively. The comparison is depicted in Fig. 4.

ProRandConv and ABA severely compromise the semantic consistency of the samples due to the utilization of convolution-based augmentation strategy. Although StyDeSty explicitly optimizes the reliability of generated samples during training, it still introduces certain distortions as it considers only spatial content. S2ECNet concentrates exclusively on domain-invariant feature extraction during training, resulting in a tendency to produce meaningless samples, which is shown in Fig. 3. In comparison, samples produced by the hyperspectral image-tailored S2AMSNNet exhibit superior realism. However, their diversity is limited due to the introduction of strict constraints based on the source-domain samples, which is depicted in Fig. 4. In contrast to comparisons, the proposed method merely establishes basic grayscale constraints with the original samples and flexibly injects perturbations in accordance with inherent heterogeneity and mixing phenomena in the spectral dimension. Therefore, our method effectively addresses the tradeoff between realism and diversity.

Ablation Study

Module Ablation Experiments We then conduct ablation experiments on Houston to verify the effect of channel-wise adaptive spectral mixer (CASM), spatial fidelity constraint (\mathcal{L}_{SF}), and spectral continuity self-constraint (\mathcal{L}_{SC}). In addition, we employ a fixed mixer as a substitute for CASM. Results are presented in Tab. 2. Except for the accuracy metrics, we also compute the mean SAM and Std SAM to quantify realism and diversity in the spectral dimension, respectively. Results are illustrated in Fig. 5.

Pairing the model solely with CASM can effectively enhance realism and diversity, thereby facilitating improvements in classifier performance in the unseen target domain. Conversely, \mathcal{L}_{SF} or \mathcal{L}_{SC} induces a drastic decline in classification accuracy. As seen in Fig. 5, this degradation likely stems from the fixed mixing pattern suppressing the diversity of generated samples, which is further aggravated by the introduction of \mathcal{L}_{SF} and \mathcal{L}_{SC} . Compared to \mathcal{L}_{SC} , \mathcal{L}_{SF} enforces augmented samples are similar to source-domain samples in the grayscale domain, causing more severe diversity loss. On the other hand, it improves the realism of sample generation, thus mitigating the adverse impact of unrealistic samples on classification performance. Therefore, the classifier performance can be effectively improved when

CASM	\mathcal{L}_{SF}	\mathcal{L}_{SC}	OA \uparrow	F1 \uparrow	Kappa \uparrow
×	×	×	74.73	0.7637	0.6068
✓	×	×	76.95	<u>0.7794</u>	0.6264
×	✓	×	70.93	0.7252	0.5408
×	×	✓	72.23	0.7414	0.5711
✓	✓	×	<u>77.87</u>	0.7724	0.6081
✓	×	✓	<u>75.94</u>	0.7643	0.6003
×	✓	✓	71.26	0.7299	0.5433
✓	✓	✓	78.60	0.7826	<u>0.6252</u>

Table 2: Ablation study of channel-wise adaptive spectral mixer (CASM), spatial fidelity constraint (\mathcal{L}_{SF}), and spectral continuity self-constraint (\mathcal{L}_{SC}) on Houston.

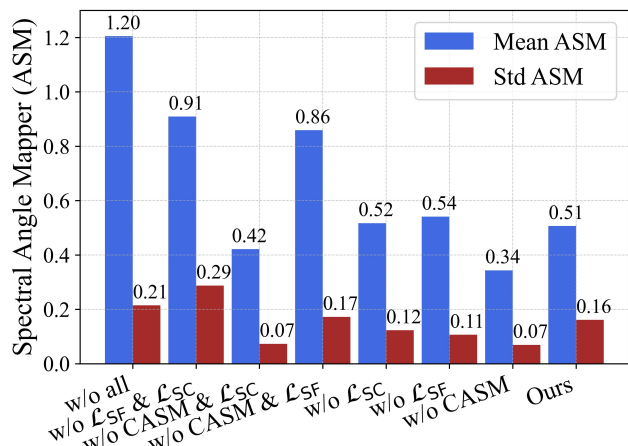


Figure 5: Effects of CASM, \mathcal{L}_{SF} , and \mathcal{L}_{SC} on spectral angle mapper (SAM). The average SAM quantifies realism of generation. A lower average SAM indicates higher realism.

\mathcal{L}_{SF} is combined with CASM. When three components are combined, the tradeoff between diversity and realism is further alleviated, and the performance achieves optimal.

λ -tuning Ablation Experiments We further conduct λ -tuning (λ -T) experiments to evaluate the effects of the adaptive tuning mechanism described in Eq. 7. Specifically, we modify the tuning strategy of λ and obtained seven variants. In the first five versions, λ is set to a fixed value, i.e., $\lambda \in \{0.1, 0.5, 1.0, 1.5, 2.0\}$, which are respectively denoted as Fix-0.1, Fix-0.5, etc. In addition, the last two versions divide the training process into early and late stages, which are denoted as Var_E and Var_S , respectively. In the early stage, λ is configured to 0.1 and to 2.0 during the late stage, where 0.1 is close to the lower bound of Eq. 7, 2.0 is the upper bound of Eq. 7. Var_E defines the early training stage as epochs less than 200, whereas Var_S considers the training to be in the early stage when $\|\mathcal{L}_{SF}\|_1 > 0.08$, which is determined based on the empirical observations.

The experiments are conducted on the Houston dataset, whose results are presented in Tab. 3. The fixed λ significantly degrades the performance of classifier in the target domain. This may be attributed to the use of fixed

λ -T	OA \uparrow	F1 \uparrow	Kappa \uparrow
Fix-0.1	75.19	0.7563	0.5677
Fix-0.5	75.18	0.7651	0.5877
Fix-1.0	72.05	0.7398	0.5374
Fix-1.5	75.27	0.7662	0.5800
Fix-2.0	74.47	0.7620	0.5729
Var_E	<u>76.15</u>	0.7632	<u>0.5906</u>
Var_S	75.62	<u>0.7687</u>	0.5817
Ours	78.60	0.7826	0.6252

Table 3: Ablation study of λ -tuning (λ -T) strategy on Houston. Fix-0.1 denotes that λ is set to a fixed value 0.1. Var_E and Var_S represent early-late stage division based on training epochs and $\|\mathcal{L}_{SF}\|_1$, respectively, with λ set to 0.1 in the early stage and 2.0 in the late stage.

weights, which hinder the generator from balancing the spatial fidelity constraint \mathcal{L}_{SF} and the spectral continuity self-constraint \mathcal{L}_{SC} during training. Specifically, \mathcal{L}_{SF} serves as an anchor during the optimization process of generator since it establish a basic grayscale constraint with source-domain samples. \mathcal{L}_{SC} serves as a regularization term through maintaining inter-channel continuity. Therefore, as training progresses, the optimization focus of generator should gradually shifts between the two constraints. Accordingly, Var_E and Var_S achieve this objective by employing a two-stage adjustment strategy, which prioritizes spatial fidelity during the early training stage and focuses on optimizing spectral continuity in the later stage. However, it remains challenging to establish a criterion for dividing these stages to achieve satisfactory results. In contrast, the proposed adaptive tuning strategy adjusts λ based on $\|\mathcal{L}_{SF}\|_1$, effectively enabling generator to shift its optimization focus between the two constraints adaptively and achieve optimal performance.

Conclusion

In this paper, we propose a spectral property-driven data augmentation (SPDDA) to mitigate the dilemma that the tradeoff between realism and diversity of the augmented samples in hyperspectral single-source domain generalization. Specifically, SPDDA develops a spectral diversity module (SDM) to simulate the variation in spectral channel counts and the spectral mixing among adjacent channels, which are inherent phenomena in real-world scenarios. To achieve diverse mixing patterns, SDM constructs a channel-wise adaptive spectral mixer through modeling a channel-wise Gaussian weighting function based on the inter-channel similarity. Moreover, we propose a spatial-spectral co-optimization mechanism (SSCOM) that comprises a spatial fidelity constraint \mathcal{L}_{SF} and a spectral continuity self-constraint \mathcal{L}_{SC} to ensure that generation maintains realism. The weight of \mathcal{L}_{SC} is further adjusted adaptively according to \mathcal{L}_{SF} , realizing a dynamical balance between two constraints during training. The effectiveness of the proposed method is identified by comprehensive experiments on three open-source datasets.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62031023 and 62331011, and in part by Shenzhen Science and Technology Project under Grant GXWD20220818170353009.

References

- Bian, L.; Wang, Z.; Zhang, Y.; Li, L.; Zhang, Y.; Yang, C.; Fang, W.; Zhao, J.; Zhu, C.; Meng, Q.; et al. 2024. A broad-band hyperspectral image sensor with high spatio-temporal resolution. *Nature*, 635(8037): 73–81.
- Blanchard, G.; Lee, G.; and Scott, C. 2011. Generalizing from several related classification tasks to a new unlabeled sample. *Advances in neural information processing systems*, 24.
- Chen, X.; Gao, L.; Zhang, M.; Chen, C.; and Yan, S. 2024. Spectral–spatial adversarial multidomain synthesis network for cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16.
- Chen, Z. S.; Pham, C.; Wang, S.; Doron, M.; Moshkov, N.; Plummer, B.; and Caicedo, J. C. 2023. CHAMMI: A benchmark for channel-adaptive models in microscopy imaging. *Advances in Neural Information Processing Systems*, 36: 19700–19713.
- Cheng, S.; Gokhale, T.; and Yang, Y. 2023. Adversarial bayesian augmentation for single-source domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11400–11410.
- Cho, D. K.; Hwang, I.; and Lee, S. 2025. PEER pressure: Model-to-Model Regularization for Single Source Domain Generalization. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 15360–15370.
- Choi, S.; Das, D.; Choi, S.; Yang, S.; Park, H.; and Yun, S. 2023. Progressive random convolutions for single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10312–10322.
- De Carvalho, O. A.; and Meneses, P. R. 2000. Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM). In *Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18*, volume 9, 2. JPL publication Pasadena, CA, USA.
- Dong, L.; Geng, J.; and Jiang, W. 2024. Spectral–spatial enhancement and causal constraint for hyperspectral image cross-scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–13.
- Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Guan, X.; Wang, Y.; Zhang, Y.; Zhang, Z.; and Zhang, Y. 2025. OT-StainNet: Optimal Transport Driven Semantic Matching for Weakly Paired H&E-to-IHC Stain Transfer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 3194–3202.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- He, Y.; Tu, B.; Jiang, P.; Liu, B.; Li, J.; and Plaza, A. 2024. IGroupSS-Mamba: Interval group spatial-spectral mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
- Hong, D.; Yokoya, N.; Chanussot, J.; and Zhu, X. X. 2018. An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Transactions on Image Processing*, 28(4): 1923–1938.
- Huang, X.; Dong, Y.; Zhang, Y.; and Du, B. 2025. Single-Source Frequency Transform for Cross-Scene Classification of Hyperspectral Image. *IEEE Transactions on Image Processing*.
- Hyttinen, J.; Fält, P.; Jäsberg, H.; Kullaa, A.; and Hauta-Kasari, M. 2020. Oral and dental spectral image database—odsi-db. *Applied Sciences*, 10(20): 7246.
- Jiang, Y.; Sha, H.; Liu, S.; Qin, P.; and Zhang, Y. 2023. AutoUnmix: an autoencoder-based spectral unmixing method for multi-color fluorescence microscopy imaging. *Biomedical Optics Express*, 14(9): 4814–4827.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673.
- Kumar, V.; Singh, R. S.; Rambabu, M.; and Dua, Y. 2024. Deep learning for hyperspectral image classification: A survey. *Computer Science Review*, 53: 100658.
- LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 2002. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278–2324.
- Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. M. 2017. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, 5542–5550.
- Li, L.; Gao, K.; Cao, J.; Huang, Z.; Weng, Y.; Mi, X.; Yu, Z.; Li, X.; and Xia, B. 2021. Progressive domain expansion network for single domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 224–233.
- Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; and Benediktsson, J. A. 2019. Deep learning for hyperspectral image classification: An overview. *IEEE transactions on geoscience and remote sensing*, 57(9): 6690–6709.
- Liu, C.; Cao, Y.; Su, X.; and Zhu, H. 2024a. Universal frequency domain perturbation for single-source domain generalization. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6250–6259.
- Liu, S.; Jin, X.; Yang, X.; Ye, J.; and Wang, X. 2024b. StyDeSty: Min-max stylization and destylization for single domain generalization. *arXiv preprint arXiv:2406.00275*.
- Qin, B.; Feng, S.; Zhao, C.; Xi, B.; Li, W.; and Tao, R. 2024. FDGNet: Frequency disentanglement and data geometry for domain generalization in cross-scene hyperspectral image

classification. *IEEE Transactions on Neural Networks and Learning Systems*.

Wang, J.; Lan, C.; Liu, C.; Ouyang, Y.; Qin, T.; Lu, W.; Chen, Y.; Zeng, W.; and Yu, P. S. 2022. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering*, 35(8): 8052–8072.

Wang, Z.; Luo, Y.; Qiu, R.; Huang, Z.; and Baktashmotlagh, M. 2021. Learning to diversify for single domain generalization. In *Proceedings of the IEEE/CVF international conference on computer vision*, 834–843.

Yang, Z.; Albrow-Owen, T.; Cai, W.; and Hasan, T. 2021. Miniaturization of optical spectrometers. *Science*, 371(6528): eabe0722.

Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; and Lin, L. 2018. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 701–710.

Zhang, W.; Suo, J.; Dong, K.; Li, L.; Yuan, X.; Pei, C.; and Dai, Q. 2023a. Handheld snapshot multi-spectral camera at tens-of-megapixel resolution. *Nature Communications*, 14(1): 5043.

Zhang, Y.; Li, W.; Sun, W.; Tao, R.; and Du, Q. 2023b. Single-source domain expansion network for cross-scene hyperspectral image classification. *IEEE Transactions on Image Processing*, 32: 1498–1512.

Zhang, Y.; Li, W.; Zhang, M.; Qu, Y.; Tao, R.; and Qi, H. 2021. Topological Structure and Semantic Information Transfer Network for Cross-Scene Hyperspectral Image Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.

Zhang, Y.; Li, W.; Zhang, M.; Wang, S.; Tao, R.; and Du, Q. 2022. Graph Information Aggregation Cross-Domain Few-Shot Learning for Hyperspectral Image Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.

Zhao, Y.; Wang, C.; Luo, X.; Luo, J.; Ju, W.; Xiao, Z.; and Zhang, M. 2025. TRACI: A Data-centric Approach for Multi-Domain Generalization on Graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 13401–13409.

Zheng, G.; Huai, M.; and Zhang, A. 2024. Advst: Revisiting data augmentations for single domain generalization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 21832–21840.

Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; and Loy, C. C. 2022. Domain generalization: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(4): 4396–4415.