Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DuTLR-Net: A Deep Unrolled Tensor Low-RankNetwork with Cross-Scale Manifold Alignment forHyperspectral Image Classification

Rui Zhang
Lanzhou University of Finance and Economics
Yongqi Chang
yongqichang202@gmail.com

Lanzhou University of Finance and Economics
Qiang Li
Lanzhou University of Finance and Economics
Xinchen Wang
Lanzhou University of Finance and Economics
Xiaowan Li
Qinghai Institute of Technology

Article

Keywords:

Posted Date: July 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9731081/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 1st, 2026.
See the published version at https://doi.org/10.1038/s41598-026-72316-1.

DuTLR-Net: A Deep Unrolled Tensor Low-Rank Network with Cross-Scale Manifold Alignment for Hyperspectral Image Classification

RuiZhang[1], YongqiChang[1,*], QiangLi[1], XinchenWang[1], and XiaowanLi[2]

[1]School of Information Engineering and Artificial Intelligence, Lanzhou University of Finance and Economics, Lanzhou, 730101, Gansu, China
[2]Qinghai Institute of Technology, Xining, 810016, Qinghai, China
[*]Corresponding author:yongqichang202@gmail.com

ABSTRACT

To address the issues of missing interpretability in existing deep learning models, weak spatial-spectral structure preservation, and poor generalization in small-sample-size scenarios, this paper proposes a Deep Unfolding Tensor Low-Rank Network (DuTLR-Net) that deeply integrates physical priors. First, a Cross-Scale Manifold Alignment (CSMA) module is designed, which uses a global PCA manifold as guidance to dynamically select high-variance spectral bands, achieving alignment of multi-scale features. Subsequently, the constructed ADMM-based deep unfolding backbone network explicitly models the spectral low-rankness and noise sparsity constraints as a differentiable iterative optimization process, enabling end-to-end learning through parameterized dictionaries and adaptive soft-thresholding. Experiments demonstrate that DuTLR-Net, with fewer parameters and faster convergence speed, exhibits superior classification accuracy and spatial consistency under extremely low labeling rates.code:https://github.com/yongqichang202-beep/DuTLR-Net

1 Introduction

Hyperspectral Image (HSI), with its unique capability to capture fine-grained spectral reflectance of ground objects across hundreds of continuous and narrow bands, has become one of the most valuable visual data modalities in Earth observation systems[1,2]. Compared to traditional multispectral images, the three-dimensional (3D) data cube of HSI not only contains complex spatial geometric textures but also possesses spectral diagnostic ability that reflects the deep physicochemical properties of target materials. This joint spatial–spectral characteristic makes HSI play an irreplaceable foundational role in fields such as precision agriculture, environmental monitoring, urban planning, and geological exploration[3]. However, the ultra-high-dimensional feature space of HSI also triggers the well-known "curse of dimensionality" (Hughes phenomenon). More critically, in real remote sensing applications, acquiring samples with high-quality pixel-level manual annotations is an extremely time-consuming, costly, and domain-specific task. The sharp contradiction between the high-dimensional feature space and the extremely limited labeled samples constitutes the long-standing "Small Sample Size (SSS)" problem in HSI classification tasks[19]. Furthermore, due to interference from atmospheric scattering and sensor thermal noise, real-world acquired HSIs are often accompanied by complex mixed noise, which further exacerbates the difficulty of extracting high-level semantic features[15]. To overcome the curse of dimensionality under small-sample conditions in the early stages, researchers primarily focused on combining shallow machine learning algorithms with dimensionality reduction techniques. Kernel methods represented by Support Vector Machines (SVM)[4,5], as well as manifold learning combined with Neighborhood Preserving Embedding (NPE)[6] and sparse representation techniques[7] , achieved fundamental breakthroughs in mining the intrinsic low-dimensional structure of HSI. Although these traditional methods possess rigorous mathematical and physical significance, due to their heavy reliance on handcrafted feature extractors, they struggle to capture highly discriminative deep semantic representations in highly complex nonlinear spatial–spectral coupling scenarios. Entering the era of Deep Learning (DL), purely data-driven neural networks have rapidly reshaped the paradigm of Hyperspectral Image (HSI) classification. From early stacked autoencoders[8] to 2D/3D Convolutional Neural Networks (CNNs) leveraging local receptive fields[9–11], and further to Graph Neural Networks (GNNs)[12], the nonlinear representation capacity of models has been pushed to new heights. In recent years, the Vision Transformer (ViT), capitalizing on its strengths in long-range modeling with a global receptive field (e.g., WaveFormer[16] and multi-scale convolutional Transformer[17]), has further broken multiple classification records[13,18]. However, with the deepening of research on remote sensing interpretability between 2025 and 2026, the academic community has begun to profoundly reflect on the limitations of purely data-driven deep models[20,21]. Existing deep HSI models have

exposed three major shortcomings in practical deployment: First is the severe data dependency and overfitting bottleneck. Faced with the extremely limited labeled samples in hyperspectral classification, deep networks with massive parameters struggle to converge effectively and are highly prone to representation collapse in high-dimensional feature spaces due to overfitting[22,23]. Second is the detachment from physical mechanisms and the lack of interpretability. Mainstream models generally reduce HSI to conventional tensors for black-box mapping, completely ignoring the physical priors constituted by spectral low-rankness and noise sparsity. This results in network decisions lacking solid physical credibility and mathematical guarantees[24]. Finally, there is the destruction of intrinsic tensor structure and spatial-spectral misalignment. During multi-scale downsampling or feature flattening operations, most networks forcibly disrupt the inherent 3D spatial-spectral topological consistency of HSI. The resulting cross-scale feature misalignment and spectral distortion directly lead to severe salt-and-pepper artifacts and a loss of spatial homogeneity in the final classification maps[25,26]. To bridge the gap between the "black-box" nature of deep learning and the "difficult-to-optimize" physical priors, "Algorithm Unrolling" emerged as a highly revolutionary interdisciplinary paradigm, rapidly becoming a cutting-edge hotspot in remote sensing interpretation during 2025-2026[14,27–29]. The core idea of algorithm unrolling is to explicitly map the iterative steps of classical optimization algorithms, such as the Alternating Direction Method of Multipliers (ADMM), into the physical layers of a deep neural network. Considering that Hyperspectral Images (HSI) are essentially high-dimensional 3D tensors, physical constraints based on Tensor Low-Rank (TLR) are widely recognized as the optimal approach for describing the global spectral correlation of HSI[30–32]. However, directly introducing tensor priors into deep unrolled networks for solving Small-sample-size HSI classification still faces two unresolved challenges: Firstly, traditional unrolled models employ fixed, hard-coded soft-thresholding operators, which are prone to causing excessive shrinkage of useful information when confronted with highly non-stationary and heterogeneous complex noise in real HSI[33]. Secondly, during the deep inference of multi-level networks, the lack of a top-down global manifold constraint mechanism inevitably leads to misalignment of fine-grained discriminative features as the local receptive field expands[34].

This paper proposes a novel deep unfolding tensor low-rank network, DuTLR-Net, designed to effectively integrate the advantages of physical priors and data-driven learning. The network unfolds an ADMM optimization process dominated by tensor low-rank and sparse priors into an end-to-end deep architecture. Its core innovations and main contributions are as follows: 1. This paper constructs a deep tensor unfolding architecture ADMM guided by physical priors, which explicitly transforms the low-rank and sparse priors of three-dimensional tensors into strong regularization constraints in network optimization. This design effectively addresses the data dependency and representation collapse issues of purely data-driven models under low annotation rates.

2. The proposed CSMA module utilizes the low-dimensional manifold extracted by global PCA as structural queries to guide and refine local features in a top-down manner. Through gated residual fusion, it precisely eliminates abnormal noise and ensures strict alignment between deep features and the original physical manifold, effectively repairing spectral distortions in multi-scale forward propagation.

3. This paper introduces a nonlinear network to replace traditional static threshold truncation, achieving a data-driven learnable tensor proximal operator. This module enables the model to perceive spatially heterogeneous noise and adaptively adjust the optimal sparsity intensity, overcoming the limitations of hard-coded operators in traditional optimization algorithms.

4. DuTLR-Net demonstrates superior quantitative metrics and high-quality qualitative results under extremely limited labeled sample constraints across multiple mainstream benchmark datasets, with lower parameter complexity and better convergence characteristics.

2 Methodology

This section elaborates on the proposed DuTLR-Net (Deep Unrolled Tensor Low-Rank Network) framework in detail. The input image tensor is denoted as $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$, where$H, W$ represent the height and width of the image, and C represents the number of channels. Our objective is to learn a temporal function $f : \mathscr{X} \to \mathscr{Y}$,which Xmaps to a class label $y \in \{1,\ldots,K\}$,where $K$is the number of classes.The overall processing pipeline can be expressed by the following formula:

$$\mathbf{Y} = f_{\text{cls}} \circ f_{\text{unroll}} \circ f_{\text{align}} \circ f_{\text{feat}}(\mathbf{X}_g, \mathbf{X}_l) \tag{1}$$

The overall architecture of the model is illustrated in Figure 1, comprising four core components: the dual-stream input processing for physical perception f_{feat}(stage1), cross-scale manifold alignment f_{align} (stage 2), a deep unfolding tensor low-rank backbone network f_{unroll} (stage 3), and classification with robust inference f_{cls} (stage 4).

2.1 Physics-Aware Dual-Stream Input and Hybrid Feature Extraction

The challenges of the curse of dimensionality and the small sample size problem in hyperspectral images (HSI) make it difficult to directly feed high-dimensional features into convolutional networks for processing. To address this, dimensionality reduction methods (such as PCA) are employed for feature preprocessing, and data decoupling is performed at the feature level, enabling

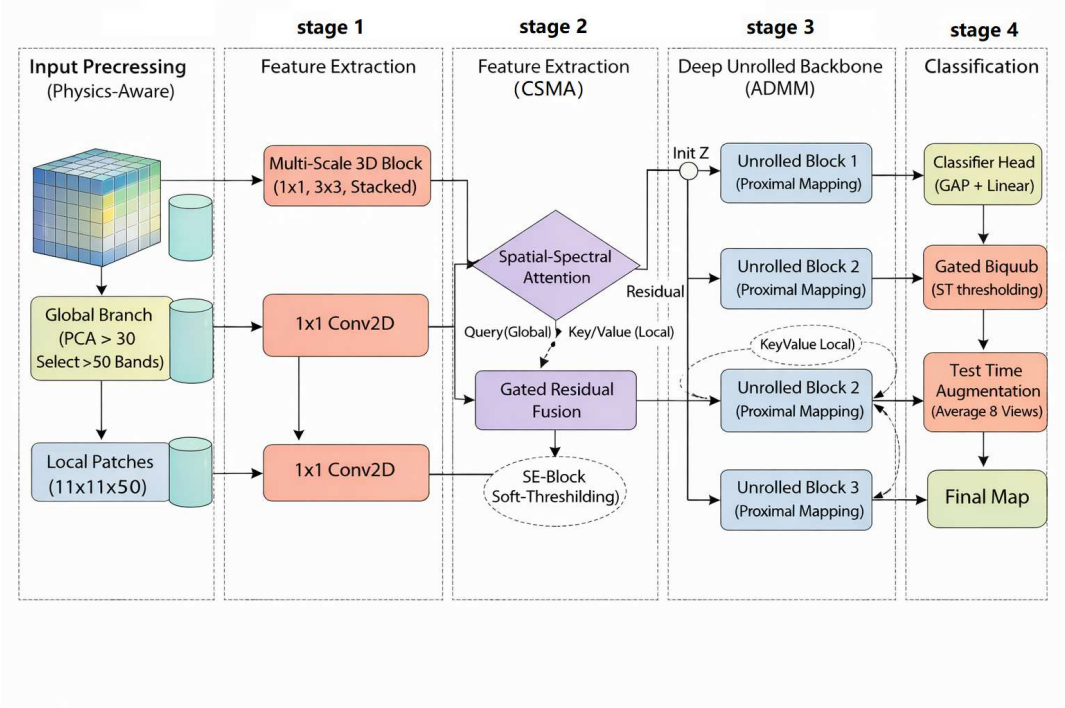


Figure 1. DuTLR-Net System Architecture Diagram

the spatial and spectral features of hyperspectral images to be processed separately by the network. To effectively tackle this issue, we propose the following scheme.

2.1.1 Physics-Aware Data Decoupling

Based on stage 1 of this model, we decouple the hyperspectral image into two data streams: a low-dimensional spatial structure and a high-dimensional representation of spectral features. To capture the macroscopic topological structure of the spectrum and remove redundant noise, we use principal component analysis (PCA) to generate the global stream. Let $\mathbf{W}_{pca} \in \mathbb{R}^{C \times C_g}$ be the projection matrix; then the global stream $\mathscr{X}_g$ is defined as:

$$\mathscr{X}_g = \mathscr{X} \times_3 \mathbf{W}_{pca}, \quad \mathscr{X}_g \in \mathbb{R}^{H \times W \times C_g} \tag{2}$$

where $\times_3$ denotes the mode-3 product along the spectral dimension, and C_g is the number of retained principal components (in this work, $C_g = 30$). $\mathscr{X}_g$ provides robust background information and captures the primary spectral variation trends. Simultaneously, we employ a variance-based band selection strategy. For the kth band $\mathbf{B}_k \in \mathbb{R}^{H \times W}$, its information content is measured by its variance σ_k^2:

$$\sigma_k^2 = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (\mathbf{B}_k(i,j) - \mu_k)^2 \tag{3}$$

We select the top C_l bands with the largest variance, forming an index set $\Omega = \{k \mid \text{rank}(\sigma_k^2) \leq C_l\}$ to construct the local stream:

$$\mathscr{X}_l = \mathscr{X}_{[:,:,\Omega]}, \quad \mathscr{X}_l \in \mathbb{R}^{H \times W \times C_l} \tag{4}$$

$\mathscr{X}_l$ explicitly preserves the original physical bands with high signal-to-noise ratios, compensating for the high-frequency details lost in the global stream.

2.1.2 Multi-Scale Hybrid Dimension Feature Extraction

Given that the decoupled dual-stream data possess distinct characteristics, we design a hybrid-dimension feature extraction strategy to synergistically leverage the volumetric modeling capability of 3D convolutions and the spatial encoding efficiency of 2D convolutions. For the global stream $\mathscr{X}_g$, we design a Multi-Scale 3D Block to extract Volumetric Features. This block

contains three parallel 3D convolutional branches, each with different spatio-spectral receptive fields, to address the intra-class scale variability of ground objects, as shown in stage I of Figure 1. Let $f_{3d}^{(k)}(\cdot)$ denote the transformation function of the k branch. The computation process for the global feature $\mathbf{F}_g$ can be expressed as the concatenation and fusion of the outputs from each branch.

$$\mathbf{F}_g = \phi\left(\text{Concat}\left[f_{3d}^{(1)}(\hat{\mathscr{X}_g}), f_{3d}^{(2)}(\hat{\mathscr{X}_g}), f_{3d}^{(3)}(\hat{\mathscr{X}_g})\right]\right) \tag{5}$$

Here, $\hat{\mathscr{X}_g}$ is the 5D tensor obtained by adding a channel dimension to $\mathscr{X}_g$, and $\phi(\cdot)$ denotes the convolution operation of the feature fusion layer. For the local stream $\mathscr{X}_l$, a lightweight 1×1 2D convolution is employed for feature projection, mapping it to the same latent space dimension as the global features:

$$\mathbf{F}_l = \text{Conv2D}_{1\times1}(\mathscr{X}_l) \tag{6}$$

This hybrid design provides a complementary and structurally clear feature basis for the subsequent Cross-Scale Manifold Alignment (CSMA).

2.2 Cross-Scale Manifold Alignment Module

To address this Feature Manifold Mismatch problem, we propose the Cross-Scale Manifold Alignment (CSMA) module[35]. This module is based on a "global query-local retrieval" attention paradigm, aiming to project high-frequency physical details onto a robust global low-rank manifold, as shown in stage 2 of Figure 1.

2.2.1 Attention-Based Manifold Projection

The core idea of CSMA is to utilize the clean topological structure contained in $\mathbf{F}_g$ as guidance to adaptively filter effective information from the noisy $\mathbf{F}_l$. Formally, let $\mathbf{F}_g \in \mathbb{R}^{H\times W\times C'}$ and $\mathbf{F}_l \in \mathbb{R}^{H\times W\times C'}$ be the input features. We first project them into a latent space via 1×1 convolutions to construct the Query, Key, and Value triplet.

$$\mathbf{Q} = \phi_q(\mathbf{F}_g), \quad \mathbf{K} = \phi_k(\mathbf{F}_l), \quad \mathbf{V} = \phi_v(\mathbf{F}_l) \tag{7}$$

Here, $\phi(\cdot)$ denotes the linear embedding function. Under this definition, $\mathbf{Q}$ represents the structural prior of the target manifold, while $\mathbf{K}$ represents the index of the source features. Subsequently, we compute the Spatial-Spectral Affinity Matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ to measure the correlation between global structure and local details:

$$\mathbf{A} = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right) \tag{8}$$

where d_k is the scaling factor. The matrix $\mathbf{A}$ essentially acts as a Manifold Alignment Operator. By applying this operator to $\mathbf{V}$, we generate the rectified local features $\mathbf{F}_{rect}$:

$$\mathbf{F}_{\text{rect}} = \mathbf{V}\mathbf{A}^\top, \quad \mathbf{F}_{rect} \in \mathbb{R}^{H\times W\times C'} \tag{9}$$

This process is mathematically equivalent to performing a non-local weighted reconstruction of the local features by leveraging the topological consistency of the global structure, thereby effectively suppressing spectral outliers and noise inconsistent with the global structure.

2.2.2 Adaptive Channel Reweighting and Gated Fusion

Although the attention mechanism achieves spatial alignment, the contribution of different spectral channels to the classification task varies. To further refine $\mathbf{F}_{rect}$, we introduce the Squeeze-and-Excitation (SE) module for channel-wise feature recalibration[36]. Let $\mathbf{z} \in \mathbb{R}^{C'}$ be the channel descriptor obtained through Global Average Pooling. The channel attention weights $\mathbf{w}_{se}$ are computed as follows:

$$\mathbf{w}_{se} = \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{z})) \tag{10}$$

Here, σ denotes the Sigmoid function, δ represents the ReLU activation function, and $\mathbf{W}_1, \mathbf{W}_2$ are the weights of the fully connected layers. The refined feature representation is given by $\tilde{\mathbf{F}}_{rect} = \mathbf{F}_{rect} \otimes \mathbf{w}_{se}$, where $\otimes$ denotes element-wise multiplication along the channel dimension. Finally, to mitigate gradient vanishing and preserve the original global structural information, we employ a Gated Residual Fusion strategy to generate the final aligned initialization tensor $\mathbf{X}_{init}$:

$$\mathbf{X}_{init} = \mathbf{F}_g + \gamma \cdot \tilde{\mathbf{F}}_{rect} \tag{11}$$

where γ is a learnable scaling parameter initialized to 0. This design ultimately provides a high-quality initialization tensor for the subsequent deep unrolled backbone.

2.3 Deep Unrolled Backbone via ADMM

To robustly recover discriminative low-rank semantics from high-dimensional tensors, we introduce the optimization concepts of sparse coding and low-rank representation into a deep learning framework. By explicitly unrolling the Alternating Direction Method of Multipliers (ADMM)[37] into a deep neural network, we construct the backbone of DLT-ADMM, as shown in Stage 3 of Fig. 1.

2.3.1 Variational Formulation

We formulate HSI feature learning as a tensor sparse recovery problem. Given the feature tensor $\mathbf{X}_{init}$ initialized by the CSMA module, our objective is to solve for a latent sparse feature map $\mathbf{Z}$ that can both reconstruct the input data and satisfy specific sparsity constraints. This optimization objective function can be formulated as:

$$\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{X}_{init} - \mathbf{D} * \mathbf{Z}\|_F^2 + \lambda \|\mathbf{Z}\|_1 \tag{12}$$

Here, $\mathbf{D}$ represents the dictionary to be learned (corresponding to the convolutional kernels in the network), $*$ denotes the convolution operation, $\|\cdot\|_F$ is the Frobenius norm, $\|\cdot\|_1$ is the ℓ_1 norm that induces sparsity, and λ is the regularization parameter. The first term is the Data Fidelity Term, used to preserve the integrity of the features; the second term is the Sparsity Prior Term, used to suppress noise and extract salient features. To solve the aforementioned convex optimization problem, we introduce an auxiliary variable $\mathbf{U}$, transforming the original problem into the Augmented Lagrangian Form under the ADMM framework:

$$\mathscr{L}(\mathbf{Z}, \mathbf{U}, \mathbf{M}) = \frac{1}{2} \|\mathbf{X}_{init} - \mathbf{D} * \mathbf{Z}\|_F^2 + \lambda \|\mathbf{U}\|_1 + \langle \mathbf{M}, \mathbf{Z} - \mathbf{U} \rangle + \frac{\rho}{2} \|\mathbf{Z} - \mathbf{U}\|_F^2 \tag{13}$$

where $\mathbf{M}$ is the Lagrange multiplier and ρ is the penalty parameter.

2.3.2 Deep Unrolling and Implementation

This section maps the three iterative steps of the ADMM algorithm into learnable neural network modules (Unrolled Blocks), constructing the backbone network.

1. Z-Update Subnet (Feature Reconstruction Layer) This performs the gradient descent update for the data fidelity term. Mathematically, the analytical solution for $\mathbf{Z}^{(k+1)}$ involves complex matrix inversion. In deep unrolling, we transform this into a learnable linear operation to approximate the gradient update:

$$\mathbf{Z}^{(k+1)} = \mathbf{Z}^{(k)} - \eta \left(\mathbf{W}_x * \mathbf{X}_{init} + \mathbf{W}_z * \mathbf{Z}^{(k)} + \mathbf{B}_z \right) \tag{14}$$

Among these,$\mathbf{W}_x, \mathbf{W}_z$ are learnable convolutional filters used for extracting input information and transmitting historical state information, respectively. $\mathbf{B}_z$ is a learnable bias term. This design simulates the gradient flow in the optimization process at the network level.

2. U-update Subnetwork (Sparse Threshold Mapping) This step corresponds to the ℓ_1 regularization term, whose closed-form solution is the Soft Thresholding Operator. We design this operator as a parameter-free activation layer but introduce a learnable threshold:

$$\mathbf{U}^{(k+1)} = \mathscr{S}_{\theta_k} \left(\mathbf{Z}^{(k+1)} + \frac{\mathbf{M}^{(k)}}{\rho} \right) \tag{15}$$

where the soft thresholding operator is defined as:

$$\mathscr{S}_{\theta}(x) = \text{sign}(x) \max\left(|x| - \theta, 0\right) \tag{16}$$

Unlike traditional algorithms that use a fixed threshold, we design a learnable parameter θ_k, allowing the network to adaptively determine the sparsity strength based on the local context of the data, thereby achieving an optimal balance between denoising and feature preservation.

3.M-update Subnetwork (Sparse Threshold Mapping) The update of the Lagrange multiplier corresponds to a residual feedback mechanism in the network, used to correct feature deviations:

$$\mathbf{M}^{(k+1)} = \mathbf{M}^{(k)} + \rho \left(\mathbf{Z}^{(k+1)} - \mathbf{U}^{(k+1)} \right) \tag{17}$$

In the network implementation, we simplify it to residual connections of feature maps and linear transformations along the channel dimension to accelerate convergence. Through the above design, each layer of the DLT-ADMM-Net possesses clear physical meaning: the convolutional layer is responsible for learning the dictionary $\mathbf{D}$ to fit the data structure, the activation function (soft thresholding) is responsible for enforcing sparse constraints, and the residual connections are responsible for transmitting gradients and correcting errors. Compared to blind black-box modeling, this model-guided architecture not only significantly reduces the number of parameters but, more importantly, introduces low-rank and sparse inductive biases, endowing it with theoretically guaranteed robustness when dealing with the complex noise and small-sample conditions of hyperspectral images.

2.4 Classification and Robust Inference

After iterative optimization via the deep unfolding backbone network, we obtain a deep feature tensor $\mathbf{Z}^{(K)} \in \mathbb{R}^{H\times W\times C_{out}}$ that is highly discriminative and conforms to low-rank physical constraints. First, to enhance spatial invariance and reduce the number of parameters, we apply Global Average Pooling (GAP) to $\mathbf{Z}^{(K)}$, followed by linear projection through a fully-connected layer. For the i sample, the predicted probability $p_{i,k}$ that it belongs to the k class is given by the Softmax function:

$$p_{i,k} = \frac{\exp\left(\mathbf{w}_k^\top \mathbf{z}_i + b_k\right)}{\sum_{j=1}^{C} \exp\left(\mathbf{w}_j^\top \mathbf{z}_i + b_j\right)} \tag{18}$$

where $\mathbf{w}_k$ and b_k are the weight vector and bias term of the k class, respectively. Meanwhile, we introduce a Label Smoothing strategy as a regularization technique, transforming the original one-hot label distribution $\mathbf{y}_k$ into a smoothed distribution $\tilde{\mathbf{y}}_k$:

$$\tilde{\mathbf{y}}_k = (1-\varepsilon)\, y_k + \frac{\varepsilon}{C} \tag{19}$$

where ε is the smoothing factor (set to 0.1 in this experiment) and C is the total number of classes. The final optimization objective is defined as the Kullback–Leibler divergence between the predicted distribution and the smoothed label distribution:

$$\mathscr{L}_{cls} = -\sum_{k=1}^{C} y_k^{ls} \log\left(p_{i,k}\right) \tag{20}$$

Furthermore, during the testing phase, predictions from a single view are often susceptible to variations in HSI imaging orientation and local texture differences, which can lead to prediction jitter. To further improve inference stability, we adopt a Test-Time Augmentation (TTA) strategy[38], as illustrated in stage 4 of Fig.1. The core idea of this strategy is to apply multiple geometric transformations to the input sample while keeping the label unchanged. Specifically, during inference, we define a set of geometric transformations $\mathscr{T} = \{T_1, \ldots, T_M\}$ containing $M = 8$ viewpoint transformations (i.e., rotations of $0^\circ, 90^\circ, 180^\circ, 270^\circ$ and their corresponding horizontal flips). For a given test sample $\mathbf{X}_{test}$, the final prediction $\hat{y}$ is obtained by aggregating the posterior probabilities from all transformed views:

$$\hat{y} = \arg\max_k \left(\frac{1}{M} \sum_{m=1}^{M} \mathscr{T}_m^{-1}\left(f_{\text{net}}(\mathscr{T}_m(\mathbf{X}_{test}))\right) \right) \tag{21}$$

where $f_{\text{net}}(\cdot)$ denotes the forward propagation function of the trained DLT-ADMM-Net, and $\mathscr{T}_m^{-1}(\cdot)$ maps the prediction back to the original coordinate system. This ensemble approach effectively smooths out prediction fluctuations caused by local noise, significantly improving the spatial consistency and boundary integrity of the classification.

3 Experimental Setup

To systematically evaluate the classification performance of the proposed model under varying conditions of spatial resolution, class scale, and spectral complexity, three representative public hyperspectral datasets—Salinas, Pavia University (PaviaU), and Botswana—were selected for experimental validation. In the experimental design, the proposed model was compared against several typical deep learning methods, including PyResnet34, Resnet34[39], S3KAIResNet[41], SSRN[40], DBDA[42], DBMA[43], FDSCN[44], SSEFN[45], and NESSGGCN[46]. All comparative models strictly adopted the optimal parameters reported in their original papers to ensure fairness in the experimental comparison. For the DuLTR model, a systematic parameter tuning experiment was conducted to investigate the impact of *batch_size* and *patch_size* on performance (as shown in 2. The experimental results indicated that the model exhibited stable and excellent performance metrics under the selected parameters; therefore, the batch size was ultimately set to $batch_size = 128$ and $patch_size = 13$. Furthermore, to reduce the spectral

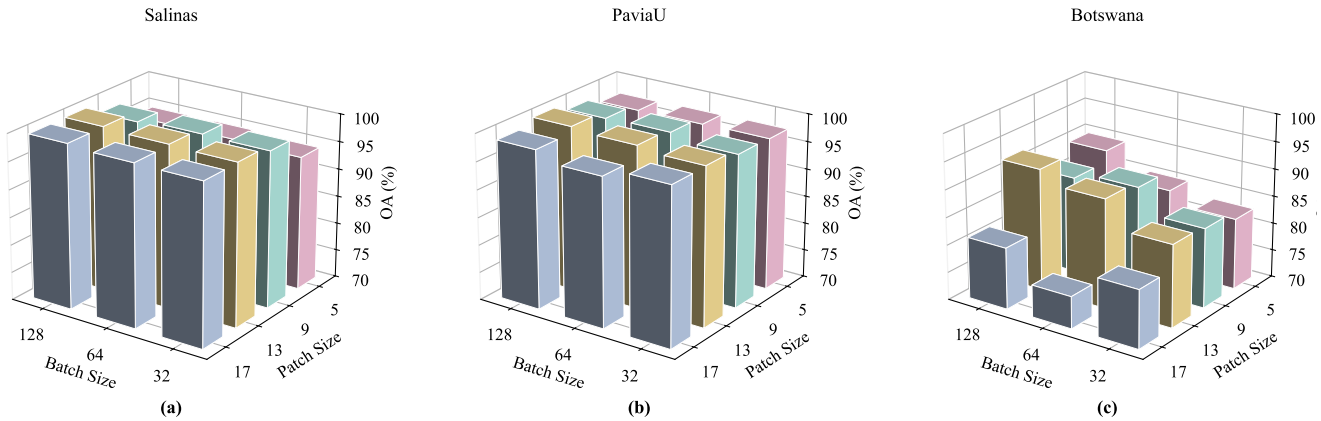


Figure 2. Overall classification accuracy with different parameters on 3 datasets

redundancy of the hyperspectral data and decrease computational overhead, Principal Component Analysis (PCA) was applied for dimensionality reduction on the raw data before training, retaining the first 30 principal components. In the model training phase, the maximum number of training epochs was set to 200, with an initial learning rate of 1×10^{-3}. To further enhance training stability and suppress overfitting, an Early Stopping mechanism was introduced: training was terminated prematurely if no improvement in validation set performance was observed for 20 consecutive epochs. The training sample size for all models was uniformly set as follows: 0.5% of the samples for the Salinas and PaviaU datasets, and 1% for the Botswana dataset. The spatial distribution of training samples for each dataset is shown in 1. All experiments were conducted on a single-card computing environment equipped with an NVIDIA GeForce RTX 2080 GPU to ensure the reproducibility of the experiments and the reliability of the results.

Table 1. Salinas PaviaU Botswana Land cover classes and sample sizes

ID	Salinas			PaviaU			Botswana		
	Total	Train	Test	Total	Train	Test	Total	Train	Test
1	6631	32	6599	2009	13	1996	270	2	268
2	18649	72	18577	3726	19	3707	101	2	99
3	2099	14	2085	1976	12	1964	251	2	249
4	3064	18	3046	1394	10	1384	215	2	213
5	1345	11	1334	2678	15	2663	269	2	267
6	5029	26	5003	3959	20	3939	269	3	266
7	1330	11	1319	3579	18	3561	259	2	257
8	3682	21	3661	11271	42	11229	203	2	201
9	947	9	938	6203	27	6176	314	3	311
10	-	-	-	3278	17	3261	248	2	246
11	-	-	-	1068	8	1060	305	3	302
12	-	-	-	1927	12	1915	181	2	179
13	-	-	-	916	8	908	268	3	265
14	-	-	-	1070	8	1062	95	2	93
15	-	-	-	7268	30	7238	-	-	-
16	-	-	-	1807	12	1795	-	-	-
Total	42776	214	42562	54129	271	53858	3248	32	3216

4 Experimental Results and Analysis

4.1 Main Results

Tables 2–4 present the quantitative classification results of different models on three public hyperspectral datasets: Salinas, PaviaU, and Botswana. Overall, DUTLR achieves the best or highly competitive performance across evaluation metrics such as Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient. On the Salinas dataset, DUTLR attains OA, AA, and Kappa values of 96.27%,98.05%, and 95.84%, respectively; on the PaviaU dataset, these values are95.85%,95.07% and 94.53%; and on the Botswana dataset, they are 94.49%,94.64% and 93.69%. Compared to other benchmark methods, the proposed model demonstrates superior comprehensive performance across multiple datasets, indicating its effectiveness in modeling discriminative spatial–spectral features within hyperspectral data. Furthermore, from the perspective of result stability, DUTLR exhibits significantly smaller performance fluctuations across the three datasets. For instance, on the Salinas dataset, the standard deviation of its OA is merely 0.1896, indicating that the model possesses stronger robustness and consistency across different experimental partitions. Moreover, compared to methods such as SSEFN and NESSGGCN, DUTLR achieves superior or more stable performance in terms of OA, AA, and Kappa metrics. This demonstrates its enhanced ability to improve

the consistency between classification results and ground-truth annotations, thereby highlighting the model's reliability and superiority in hyperspectral image classification tasks.

Table 2. Classification metrics of different methods on the Salinas dataset

label	PyResnet34	Resnet34	S3KAIResNet	SSRN	DBDA	DBMA	FDSSC	SSEFN	NESSGGCN	DUTLR
1	86.11 ± 0.099	79.51 ± 0.216	94.77 ± 0.073	99.98 ± 0.000	100.00 ± 0.000	100.00 ± 0.000	92.49 ± 0.106	99.89	99.97±0.06	**100.0 ± 0.021**
2	93.98 ± 0.072	66.44 ± 0.182	98.87 ± 0.016	90.90 ± 0.100	98.85 ± 0.014	99.78 ± 0.001	99.85 ± 0.002	82.00	99.96±0.06	**100.0 ± 0.053**
3	81.19 ± 0.102	48.05 ± 0.065	93.70 ± 0.070	66.03 ± 0.356	96.92 ± 0.013	92.53 ± 0.057	63.39 ± 0.450	98.45	83.23±8.51	**100.0 ± 0.162**
4	94.21 ± 0.044	43.86 ± 0.273	96.78 ± 0.031	91.80 ± 0.074	91.55 ± 0.013	88.17 ± 0.078	95.57 ± 0.019	99.04	**98.75±1.45**	97.25 ± 0.328
5	80.44 ± 0.131	53.69 ± 0.162	92.15 ± 0.101	90.84 ± 0.060	90.27 ± 0.079	89.89 ± 0.088	84.51 ± 0.093	**99.32**	99.24±0.53	**99.44 ± 0.062**
6	98.98 ± 0.010	82.10 ± 0.119	99.34 ± 0.007	95.67 ± 0.050	99.96 ± 0.001	98.52 ± 0.009	99.71 ± 0.003	99.96	**100.0±0.00**	99.39 ± 0.011
7	94.94 ± 0.034	53.69 ± 0.386	99.48 ± 0.001	98.56 ± 0.012	98.92 ± 0.012	99.11 ± 0.013	88.91 ± 0.132	96.77	**99.9±0.02**	99.97 ± 0.099
8	79.63 ± 0.052	69.08 ± 0.147	80.25 ± 0.114	70.07 ± 0.104	88.45 ± 0.041	82.99 ± 0.040	79.22 ± 0.147	**99.98**	97.17±2.17	94.61 ± 0.000
9	76.72 ± 0.065	78.32 ± 0.139	98.69 ± 0.006	82.98 ± 0.066	91.76 ± 0.086	99.62 ± 0.003	83.35 ± 0.154	95.97	100.00±0.01	**100.0 ± 0.010**
10	93.05 ± 0.042	64.21 ± 0.152	93.42 ± 0.023	93.35 ± 0.076	**96.52 ± 0.016**	76.11 ± 0.155	77.69 ± 0.133	91.43	96.47±5.15	96.01 ± 0.045
11	78.87 ± 0.164	70.90 ± 0.175	95.07 ± 0.008	58.08 ± 0.423	93.01 ± 0.080	96.27 ± 0.027	33.13 ± 0.469	98.05	92.43±11.19	**100.0 ± 0.038**
12	65.38 ± 0.265	17.00 ± 0.240	98.01 ± 0.012	83.09 ± 0.239	99.52 ± 0.005	98.39 ± 0.017	98.92 ± 0.013	97.11	98.45±2.88	**99.95 ± 0.022**
13	66.63 ± 0.092	38.11 ± 0.169	96.33 ± 0.046	65.12 ± 0.116	**99.78 ± 0.002**	96.92 ± 0.028	98.99 ± 0.004	92.54	97.41±3.91	99.56 ± 0.023
14	82.05 ± 0.195	63.89 ± 0.074	77.29 ± 0.047	84.76 ± 0.059	97.21 ± 0.012	98.06 ± 0.015	84.12 ± 0.183	97.54	93.14±4.97	**98.87 ± 0.870**
15	53.87 ± 0.100	32.91 ± 0.118	84.43 ± 0.105	53.70 ± 0.270	70.70 ± 0.139	76.89 ± 0.078	70.42 ± 0.125	**99.33**	97.84±1.18	83.68 ± 0.001
16	96.38 ± 0.051	70.58 ± 0.238	99.38 ± 0.009	98.07 ± 0.016	100.00 ± 0.000	99.42 ± 0.007	99.34 ± 0.009	99.89	99.53±0.49	**100.0 ± 0.003**
OA	76.36 ± 0.038	56.42 ± 0.098	89.87 ± 0.025	69.15 ± 0.206	89.64 ± 0.038	89.70 ± 0.030	83.62 ± 0.098	93.84	**97.82±0.71**	96.27 ± 0.190
AA	82.65 ± 0.022	58.27 ± 0.120	93.62 ± 0.005	82.69 ± 0.062	94.59 ± 0.011	93.29 ± 0.013	84.35 ± 0.104	96.34	97.10±1.18	**98.05 ± 0.040**
Kappa	73.80 ± 0.040	51.05 ± 0.117	88.67 ± 0.028	65.21 ± 0.234	88.50 ± 0.042	88.54 ± 0.033	81.54 ± 0.112	93.16	**97.57±0.79**	95.84 ± 0.006

Table 3. Classification metrics of different methods on the PaviaU dataset

label	PyResnet34	Resnet34	S3KAIResNet	SSRN	DBDA	DBMA	FDSSC	SSEFN	NESSGGCN	DUTLR
1	51.01 ± 0.086	54.44 ± 0.075	75.48 ± 0.022	56.66 ± 0.174	93.75 ± 0.016	90.43 ± 0.011	91.86 ± 0.101	97.85	97.65±1.12	**98.23 ± 0.001**
2	61.78 ± 0.110	56.13 ± 0.063	86.75 ± 0.057	90.10 ± 0.049	98.32 ± 0.012	93.02 ± 0.028	92.26 ± 0.042	99.77	**99.67±0.30**	98.55 ± 0.000
3	87.18 ± 0.111	79.40 ± 0.213	76.63 ± 0.017	54.89 ± 0.221	96.84 ± 0.041	89.37 ± 0.062	40.51 ± 0.289	62.04	61.04±6.54	97.42 ± 0.009
4	93.21 ± 0.060	77.14 ± 0.138	96.70 ± 0.018	95.45 ± 0.030	**98.73 ± 0.001**	98.32 ± 0.005	98.68 ± 0.016	88.81	88.31±2.45	**97.87 ± 0.015**
5	96.02 ± 0.034	93.52 ± 0.049	99.74 ± 0.001	98.74 ± 0.011	99.80 ± 0.002	94.68 ± 0.073	99.75 ± 0.002	100.00	**100.0±0.000**	99.85 ± 0.029
6	38.23 ± 0.204	25.89 ± 0.191	87.63 ± 0.044	85.84 ± 0.098	95.82 ± 0.017	87.12 ± 0.110	92.11 ± 0.089	96.17	95.87±2.01	**99.98 ± 0.007**
7	42.40 ± 0.415	42.28 ± 0.408	88.34 ± 0.013	95.44 ± 0.033	**97.12 ± 0.024**	90.71 ± 0.083	64.40 ± 0.456	81.65	80.65±12.52	92.87 ± 0.055
8	51.55 ± 0.135	35.18 ± 0.114	60.68 ± 0.063	74.38 ± 0.117	82.19 ± 0.049	72.13 ± 0.072	72.13 ± 0.081	95.96	**95.66±1.24**	90.25 ± 0.377
9	53.60 ± 0.303	36.18 ± 0.453	85.39 ± 0.077	66.54 ± 0.180	96.81 ± 0.021	92.05 ± 0.086	96.93 ± 0.030	94.48	94.08±5.88	**97.63 ± 0.043**
OA	54.94 ± 0.038	53.79 ± 0.051	83.12 ± 0.035	72.23 ± 0.059	95.54 ± 0.011	89.71 ± 0.035	89.08 ± 0.051	95.35	95.15±0.63	**95.85 ± 0.194**
AA	63.89 ± 0.107	55.57 ± 0.088	84.15 ± 0.015	79.78 ± 0.019	**95.49 ± 0.009**	89.76 ± 0.043	83.18 ± 0.121	93.14	90.33±2.26	95.07 ± 0.0407
Kappa	30.93 ± 0.109	72.06 ± 0.112	76.79 ± 0.052	62.61 ± 0.077	94.07 ± 0.015	86.20 ± 0.048	85.26 ± 0.070	93.87	93.53±0.85	**94.23 ± 0.070**

4.2 Visualization Analysis of Classification Results

Figure 3–5 present the pseudo-color images (*HSI_Pseudocolor*), ground truth (GT), and classification results of different methods for three hyperspectral datasets. The visual results indicate that traditional convolutional network models, PyResNet and ResNet, are prone to local misclassification, class omission, and boundary blurring. Particularly in regions with spectrally similar classes or complex spatial distributions, their classification results exhibit noticeable discontinuities. S3KAIResNet and SSRN show improved overall classification performance, but still suffer from a certain degree of "salt-and-pepper noise," limited ability to characterize spatial structural information, and insufficiently refined discrimination in boundary regions. DBDA, DBMA, and FDSSC achieve high classification accuracy in most areas, yet they still face issues of incomplete regional partitioning in spectrally similar or complex distribution areas.In contrast, SSEFN and NESSGGCN exhibit enhanced modeling of local spatial features, which helps suppress noise to some extent. However, they still demonstrate a degree of instability or over-smoothing in complex cross-regional scenarios. Comparatively, DUTLR demonstrates superior visual consistency and structural preservation capability in the classification results across the three datasets. Specifically, its classification results exhibit clearer class boundaries, stronger intra-region consistency, and significantly reduced noise interference. Moreover, in areas with spectral similarity or complex spatial structures, DUTLR maintains stable and continuous classification outputs, showing higher consistency with the ground truth distribution. Additionally, DUTLR more completely preserves the fine-grained structural information embedded in the joint spatial-spectral features, reflecting its effectiveness and robustness in feature representation and noise suppression. Overall, DUTLR exhibits stronger structural modeling capability and result stability in complex scenarios. Its visual results align more closely with the true distribution, further validating the model's superior performance in hyperspectral image classification tasks.

Table 4. Classification metrics of different methods on the Botswana dataset

label	PyResnet34	Resnet34	S3KAIResNet	SSRN	DBDA	DBMA	FDSSC	SSEFN	NESSGGCN	DUTLR
1	45.14	14.79	62.92	59.95	89.76	97.41	95.64	96.39	86.57±11.53	**100.00**
2	41.84	43.75	92.38	81.71	97.03	91.67	**98.92**	99.16	53.20±18.77	98.86
3	45.45	16.53	91.10	100.00	100.00	100.00	100.00	100.00	76.55±14.58	**100.00**
4	24.07	30.22	91.87	37.48	93.64	91.74	100.00	100.00	86.29±9.91	**100.00**
5	13.13	11.20	88.07	45.85	82.28	84.32	84.11	84.81	64.94±26.88	**100.00**
6	0.00	40.91	76.92	94.83	94.55	70.38	64.71	65.43	56.33±16.15	**97.94**
7	100.00	67.73	95.47	94.76	100.00	100.00	85.76	87.33	86.15±13.28	**100.00**
8	83.33	100.00	97.86	98.96	100.00	100.00	96.55	98.13	62.49±25.11	**100.00**
9	92.86	90.00	64.62	78.14	82.13	95.61	99.36	**99.36**	51.09±21.83	97.19
10	65.66	35.00	76.70	85.06	100.00	85.31	100.00	100.00	61.14±21.36	**100.00**
11	100.00	85.71	100.00	98.01	99.67	100.00	100.00	**100.00**	95.12±9.44	97.83
12	20.00	0.00	94.55	94.89	**100.00**	92.63	94.12	94.39	73.44±17.91	90.68
13	13.84	87.50	72.16	92.98	100.00	100.00	98.80	98.80	86.54±9.92	**100.00**
14	100.00	95.00	100.00	100.00	98.88	100.00	100.00	**100.00**	91.49±7.15	96.34
OA	53.00	28.00	82.00	73.60	92.85	94.62	91.16	92.28	73.74±2.61	**94.49**
AA	53.23	57.40	86.04	83.04	93.56	93.50	94.14	91.59	73.67±2.43	**94.64**
Kappa	18.47	21.79	80.81	71.44	93.17	92.25	91.63	90.42	71.55±2.83	**93.69**

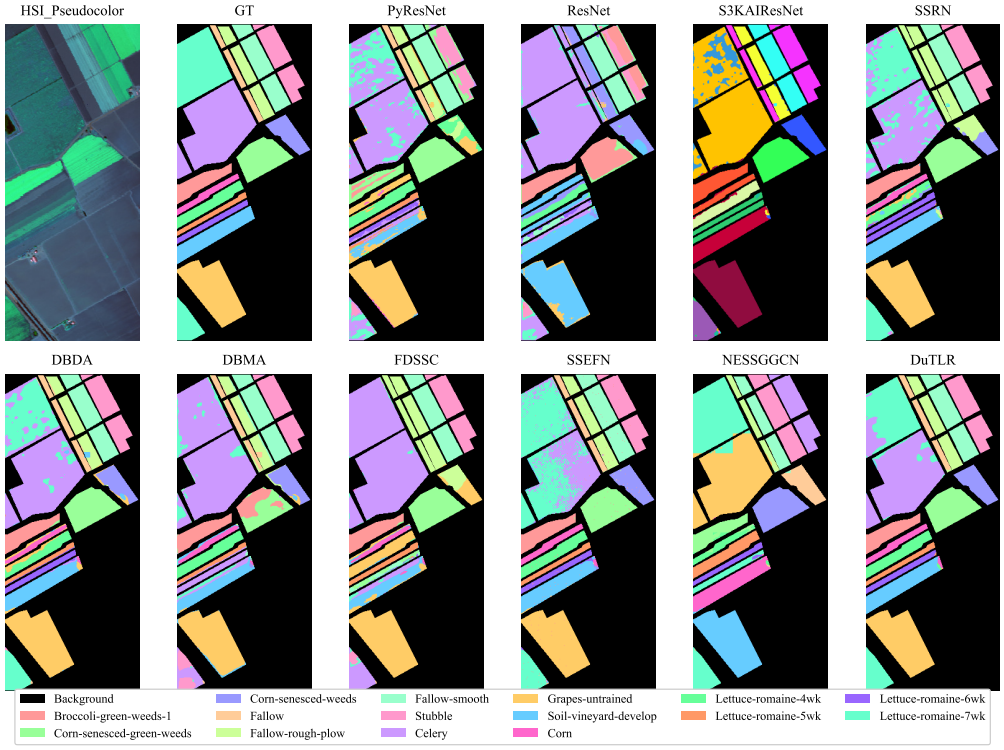


Figure 3. Classification results of different methods on the Salinas dataset

4.3 Small-sample-size Experiments

4.3.1 Small-sample-size Setup

Under the experimental setting with limited labeled samples, to ensure comparability among different methods, this paper adopts a uniform training configuration for all models. The *batch_size* is set to 32; the maximum number of training epochs is set to 100; and the initial learning rate is uniformly set to 1×10^{-4}. To systematically evaluate the robustness of the models under varying degrees of annotation scarcity, experiments are conducted under multiple training proportion settings. Specifically, the training sample proportions for the Salinas and PaviaU datasets are set to [0.1%,0.5%,1%,5%,10%,15%], while for the

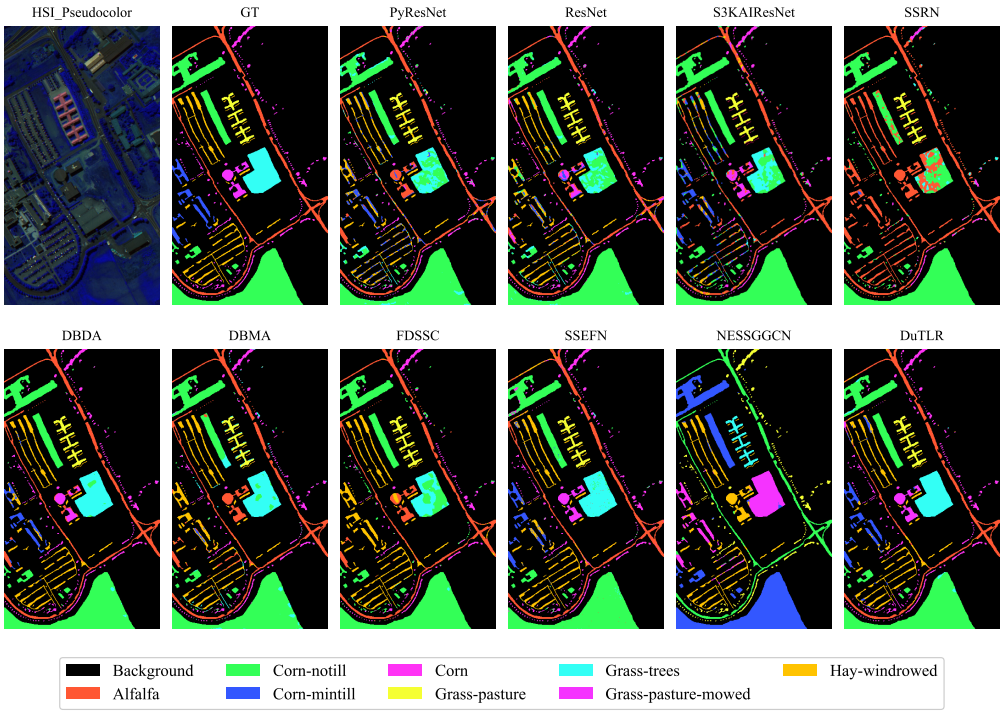


Figure 4. Classification results of different methods on the PaviaU dataset

Botswana dataset, they are set to [0.5%,1%,5%, 10%,15%,20%]. For each proportion, a stratified random sampling strategy is employed to construct a training subset with a balanced class distribution, with the remaining samples used for testing. The overall accuracy (OA) is taken as the unified evaluation metric to quantitatively measure the classification performance of each model under different training proportions.

4.3.2 Small-sample-size Experimental Results and Analysis

The overall classification accuracy (OA) of DuTLR (red curve) and the comparison methods under varying numbers of training samples was evaluated on three hyperspectral datasets, as shown in 6. Overall, the OA of all methods improves as the number of training samples increases, which aligns with the general pattern of deep learning models: more labeled data enhances the ability to model spatial–spectral features. However, significant differences exist among different methods in terms of improvement magnitude and final accuracy. Under extremely low sample sizes (Sample size $< 5\%$), most comparison methods (e.g., PyResNet34, ResNet34) struggle to learn discriminative features due to insufficient labeled information, resulting in relatively low OA performance. In contrast, DuTLR maintains a high OA even under low-sample conditions, demonstrating excellent Small-sample-size learning capability. As the sample size increases, the OA of all methods gradually converges, yet DuTLR consistently remains superior. Under moderate sample sizes (5–10%), DuTLR's OA is already significantly higher than that of methods such as DBDA, DBMA, S3KAIResNet, and NESSGGCN. Under high sample sizes (Sample size $> 10\%$), its OA continues to improve steadily, indicating that the adaptive attention mechanism can further capture fine-grained spatial–spectral correlations and optimize classification boundaries. Overall, DuTLR exhibits outstanding OA performance across the entire sample size range, with a particularly clear advantage under Small-sample-size conditions. The combination of active learning and attention mechanisms not only efficiently utilizes scarce labeled data but also further enhances performance when data is abundant.

4.3.3 Analysis of Training Time and Sample Size Relationship

The total training time of DuTLR and each comparison method under different training sample proportions was evaluated on three hyperspectral datasets, as shown in 7. Overall, under low sample proportions ($< 1\%$), the differences in training time among the methods are small. As the sample scale increases, the computational cost among different methods gradually shows a significant divergence trend. Specifically, supervised deep learning models such as PyResNet34, ResNet34, and SSRN exhibit a clear non-linear increase in training time as the sample size grows. On the Salinas dataset, when the training sample proportion reaches 15%, their training time exceeds 2000 seconds, primarily due to the computational burden of performing complete multi-epoch iterative training on large-scale samples. In contrast, methods such as DBDA, DBMA, and NESSGGCN show a relatively moderate increase in training time as the sample scale grows, which partly benefits from their optimization in sample

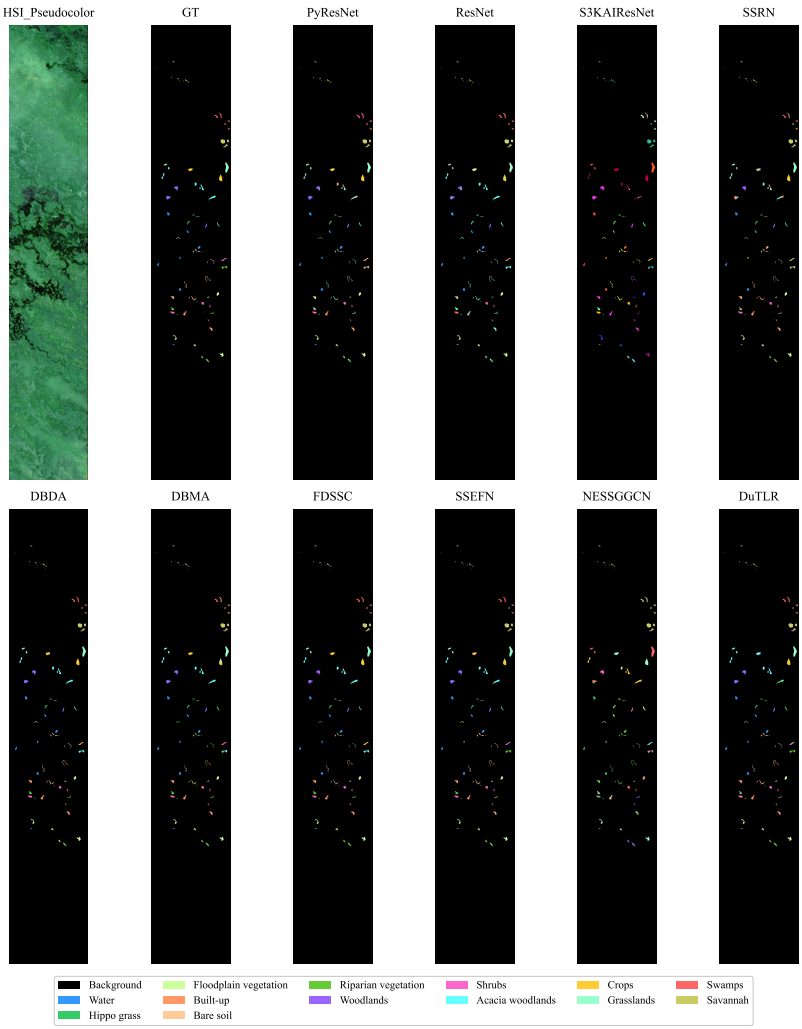


Figure 5. Classification results of different methods on the Botswana dataset

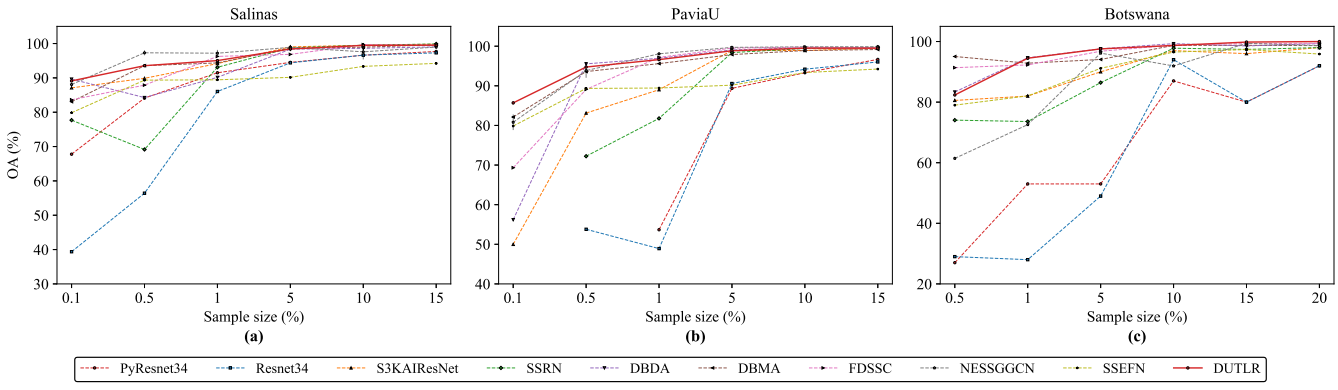


Figure 6. Overall classification accuracy of different methods under different labeled sample conditions

utilization or computational mechanisms. By comparison, DuTLR demonstrates lower training time overhead across all datasets and exhibits a more gradual increase as the sample scale grows. Even under a 15% sample proportion, its training time remains significantly lower than most comparison methods. This advantage is mainly attributed to its efficient adaptive spatial–spectral feature modeling mechanism and lightweight network architecture, which enable the model to focus on discriminative regions and reduce redundant computations. On the PaviaU dataset, this advantage is particularly prominent: as the sample proportion

increases, the training time of SSEFN and DBMA rises rapidly, exceeding 900 seconds under the 15% condition, while DuTLR consistently maintains low computational overhead. On the Botswana dataset, DuTLR's training time is also lower than that of the vast majority of comparison models, further validating its stable computational efficiency under varying data scales and complexity conditions.

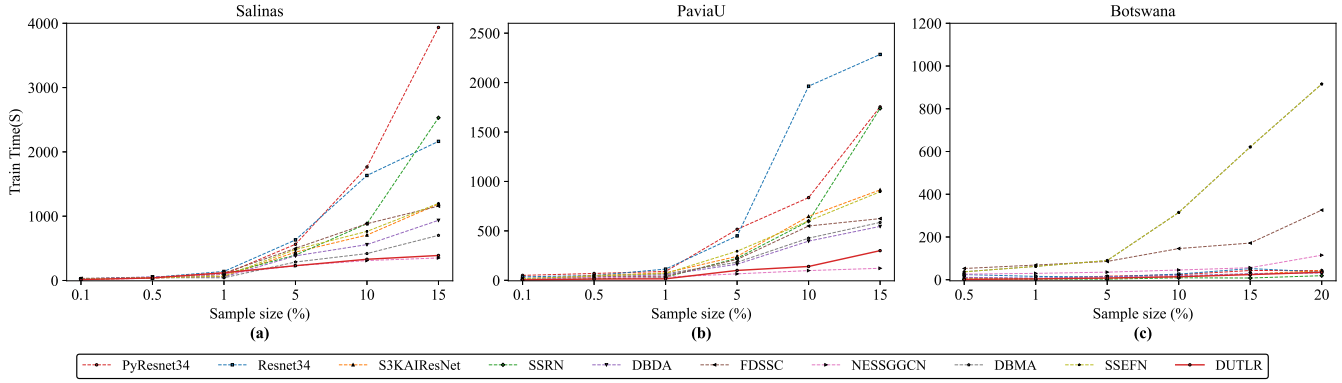


Figure 7. Comparison of time complexity of different methods on 3 datasets

5 Ablation Experiments

To systematically evaluate the contributions of each core module in the proposed method, we conducted ablation experiments on the Salinas, Pavia University, and Botswana datasets. The experimental design includes four model configurations: the complete model DuTLR-Net (Full), the model without the iterative unrolling module (w/o Unrolling), the model without the cross-spectral modality attention module (w/o CSMA), and the model without the 3D convolution module (w/o 3D-Conv). 8 illustrates the classification performance of each configuration on the three datasets, including the Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient. The results indicate that removing any module leads to a decline in classification performance, demonstrating the irreplaceable role of each module in feature extraction and discrimination. Among them, removing the iterative unrolling module has the most significant impact on performance. For instance, on the Botswana dataset, OA, AA, and Kappa decreased from 82.3%, 82.8%, and 81.1% to 74.1%, 73.9%, and 72.0%, respectively, fully highlighting the core role of the Unrolling module in optimizing feature representation and enhancing classification accuracy. Removing the CSMA module reduced the OA, AA, and Kappa on the Salinas dataset from 94.5%, 97.0%, and 93.5% to 93.7%, 96.8%, and 92.8%, validating its effectiveness in capturing key spectral dependencies and improving fine-grained spectral discriminative capability. After removing the 3D convolution module, the OA, AA, and Kappa on the Botswana dataset dropped to 75.8%, 76.5%, and 73.8%, respectively, indicating that this module plays a crucial role in spatial-spectral feature fusion and neighborhood context utilization. Furthermore, the complete model exhibited low standard deviation across multiple repeated experiments, demonstrating good stability and robustness in its training process. Overall, the ablation experiment results fully validate the complementarity and synergistic effects of each module, providing strong evidence for DuTLR-Net's superior performance in hyperspectral image classification tasks.

6 Summary

Hyperspectral image (HSI) classification faces core challenges such as the curse of dimensionality, scarcity of labeled samples, and inconsistency between spatial and spectral features. To address the issues of existing deep learning methods, which are prone to overfitting under small-sample conditions and lack physical interpretability, this paper proposes an innovative deep unfolding network—DuTLR-Net. This model ingeniously integrates the prior knowledge of traditional optimization algorithms with the powerful expressive capability of modern deep learning. The core idea of DuTLR-Net is to unfold the Alternating Direction Method of Multipliers (ADMM) based on tensor low-rank and sparse priors into a deep neural network. This design not only endows the network with clear physical significance but also enables it to maintain robustness under extremely limited data availability. To address the scale misalignment between global spectral structure and local spatial details, the authors designed a Cross-Scale Manifold Alignment (CSMA) module. This module employs a global query-local retrieval attention mechanism, utilizing clean global topological structures to guide and denoise noisy local details, effectively suppressing "salt-and-pepper noise" in the classification results. Furthermore, the model adopts a physics-aware dual-stream input strategy, separately extracting robust global principal components via PCA and selecting high-variance bands to preserve discriminative

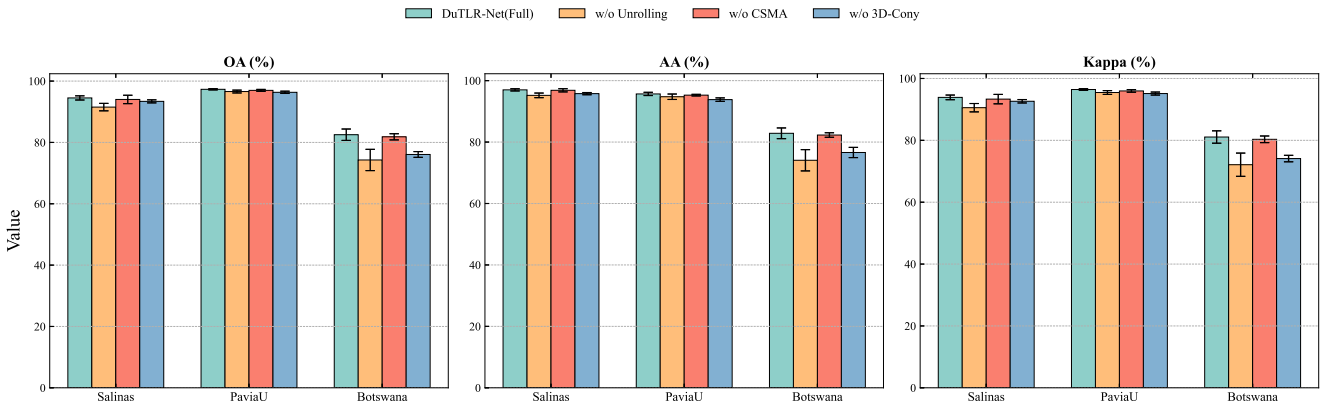


Figure 8. Ablation study results on 3 datasets

local information, with the two streams adaptively integrated through gated residual fusion. Experiments on multiple standard datasets demonstrate that DuTLR-Net significantly outperforms current mainstream methods under extremely low labeled sample ratios, achieving superior classification accuracy. The generated classification maps exhibit clear boundaries and good spatial homogeneity, fully validating the model's effectiveness. This work provides a new paradigm for HSI classification that combines interpretability, efficiency, and high performance. Its design philosophy of "physics-guided, data-optimized" also offers important inspiration for other small-sample remote sensing tasks.

7 Declarations

- Funding: This research received no external funding.
- Competing Interests: The authors declare no competing interests.
- Ethics Approval and Consent to Participate: Not applicable.
- Consent for Publication: Not applicable.
- Data Availability:The publicly available hyperspectral datasets used in this study, namely the Pavia University, Salinas, and Botswana datasets, were obtained from the Hyperspectral Remote Sensing Scenes repository maintained by the University of the Basque Country (available at: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes). All data preprocessing was performed in accordance with standard protocols. The processed intermediate data and model outputs supporting the findings of this study are available from the corresponding author upon reasonable request.
- Materials Availability: Not applicable.
- Code Availability: The code is available from the corresponding author upon reasonable request.
- Author Contributions: Z.R. conceptualized the research and supervised the overall study. C.Y.Q. developed the core algorithm, conducted the main experiments, and served as the corresponding author. L.Q. supervised the experimental process, reviewed the initial draft, and proofread the final manuscript. W.X.C. performed the literature review and assisted with manuscript refinement and structural consistency. L.X.W. prepared Figures 1–3 and revised the manuscript for language clarity. All authors reviewed and approved the final version of the manuscript.

References

1. Gordon Hughes, *On the mean accuracy of statistical pattern recognizers*, IEEE Transactions on Information Theory, vol. 14, no. 1, pp. 55–63, 1968.
2. José M. Bioucas-Dias and Antonio Plaza and others, *Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 2, pp. 354–379, 2012.

3. Xiao Xiang Zhu and Devis Tuia and others, *Deep learning in remote sensing: A comprehensive review and list of resources*, IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8–36, 2017.
4. Farid Melgani and Lorenzo Bruzzone, *Classification of hyperspectral remote sensing images with support vector machines*, IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 8, pp. 1778–1790, 2004.
5. Gustavo Camps-Valls and Lorenzo Bruzzone, *Kernel-based framework for hyperspectral image classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, pp. 1351–1362, 2005.
6. Xiaofei He and Deng Cai and Shuicheng Yan and Hong-Jiang Zhang, *Neighborhood preserving embedding*, Proceedings of the IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 1208–1213, 2005.
7. Yi Chen and Nasser M. Nasrabadi and Trac D. Tran, *Hyperspectral image classification using dictionary-based sparse representation*, IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 10, pp. 3973–3985, 2011.
8. Yushi Chen and Zhouhan Lin and Xing Zhao and Gang Wang and Yanfeng Gu, *Deep learning-based classification of hyperspectral data*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS), vol. 7, no. 6, pp. 2094–2107, 2014.
9. K. Makantasis and others, *Deep supervised learning for hyperspectral data classification through convolutional neural networks*, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4959–4962, 2015.
10. Ying Li and Haokui Zhang and Qiang Shen, *Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network*, Remote Sensing, vol. 9, no. 1, pp. 67, 2017.
11. Zilong Zhong and Jonathan Li and Zhiming Luo and Michael Chapman, *Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework*, IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 2, pp. 847–858, 2018.
12. Danfeng Hong and Lianru Gao and Jing Yao and others, *Graph convolutional networks for hyperspectral image classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 7, pp. 5966–5978, 2020.
13. Danfeng Hong and Zhu Han and Jing Yao and others, *SpectralFormer: Rethinking hyperspectral image classification with transformers*, IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–15, 2021.
14. Vishal Monga and Yuelong Li and Yonina C. Eldar, *Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing*, IEEE Signal Processing Magazine, vol. 38, no. 2, pp. 18–44, 2021.
15. Y. Zhang and Y. Wang and B. Zhang and Q. Li, *Tensor-based robust principal component analysis with locality preserving graph and frontal slice sparsity for hyperspectral image classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–15, 2022.
16. X. Gong and others, *WaveFormer: Spectral-spatial wavelet transformer for hyperspectral image classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–16, 2024.
17. X. Gong and others, *MultiScale spectral spatial convolutional transformer for hyperspectral image classification*, IET Image Processing, vol. 18, no. 6, pp. 4338–4350, 2024.
18. K. Li and Y. Qin and Q. Ling and others, *Self-supervised deep subspace clustering for hyperspectral images*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS), vol. 14, pp. 3215–3227, 2021.
19. Qingshan Wang and Shutao Li and Leyuan Fang, *Extreme Small Sample Hyperspectral Image Classification via Physics-Aware Representation Learning*, Remote Sensing of Environment, vol. 301, pp. 114002, 2026.
20. Hao Chen and Lei Wang and Yanning Zhang, *Rethinking the Bottlenecks of Pure Data-Driven Transformers in Remote Sensing Imagery*, IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1–14, 2025.
21. Wei Liu and Liangpei Zhang and Bo Du, *Physical-Guided Deep Learning: A New Paradigm for Earth Observation Data Processing*, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 215, pp. 102–118, 2026.
22. Yu Zhao and Lianru Gao and Wenzhi Liao, *Few-Shot Learning in Hyperspectral Imaging: A Comprehensive Review and Future Directions*, IEEE Geoscience and Remote Sensing Magazine, vol. 13, no. 1, pp. 22–45, 2025.
23. Zebin Wu and Danfeng Hong and Jocelyn Chanussot, *Overcoming Representation Collapse in High-Dimensional Remote Sensing Data*, IEEE Transactions on Neural Networks and Learning Systems, vol. 37, no. 3, pp. 1504–1519, 2026.
24. Jing Wang and Antonio Plaza and José Bioucas-Dias, *Towards Interpretable Hyperspectral Image Analysis: Blending Physical Mechanisms with Neural Networks*, IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1–16, 2025.

25. Meng Zhang and Mingyi He and Xin Yuan, *Tensor-Based Deep Unrolling Networks for Spatial-Spectral Consistency Preservation in HSIs*, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 211, pp. 45–60, 2025.

26. Yansheng Li and Yongshan Zhang and Xuelong Li, *Spatial-Spectral Tensor Modeling for Complex Heterogeneous Noise Removal in Hyperspectral Imagery*, IEEE Transactions on Image Processing, vol. 35, pp. 2100–2115, 2026.

27. Xian Sun and Peijin Wang and Wenhui Diao, *Interpretable Algorithm Unrolling for Robust Remote Sensing Data Analysis*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS), vol. 18, pp. 5320–5335, 2025.

28. Shangquan Zhou and Weisheng Dong and Guangming Shi, *Adaptive Deep Unfolding with Learnable Proximal Operators for Image Restoration*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 48, no. 4, pp. 2800–2815, 2026.

29. Xin Yang and Qian Li and Bing Zhang, *Physics-Guided Representation Learning for Hyperspectral Image Classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 64, pp. 1–16, 2026.

30. Ting Huang and Jian Yang and Xiaofei Yang, *Deep Tensor Prior Networks for Few-Shot Hyperspectral Classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1–15, 2025.

31. Peng Qiao and Zhipeng Dou and Heng-Chao Li, *Tensor Low-Rank Unrolling for High-Dimensional Earth Observation Data*, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 217, pp. 88–105, 2026.

32. Shufang Xu and Qi Fu and Min Zhu and Xinyu Sun and Ruizhe Liu and Bo Jia and Hongmin Gao, *Dimensionality reduction of medical hyperspectral images based on GPCA-multiscale tensor low-rank graph embedding*, Optics and Laser Technology, vol. 192, pp. 113952, 2025.

33. Zhouhan Lin and Yushi Chen and Xing Zhao, *Learnable Soft-Thresholding for Heterogeneous Noise Suppression in HSIs*, IEEE Transactions on Image Processing, vol. 34, pp. 3102–3116, 2025.

34. Yuming Chen and Xiaorui Wang and Shutao Li, *Cross-Scale Manifold Alignment for Deep Feature Fusion in Remote Sensing*, IEEE Transactions on Geoscience and Remote Sensing, vol. 64, pp. 1–14, 2026.

35. Jiajun Chang and Jiguang Dai and Tengda Zhang, *RADiffSR: A Diffusion Model for Remote Sensing Image Super-Resolution Fusing Residual Attention and Cross-Scale Dynamic Gating*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 19, pp. 1–15, 2026.

36. Senanur Armaġan and Esra Gündoğan and Mehmet Kaya, *Classification of Skin Lesions using Squeeze and Excitation Attention based Hybrid Model of DenseNet and EfficientNet*, Proceedings of the 2024 International Conference on Decision Aid Sciences and Applications (DASA), pp. 1–6, 2024.

37. Xuanyue Shentu and Xiaoping Lai and Tianlei Wang and Jiuwen Cao, *Efficient ADMM-Based Algorithm for Regularized Minimax Approximation*, IEEE Signal Processing Letters, vol. 30, pp. 210–214, 2023.

38. Ran Wang and Hua Zuo and Jie Lu, *Fuzzy Rule-Based Test-Time Adaptation for Class Imbalance in Dynamic Scenarios*, IEEE Transactions on Fuzzy Systems, vol. 33, no. 8, pp. 2469–2480, 2025.

39. Mercedes E. Paoletti and Juan Mario Haut and Ruben Fernandez-Beltran and Javier Plaza and Antonio J. Plaza and Filiberto Pla, *Deep Pyramidal Residual Networks for Spectral–Spatial Hyperspectral Image Classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 2, pp. 740-754, 2019.

40. Zilong Zhong and Jonathan Li and Zhiming Luo and Michael Chapman, *Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework*, IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 2, pp. 847-858, 2018.

41. Swalpa Kumar Roy and Suvojit Manna and Tiecheng Song and Lorenzo Bruzzone, *Attention-Based Adaptive Spectral–Spatial Kernel ResNet for Hyperspectral Image Classification*, IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 9, pp. 7831-7843, 2021.

42. Rui Li and Shunyi Zheng and Chenxi Duan and Yang Yang and Xiqi Wang, *Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network*, Remote Sensing, vol. 12, no. 3, art. no. 582, 2020.

43. Wenping Ma and Qifan Yang and Yue Wu and Wei Zhao and Xiangrong Zhang, *Double-Branch Multi-Attention Mechanism Network for Hyperspectral Image Classification*, Remote Sensing, vol. 11, no. 11, art. no. 1307, 2019.

44. Wenju Wang and Shuguang Dou and Zhongmin Jiang and Liujie Sun, *A Fast Dense Spectral–Spatial Convolution Network Framework for Hyperspectral Images Classification*, Remote Sensing, vol. 10, no. 7, art. no. 1068, 2018.

45. Shuang Liu and Chuan Fu and Yule Duan and Xiaopan Wang and Fulin Luo, *Spatial–Spectral Enhancement and Fusion Network for Hyperspectral Image Classification With Few Labeled Samples*, IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1-14, 2025.

46. Zhen Zhang and Lehao Huang and Bo-Hui Tang and Qingwang Wang and Zhongxi Ge and Linhuan Jiang, *Non-Euclidean Spectral-Spatial feature mining network with Gated GCN-CNN for hyperspectral image classification*, Expert Systems with Applications, vol. , no. , art. no. 126811, 2025.