QuHIC: Quantization-Consistent Multi-Prior Unrolled Reconstruction for Hyperspectral Image Compression

Anonymous authors
Paper under double-blind review

Abstract

Recent advances have demonstrated the great potential of lossy hyperspectral image (HSI) compression for airborne remote-sensing applications. However, given the limited storage capacity and transmission bandwidth of airborne sensing platforms, most existing methods insufficiently integrate compression-induced constraints with spectral–spatial priors, making it difficult to balance low bit rates and high-fidelity reconstruction. To address this challenge, we propose QuHIC, which synergizes bit-level quantized compression with multi-stage unrolled reconstruction, establishing a new paradigm for HSI compression and reconstruction. Specifically, during encoding, QuHIC employs a learnable lightweight encoder to compress an HSI into a low-dimensional quantized representation for efficient storage and transmission. During decoding, HSI reconstruction is formulated as an iterative optimization problem jointly constrained by quantization consistency and multi-prior regularization. Each iteration alternates between a quantization-consistency module (QCM) that explicitly corrects deviations from the feasible quantization interval and a multi-prior refinement network (MPRN) that serves as a learnable proximal operator integrating raw-band spectral continuity and curvature modeling, multi-scale spatial detail enhancement, and gated spectral–spatial contextual interaction. Extensive experiments demonstrate that the proposed QuHIC not only achieves efficient HSI compression at competitive bit rates but also outperforms many state-of-the-art methods in both reconstruction fidelity and spectral detail preservation.

1 Introduction

Hyperspectral images (HSIs) comprise hundreds of contiguous spectral bands and capture rich spectral information together with detailed spatial structures, enabling fine-grained land-cover characterization and material identification. Because of these capabilities, HSIs have demonstrated vast potential in remote sensing interpretation (Hong et al., 2021; Yu et al., 2024), environmental monitoring (Hong et al., 2024; Dong et al., 2025), precision agriculture (Zhou et al., 2025; Li et al., 2025), target and anomaly detection (Lian et al., 2025; Wang et al., 2025). However, the high dimensionality of HSIs inevitably gives rise to massive data volumes, imposing substantial burdens on storage and transmission, particularly for resource-constrained airborne sensing platforms. Meanwhile, low-bit-rate compression often compromises fine spatial structures and spectral correlations, making the reconstruction of high-fidelity HSIs from compact representations considerably challenging.

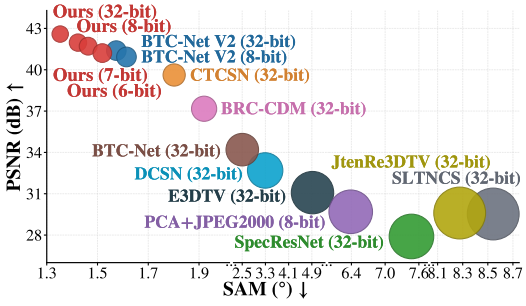

Figure 1: Comparison of HSI reconstruction quality and spectral fidelity, with bubble size representing reconstruction error (RMSE).

To address this challenge, HSI compression methods have evolved from traditional transform coding that exploits transform-domain decorrelation (Du et al., 2014) and model-driven compressed sensing

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

built upon handcrafted sparsity and low-rank priors (Yang et al., 2015; Peng et al., 2020) to data-driven deep learning methods, including deep compressed sensing that jointly learns compressive sampling and nonlinear reconstruction (Hsu et al., 2021; Zhang et al., 2024) and end-to-end learned compression that jointly optimizes quantized representations and neural reconstruction (Zhou et al., 2024; 2026). However, traditional methods rely heavily on handcrafted transforms or priors, whereas most existing learning-based methods adopt one-shot feed-forward decoding, leaving compression-induced constraints insufficiently exploited during reconstruction. This limits progressive reconstruction refinement and the recovery of fine spectral-spatial details at low bit rates.

Bearing the above concerns in mind, we propose QuHIC, a novel deep unrolling framework for efficient HSI compression and high-fidelity reconstruction that synergizes lightweight quantized encoding with multi-stage iterative reconstruction. Specifically, QuHIC first employs a learnable single-layer encoder to compress the input HSI into a low-dimensional quantized representation. It then unrolls an iterative optimization process jointly constrained by quantization consistency and multiple priors into a sequence of learnable reconstruction stages, enabling progressive reconstruction under explicit quantization guidance and thereby striking a favorable balance between low bit rates and high reconstruction fidelity, as illustrated in Figure 1. At the core of QuHIC are two complementary modules embedded in each unrolling stage: the quantization-consistency module (QCM) and the multi-prior refinement network (MPRN). QCM explicitly enforces quantization consistency through feasible-interval projection and representation-residual back-projection. MPRN serves as a learnable proximal operator that employs the raw-band spectral prior to model spectral continuity and curvature, the multi-scale spatial detail prior to restore edges and textures, and the asymmetric spectral-spatial context prior to capture long-range dependencies through gated interactions at low resolution. To the best of our knowledge, this is the first work to introduce deep unrolling into bit-level learned HSI compression. The main contributions of this paper are summarized as follows:

(1) We propose QuHIC, a novel deep unrolling framework for bit-level learned HSI compression that integrates lightweight quantized encoding with quantization-guided iterative refinement for efficient compression and high-fidelity reconstruction.

(2) We develop QCM to promote quantization consistency through feasible-interval projection and representation-residual back-projection, which progressively correct intermediate reconstruction estimates.

(3) We design MPRN as a learnable proximal operator that jointly models raw-band spectral continuity and curvature, multi-scale spatial details, and long-range dependencies through gated spectral-spatial interaction at low resolution.

(4) Extensive experiments demonstrate that QuHIC consistently outperforms a range of state-of-the-art methods at low bit rates, achieving superior reconstruction quality and spectral fidelity.

2 Related Work

2.1 HSI Compression

HSI compression is inherently a challenging rate-distortion optimization problem due to the high dimensionality and strong spectral-spatial redundancy of hyperspectral data. Among early methods, PCA+JPEG2000 (Du et al., 2014) combines PCA-based spectral dimensionality reduction with JPEG2000 coding. SLTNCS (Yang et al., 2015) integrates sparse tensor representation with nonlinear compressed sensing (CS). JTenRe3DTV (Wang et al., 2017) combines Tucker decomposition with weighted three-dimensional total variation (3DTV) to model global low-rank correlations and local spectral-spatial smoothness, and E3DTV (Peng et al., 2020) enhances 3DTV regularization by exploiting correlated sparsity in spectral-spatial gradient maps. More recently, two main paradigms have emerged in deep learning-based HSI compression: deep CS and learned compression. Within the deep CS paradigm, DCSN (Hsu et al., 2021) jointly learns compressive sampling and nonlinear reconstruction. CTCSN (Zhang et al., 2024) combines non-square convolutions with CNN-Transformer mixture modules to construct a lightweight encoder and reconstruction backbone, followed by spectral and spatial super-resolution. Within the learned compression paradigm, BTC-Net (Zhou et al., 2024) employs a lightweight quantized encoder and an attention-based decoder for bit-level compression. BRC-CDM (Liu et al., 2025) combines biresidual compression

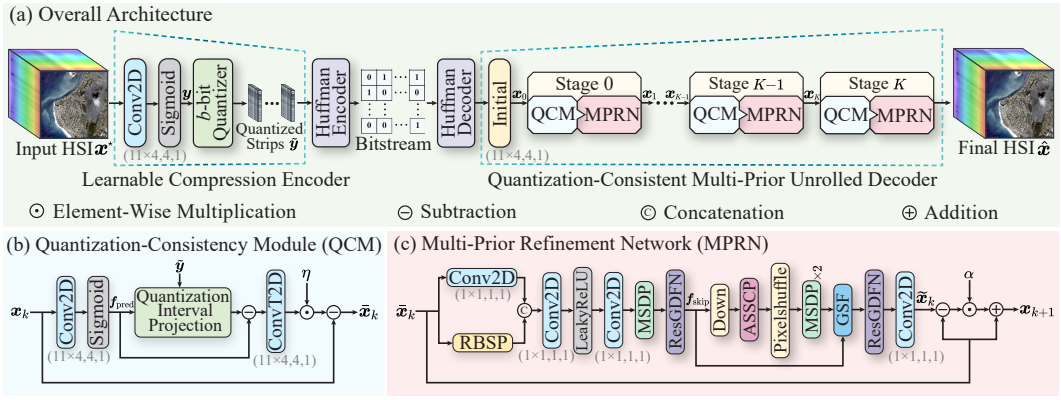

Figure 2: (a) Overview of QuHIC. (b) QCM with re-encoding, quantization-interval projection, and residual back-projection. (c) MPRN integrating RBSP, MSDP, ASSCP, and GSF for progressive spectral-spatial refinement. '$(11 \times 4,\ 4,\ 1)$' indicates a convolution with kernel size 11×4, stride 4, and dilation rate 1.

with conditional diffusion to reconstruct HSIs from compressed spectral-spatial residuals. BTC-Net V2 (Zhou et al., 2026) improves spatial detail recovery using a single-layer quantized encoder and spatial-priority hierarchical reconstruction, and SpecResNet (Saeed et al., 2026) combines hybrid residual learning with spectral calibration to enhance reconstruction efficiency and spectral fidelity. Despite this progress, conventional methods rely on handcrafted priors and iterative solvers, whereas most deep compression models perform one-shot feed-forward reconstruction with limited explicit modeling of compression constraints. In contrast, QuHIC introduces a new paradigm for HSI compression based on quantization-consistent multi-prior unrolling, enabling progressive high-fidelity reconstruction at low bit rates.

3 Method

3.1 Problem Formulation and Unrolling Strategy

Learnable Compression and Quantization Model. Given an input HSI $\boldsymbol{x}^\star \in \mathbb{R}^{C\times H\times W}$, where C is the number of spectral bands, H and W denote the spatial height and width, respectively. A lightweight learnable encoder $\mathcal{F}_\theta(\cdot)$ jointly compresses its spectral and spatial information through convolutional sensing and a Sigmoid activation, yielding a continuous compact representation $\boldsymbol{y} = \mathcal{F}_\theta(\boldsymbol{x}^\star) \in \mathbb{R}^{C'\times H/4\times W/4}$, where C' denotes the number of channels in the compact representation. Subsequently, a b-bit uniform quantizer $\mathcal{Q}_b(\cdot)$ discretizes the continuous representation, producing a quantized representation:

$$\widetilde{\boldsymbol{y}} = \mathcal{Q}_b(\boldsymbol{y}) = \frac{\mathrm{round}\big((2^b-1)\boldsymbol{y}\big)}{2^b-1} \in \mathbb{R}^{C'\times H/4\times W/4}, \tag{1}$$

where $\mathrm{round}(\cdot)$ denotes element-wise rounding to the nearest integer. Since both compression and quantization are many-to-one mappings, recovering $\boldsymbol{x}^\star$ from $\widetilde{\boldsymbol{y}}$ constitutes a non-unique and highly ill-posed inverse problem.

Quantization-Consistent Multi-Prior Unrolling Strategy. Each transmitted quantized value corresponds to a quantization-feasible interval determined by the quantization precision. Therefore, we only require the continuous representation predicted from a candidate reconstruction to lie within the corresponding interval, rather than forcing it to exactly equal $\widetilde{\boldsymbol{y}}$. Let $\Delta = 1/(2^b-1)$ denote the quantization step size. The corresponding element-wise quantization-feasible interval is defined as:

$$\mathcal{C}_b\left(\widetilde{\boldsymbol{y}}\right) = \left[\widetilde{\boldsymbol{y}} - \frac{\Delta}{2}, \widetilde{\boldsymbol{y}} + \frac{\Delta}{2}\right]. \tag{2}$$

For an HSI $\boldsymbol{x}$ to be recovered, the same learnable compression mapping used at the encoder produces the continuous predicted representation $\boldsymbol{f}_{\text{pred}} = \mathcal{F}_\theta(\boldsymbol{x})$. To quantify the deviation of the predicted representation from the quantization-feasible interval, we define the quantization-consistency

fidelity term as:

$$\mathcal{D}\left(\boldsymbol{x}, \widetilde{\boldsymbol{y}}\right) = \frac{1}{2}\left\|\boldsymbol{f}_{\text{pred}} - \Pi_{\mathcal{C}_b(\widetilde{\boldsymbol{y}})}\left(\boldsymbol{f}_{\text{pred}}\right)\right\|_2^2, \tag{3}$$

where $\Pi_{\mathcal{C}_b(\widetilde{\boldsymbol{y}})}(\cdot)$ denotes projection onto the quantization-feasible interval. This projection yields zero resi dual for components within the interval and retains only the errors of out-of-range components. By further introducing a multi-prior regularizer $\mathcal{R}(\boldsymbol{x})$, the final reconstruction objective is formulated as:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{\mathcal{D}\left(\boldsymbol{x}, \widetilde{\boldsymbol{y}}\right) + \lambda \mathcal{R}(\boldsymbol{x})\right\}, \tag{4}$$

where $\mathcal{D}$ enforces consistency between the candidate reconstruction and the transmitted quantized representation, $\mathcal{R}$ characterizes the spectral continuity, multi-scale spatial details, and long-range spectral-spatial dependencies of the HSI, and λ balances the quantization-consistency fidelity term and the multi-prior regularization term. To solve the above optimization problem, we unroll the iterative reconstruction process into K learnable stages, each sequentially performing quantization-consistency correction and multi-prior proximal refinement.

3.2 Overall Architecture

QuHIC consists of a learnable compression encoder, a bitstream representation, and a quantization-consistent multi-prior unrolled decoder, as illustrated in Figure 2(a). On the encoder side, the input HSI $\boldsymbol{x}^\star$ is first transformed into a compact continuous representation $\boldsymbol{y}$ and then quantized according to Eq. 1 to obtain $\widetilde{\boldsymbol{y}}$. For practical storage or transmission, $\widetilde{\boldsymbol{y}}$ is further losslessly entropy-coded into a bitstream using Huffman coding. On the decoder side, Huffman decoding exactly recovers the quantized strip representation $\widetilde{\boldsymbol{y}}$, which is subsequently mapped back to the HSI domain through an independently parameterized transposed convolution to produce an informative initial estimate $\boldsymbol{x}_0 \in \mathbb{R}^{C\times H\times W}$. QuHIC then performs K reconstruction stages, in which QCM explicitly corrects inconsistencies between the current reconstruction and the transmitted quantized representation, and MPRN restores the missing spectral-spatial information. These two modules are applied alternately to progressively improve reconstruction fidelity. During training, the quantizer employs a straight-through estimator with actual quantized values in the forward pass and a unit gradient in the backward pass. The entire framework is jointly optimized using an ℓ_1 reconstruction loss between the final output $\hat{\boldsymbol{x}}$ and the ground-truth HSI $\boldsymbol{x}^\star$.

3.3 Quantization-Consistency Module

QCM explicitly handles the quantization-consistency fidelity term in Eq. 3. At the k-th stage, the current estimate $\boldsymbol{x}_k \in \mathbb{R}^{C\times H\times W}$ is re-encoded into the compressed-representation domain using the parameter-shared encoder $\mathcal{F}_\theta(\cdot)$, yielding the predicted representation $\boldsymbol{f}_{\text{pred}} \in \mathbb{R}^{C'\times H/4\times W/4}$. This weight-sharing strategy ensures that the intermediate reconstruction at each stage is constrained under the same encoding model as the original compression process. To explicitly enforce consistency between the reconstruction and the transmitted quantized representation, we propose quantization-interval projection $\Pi_{\mathcal{C}_b(\widetilde{\boldsymbol{y}})}(\cdot)$, which projects each element of $\boldsymbol{f}_{\text{pred}}$ onto the feasible interval defined in Eq. 2. Components already within this interval remain unchanged, whereas out-of-range components are projected onto the nearest feasible boundary, yielding the projected representation:

$$\Pi_{\mathcal{C}_b(\widetilde{\boldsymbol{y}})}\left(\boldsymbol{f}_{\text{pred}}\right) = \begin{cases} \widetilde{\boldsymbol{y}} - \Delta/2, & \boldsymbol{f}_{\text{pred}} < \widetilde{\boldsymbol{y}} - \Delta/2, \\ \boldsymbol{f}_{\text{pred}}, & \widetilde{\boldsymbol{y}} - \Delta/2 \le \boldsymbol{f}_{\text{pred}} \le \widetilde{\boldsymbol{y}} + \Delta/2, \\ \widetilde{\boldsymbol{y}} + \Delta/2, & \boldsymbol{f}_{\text{pred}} > \widetilde{\boldsymbol{y}} + \Delta/2, \end{cases} \in \mathbb{R}^{C'\times H/4\times W/4}. \tag{5}$$

The difference between the representations before and after projection directly reflects the violation of the quantization-interval constraint. QCM then back-projects this residual into the HSI domain to update the current estimate:

$$\bar{\boldsymbol{x}}_k = \boldsymbol{x}_k - \eta\, \text{ConvT2D}(\boldsymbol{f}_{\text{pred}} - \Pi_{\mathcal{C}_b(\widetilde{\boldsymbol{y}})}\left(\boldsymbol{f}_{\text{pred}}\right)), \tag{6}$$

where η is a learnable stage-dependent step size that controls the strength of the quantization-consistency correction, and $\bar{\boldsymbol{x}}_k \in \mathbb{R}^{C\times H\times W}$ denotes the corrected HSI estimate. As illustrated in Figure 2(b), QCM selectively corrects quantization-inconsistent components, providing an approximate descent step for the fidelity term while preserving components within the feasible interval.

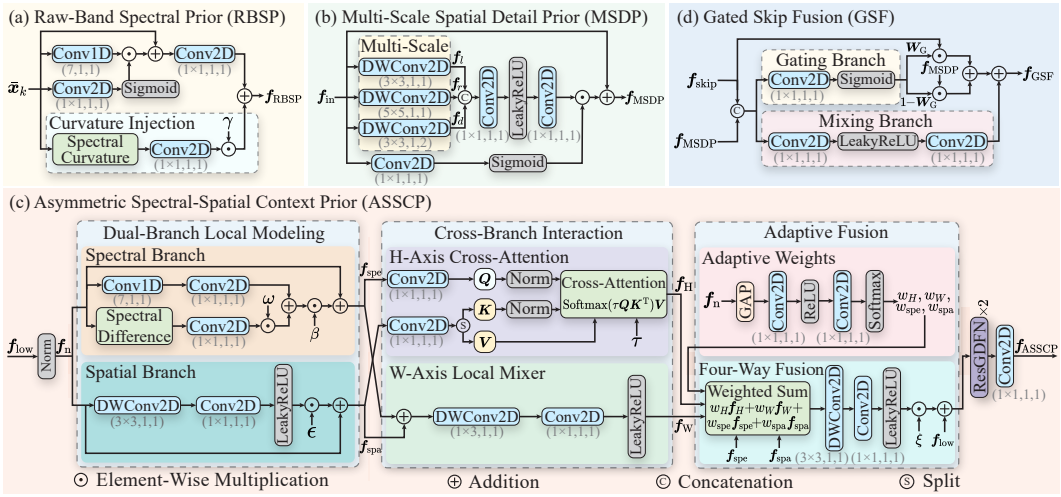

Figure 3: (a) RBSP models raw-band dependencies and spectral curvature. (b) MSDP captures multi-scale spatial details through parallel depth-wise convolutions. (c) ASSCP performs asymmetric spectral-spatial interaction through dual-branch local modeling, cross-branch interaction, and adaptive fusion. (d) GSF adaptively fuses high-resolution skip and spatially refined features.

3.4 Multi-Prior Refinement Network

At the k-th unrolling stage, QCM corrects the current reconstruction using the transmitted quantized representation to obtain the quantization-consistent estimate $\bar{\boldsymbol{x}}_k$, which is subsequently refined by MPRN using spectral-spatial priors to produce the multi-prior estimate $\widetilde{\boldsymbol{x}}_k$. Each stage employs an independently parameterized MPRN to adapt to stage-specific reconstruction states. As illustrated in Figure 2(c), MPRN comprises four core components: the raw-band spectral prior (RBSP), multi-scale spatial detail prior (MSDP), asymmetric spectral–spatial context prior (ASSCP), and gated skip fusion (GSF). Specifically, $\bar{\boldsymbol{x}}_k$ is processed by a dual-branch module consisting of a convolutional branch and an RBSP branch, which respectively extract local shallow features and raw-band features characterizing spectral continuity and curvature. After concatenation and convolutional fusion, MSDP and ResGDFN (Zamir et al., 2022) enhance multi-scale spatial details and generate the full-resolution skip feature. This feature is retained for skip fusion, and its downsampled counterpart is processed by ASSCP to model spectral-spatial context at low resolution. PixelShuffle then restores the full spatial resolution, followed by two cascaded MSDP blocks for further detail refinement. GSF adaptively fuses the restored feature with the skip feature, after which ResGDFN and a convolutional layer generate $\widetilde{\boldsymbol{x}}_k$. Finally, an adaptive residual update controlled by the learnable coefficient α combines $\widetilde{\boldsymbol{x}}_k$ with $\bar{\boldsymbol{x}}_k$ to produce the current stage output $\boldsymbol{x}_{k+1}$:

$$\boldsymbol{x}_{k+1} = \bar{\boldsymbol{x}}_k + \alpha\left(\widetilde{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}_k\right). \tag{7}$$

Raw-Band Spectral Prior. RBSP models inter-band correlations across the original physical spectral bands, as illustrated in Figure 3(a). Given the input $\bar{\boldsymbol{x}}_k$, a one-dimensional convolution first aggregates information from neighboring bands, and an input-dependent gate adaptively modulates the local spectral responses. To further capture higher-order variations in spectral shape, we introduce a curvature injection branch that computes the second-order spectral difference of $\bar{\boldsymbol{x}}_k$ along the physical spectral dimension and projects it into the latent feature space. The projected curvature feature is scaled by a learnable coefficient γ and injected into the local spectral representation. Let C_k denote the number of latent feature channels at the k-th stage. The resulting RBSP feature $\boldsymbol{f}_{\mathrm{RBSP}} \in \mathbb{R}^{C_k \times H \times W}$ is formulated as:

$$\boldsymbol{f}_{\mathrm{RBSP}} = \mathrm{Conv2D}\left[\bar{\boldsymbol{x}}_k + \mathrm{Conv1D}\left(\bar{\boldsymbol{x}}_k\right) \odot \sigma\left(\mathrm{Conv2D}(\bar{\boldsymbol{x}}_k)\right)\right] + \gamma\,\mathrm{Conv2D}\left[\mathcal{D}^2\left(\bar{\boldsymbol{x}}_k\right)\right], \tag{8}$$

where $\sigma(\cdot)$ denotes the Sigmoid activation function, and $\mathcal{D}^2(\cdot)$ denotes the second-order finite difference along the physical spectral dimension. For an interior band $c = \{2, \ldots, C-1\}$, it is defined as $\mathcal{D}^2(\bar{\boldsymbol{x}}_k)_c = \bar{\boldsymbol{x}}_{k,c-1} - 2\bar{\boldsymbol{x}}_{k,c} + \bar{\boldsymbol{x}}_{k,c+1}$, with one-sided differences applied at the spectral boundaries. The local branch enhances spectral continuity through gated neighboring-band aggregation, and the curvature injection branch supplements spectral curvature information that may be weakened during compression.

Multi-Scale Spatial Detail Prior. MSDP recovers spatial details at the original feature resolution through complementary multi-scale modeling, as illustrated in Figure 3(b). Given the input $\boldsymbol{f}_{\text{in}} \in \mathbb{R}^{C_k \times H \times W}$, three parallel depth-wise convolution branches extract spatial information over different receptive fields. Specifically, a 3×3 depth-wise convolution extracts local structures to produce $\boldsymbol{f}_l \in \mathbb{R}^{C_k \times H \times W}$, a 5×5 depth-wise convolution captures regional structures to obtain $\boldsymbol{f}_r \in \mathbb{R}^{C_k \times H \times W}$, and a dilated 3×3 depth-wise convolution with dilation rate 2 expands the sampling neighborhood to generate $\boldsymbol{f}_d \in \mathbb{R}^{C_k \times H \times W}$. The three features are concatenated along the channel dimension and fused by two point-wise convolutions, which project them back to C_k channels. Meanwhile, an input-dependent gate is generated from $\boldsymbol{f}_{\text{in}}$ to adaptively regulate the injection of multi-scale details into the residual path. The MSDP output $\boldsymbol{f}_{\text{MSDP}}$ is formulated as:

$$\boldsymbol{f}_{\text{MSDP}} = \boldsymbol{f}_{\text{in}} + \sigma\left(\text{Conv2D}\left(\boldsymbol{f}_{\text{in}}\right)\right) \odot \text{Conv2D}\left[\phi\left(\text{Conv2D}\left(\text{Concat}\left[\boldsymbol{f}_{\text{l}}, \boldsymbol{f}_{\text{r}}, \boldsymbol{f}_{\text{d}}\right]\right)\right)\right], \tag{9}$$

where $\phi(\cdot)$ denotes the LeakyReLU activation function.

Asymmetric Spectral–Spatial Context Prior. ASSCP performs complementary spectral-spatial contextual interaction at the low-resolution bottleneck, as illustrated in Figure 3(c). At the k-th stage, the input bottleneck feature $\boldsymbol{f}_{\text{low}} \in \mathbb{R}^{2C_k \times H/2 \times W/2}$ is first normalized to obtain $\boldsymbol{f}_n \in \mathbb{R}^{2C_k \times H/2 \times W/2}$, and then processed by parallel spectral and spatial branches to extract complementary representations. The spectral branch models channel correlations in the low-resolution latent space, where a one-dimensional convolution captures dependencies among neighboring latent channels and a second-order difference supplements local variation information, yielding the spectral representation $\boldsymbol{f}_{\text{spe}} \in \mathbb{R}^{2C_k \times H/2 \times W/2}$:

$$\boldsymbol{f}_{\text{spe}} = \boldsymbol{f}_n + \beta\left[\text{Conv2D}\left(\text{Conv1D}(\boldsymbol{f}_n)\right) + \omega\,\text{Conv2D}(\mathcal{D}^2(\boldsymbol{f}_n))\right], \tag{10}$$

where β controls the residual enhancement strength, and ω balances the contribution of the second-order variation. The spatial branch sequentially applies a depth-wise convolution and a point-wise projection, followed by residual enhancement, to generate the spatial representation $\boldsymbol{f}_{\text{spa}} \in \mathbb{R}^{2C_k \times H/2 \times W/2}$. The two representations then interact along different spatial axes. For H-axis cross-attention, $\boldsymbol{f}_{\text{spe}}$ is projected into the query $\boldsymbol{Q}$, and $\boldsymbol{f}_{\text{spa}}$ provides the key $\boldsymbol{K}$ and value $\boldsymbol{V}$. After normalizing $\boldsymbol{Q}$ and $\boldsymbol{K}$, the cross-branch interaction along the height axis is formulated as:

$$\boldsymbol{f}_H = \text{Softmax}\left(\tau \boldsymbol{Q}\boldsymbol{K}^\top\right)\boldsymbol{V} \in \mathbb{R}^{2C_k \times H/2 \times W/2}, \tag{11}$$

where τ denotes a learnable head-wise temperature parameter. In parallel, the W-axis interaction employs a lightweight local mixer. Specifically, $\boldsymbol{f}_{\text{spe}}$ and $\boldsymbol{f}_{\text{spa}}$ are added and then processed by a depth-wise convolution and a point-wise projection to obtain $\boldsymbol{f}_W \in \mathbb{R}^{2C_k \times H/2 \times W/2}$. To adaptively fuse the different branches, the normalized feature $\boldsymbol{f}_n$ is passed through global average pooling and two point-wise convolutions, followed by a Softmax function to generate the weights w_H, w_W, w_{spe}, and w_{spa}. These weights are used to aggregate the four complementary features $\boldsymbol{f}_H$, $\boldsymbol{f}_W$, $\boldsymbol{f}_{\text{spe}}$, and $\boldsymbol{f}_{\text{spa}}$: $w_H\boldsymbol{f}_H + w_W\boldsymbol{f}_W + w_{\text{spe}}\boldsymbol{f}_{\text{spe}} + w_{\text{spa}}\boldsymbol{f}_{\text{spa}}$. The fused feature is further refined through convolutional projection and injected into $\boldsymbol{f}_{\text{low}}$ under the control of a learnable residual scale ξ. The resulting feature is then processed by ResGDFN to obtain the final ASSCP representation $\boldsymbol{f}_{\text{ASSCP}} \in \mathbb{R}^{2C_k \times H/2 \times W/2}$. Through this design, ASSCP efficiently models latent spectral dependencies, local spatial structures, dual-axis cross-branch interactions, and content-adaptive fusion at low resolution.

Gated Skip Fusion. GSF adaptively fuses the high-resolution skip feature with the spatially refined feature, as illustrated in Figure 3(d). Specifically, the skip feature $\boldsymbol{f}_{\text{skip}} \in \mathbb{R}^{C_k \times H \times W}$ is taken from the output of the first ResGDFN in MPRN to preserve high-resolution structural information, and the final MSDP produces the complementary feature $\boldsymbol{f}_{\text{MSDP}} \in \mathbb{R}^{C_k \times H \times W}$. The gating branch predicts an adaptive gating map $W_G \in \mathbb{R}^{C_k \times H \times W}$ through a convolution to dynamically balance the two features. Meanwhile, the mixing branch employs two point-wise convolutions to extract interactions from their concatenation. The gated interpolation preserves reliable information from both inputs, and the mixing branch captures cross-feature interactions, yielding $\boldsymbol{f}_{\text{GSF}}$:

$$\boldsymbol{f}_{\text{GSF}} = W_G \odot \boldsymbol{f}_{\text{skip}} + (1 - W_G) \odot \boldsymbol{f}_{\text{MSDP}} + \text{Conv2D}\left[\phi\left(\text{Conv2D}\left(\text{Concat}(\boldsymbol{f}_{\text{skip}}, \boldsymbol{f}_{\text{MSDP}})\right)\right)\right]. \tag{12}$$

4 Experimental Results

Datasets. For HSI compression, we follow the dataset setting of BTC-Net V2 (Zhou et al., 2026) and adopt its preprocessed AVIRIS (Green et al., 1998) and WHU-Hi (Zhong et al., 2020) datasets.

Table 1: Quantitative comparison of different HSI compression methods on the AVIRIS dataset under a 1% feature compression ratio. Each scene cell reports PSNR (dB)↑, RMSE↓, and SAM (°)↓ from top to bottom. bpppb denotes bits per pixel per band. Params (M) and FLOPs (G) measure model complexity, and the best overall results are highlighted in **bold**.

Algorithms	bpppb	Params	FLOPs	HyspIRI17(F)	HyspIRI40(M)	California(U)	FtMyers(L)	Moffett	Average
DCSN (32-bit) (Hsu et al., 2021)	0.314	11.95	110.66	34.09 84.09 2.162	35.02 40.25 1.328	31.05 123.81 3.466	30.49 194.00 5.228	32.80 184.59 4.270	32.69 125.35 3.291
BTC-Net (32-bit) (Zhou et al., 2024)	0.314	12.40	156.71	35.43 69.95 1.820	36.32 33.68 0.986	32.63 105.95 2.851	32.85 152.60 2.831	33.76 165.33 4.023	34.20 105.50 2.502
CTCSN (32-bit) (Zhang et al., 2024)	0.314	11.61	121.22	39.93 40.62 1.297	42.67 15.13 0.742	38.08 53.98 1.995	38.46 74.71 1.926	38.97 94.57 3.051	39.62 55.80 1.802
BRC-CDM (32-bit) (Liu et al., 2025)	0.314	23.83	778.94	38.95 41.67 1.347	38.59 22.55 0.842	35.03 71.27 2.324	36.02 92.47 2.296	37.27 105.18 2.789	37.17 66.63 1.920
SpecResNet (32-bit) (Saeed et al., 2026)	0.314	11.81	265.00	30.42 122.50 3.742	30.65 65.50 2.609	26.71 197.59 6.544	23.34 374.58 15.946	28.39 280.59 8.497	27.90 208.15 7.468
BTC-Net V2 (32-bit) (Zhou et al., 2026)	0.314	12.55	565.03	43.17 27.20 0.992	44.67 12.68 0.644	39.34 47.90 1.774	40.06 65.71 1.592	39.73 88.33 2.866	41.39 48.36 1.574
BTC-Net V2 (8-bit) (Zhou et al., 2026)	0.078	12.55	565.03	42.70 28.13 1.031	43.67 13.61 0.691	38.98 48.15 1.806	39.81 65.16 1.669	39.51 88.25 2.876	40.93 48.66 1.614
Ours (32-bit)	0.314	11.58	876.53	**44.61 23.01 0.849**	**45.83 10.58 0.539**	**40.60 41.79 1.527**	**41.12 58.51 1.318**	**40.76** 80.51 **2.533**	**42.59 42.88 1.353**
Ours (8-bit)	0.078	11.58	876.53	43.85 25.10 0.913	44.48 12.29 0.613	39.98 43.53 1.611	40.80 58.93 1.430	40.74 **80.06** 2.546	41.97 43.98 1.423
Ours (7-bit)	0.069	11.58	876.53	43.51 25.80 0.951	44.16 12.63 0.622	39.85 44.10 1.631	40.46 60.25 1.500	40.57 80.74 2.611	41.71 44.70 1.463
Ours (6-bit)	0.059	11.58	876.53	43.04 27.02 0.981	43.23 14.07 0.673	39.44 45.79 1.702	40.12 62.87 1.613	40.21 82.52 2.629	41.21 46.45 1.519

The preprocessed AVIRIS dataset is derived from original airborne hyperspectral data comprising 224 bands over approximately 400-2500 nm. After removing water-absorption and low-SNR bands, 172 valid bands are retained. This processed version contains 1,183 HSI samples covering diverse scenes, including California, Ft. Myers, HyspIRI, and Moffett Field. For WHU-Hi, we use the HongHu and LongKou high-resolution agricultural scenes, each comprising 270 bands over 400-1000 nm. For both datasets, we adopt the training and test splits predefined by BTC-Net V2.

Implementation Details. We adopt the Adam optimizer with a learning rate of 1×10^{-5}. The unrolled reconstruction network comprises nine stages for progressive refinement. Training is conducted for 5,000 epochs on a single NVIDIA A100 GPU with 40 GB of memory.

4.1 Quantitative Results

We compare QuHIC with six representative state-of-the-art HSI compression methods. Table 1 summarizes the PSNR, RMSE, and SAM results on five AVIRIS test scenes under different quantization precisions. Orange and green denote end-to-end and deep unrolling networks, respectively. Under the same 32-bit setting at 0.314 bpppb, QuHIC achieves higher PSNR and lower RMSE and SAM across all scenes. Under the same 8-bit setting at 0.078 bpppb, QuHIC also outperforms BTC-Net V2 across all scenes and metrics. Even at 7 and 6 bits, QuHIC achieves better average results across all metrics than 8-bit BTC-Net V2, demonstrating its advantages in reconstruction quality and spectral fidelity at lower bit rates. Moreover, QuHIC has the lowest parameter count among all compared methods. These results demonstrate that quantization-consistent reconstruction and multi-prior refinement effectively mitigate spectral and spatial information loss caused by compression.

We further evaluate QuHIC on the WHU-Hi dataset and AVIRIS scene groups from geographic regions unseen during training, with all methods evaluated under the 32-bit setting. As shown in Table 2, QuHIC outperforms the competing methods in both evaluations. The WHU-Hi results validate its effectiveness for UAV-borne HSI compression, and the unseen AVIRIS results further confirm its

Table 2: Quantitative comparison of different HSI compression methods on the WHU-Hi and unseen AVIRIS datasets. The unseen AVIRIS results are averaged over five geographically unseen scenes.

Algorithms	WHU-Hi (Zhong et al., 2020)			unseen AVIRIS (Green et al., 1998)		
	PSNR (dB)↑	RMSE↓	SAM (°)↓	PSNR (dB)↑	RMSE↓	SAM (°)↓
DCSN (Hsu et al., 2021)	28.42	0.290	5.770	28.28	113.30	3.523
BTC-Net (Zhou et al., 2024)	26.56	0.360	7.214	30.61	92.51	2.411
CTCSN (Zhang et al., 2024)	30.19	0.258	5.341	36.71	35.37	1.628
BRC-CDM (Liu et al., 2025)	30.55	0.289	5.144	32.80	61.18	2.206
SpecResNet (Saeed et al., 2026)	26.04	0.387	6.808	21.27	221.54	12.019
BTC-Net V2 (Zhou et al., 2026)	26.80	0.322	6.615	40.48	24.12	1.082
Ours	**31.81**	**0.226**	**4.466**	**41.21**	**23.36**	**0.955**

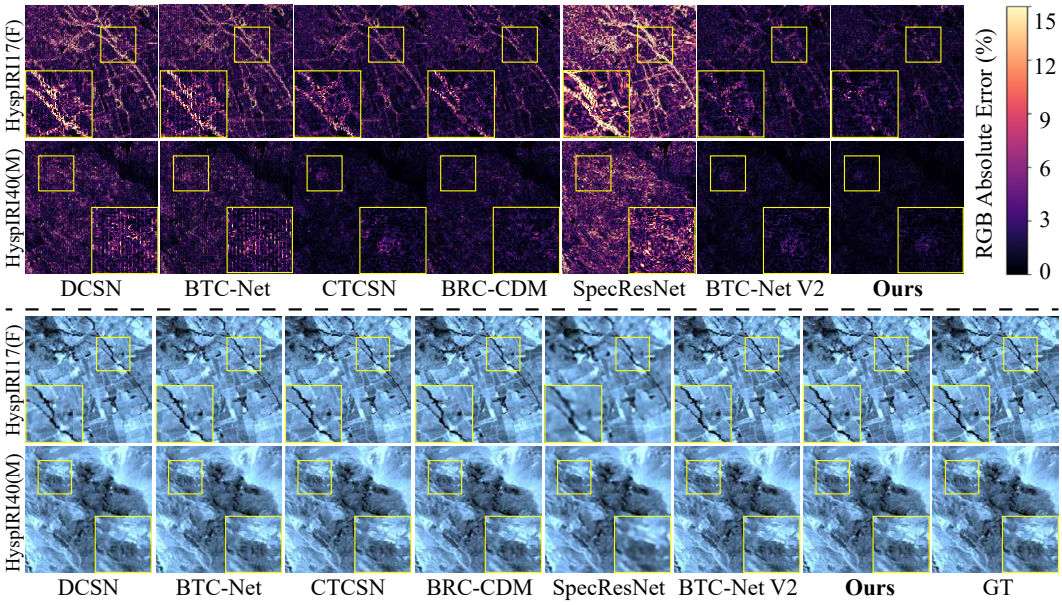

Figure 4: Qualitative comparison of reconstruction results and corresponding error maps on the AVIRIS dataset. The lower part presents the false-color RGB reconstruction results together with the ground truth, and the upper part shows the corresponding RGB absolute error maps of different methods with respect to the ground truth, where brighter colors indicate larger reconstruction errors, as indicated by the error bar. The yellow boxes mark representative regions, with zoomed-in views provided for detailed comparison.

cross-region generalization capability. Additional experimental results and detailed descriptions are provided in the appendix.

4.2 Qualitative Results

As shown in Figure 4, QuHIC produces RGB reconstructions on AVIRIS that most closely match the ground truth, with clearer structural boundaries and richer local textures. Its darker error maps indicate lower reconstruction errors and fewer compression artifacts. As shown in Figure 5, QuHIC also yields RGB reconstructions closer to the ground truth and lower reconstruction errors on WHU-Hi and unseen AVIRIS scenes, validating its effectiveness for UAV-borne HSI compression and its cross-region generalization capability, respectively. Overall, these visual results highlight the robust spatial-detail recovery of QuHIC across different sensing platforms and geographic regions.

4.3 Ablation Studies

As shown in Table 3, components are incrementally added from A0 to A8 to assess their respective contributions. A0 is the basic compression–backprojection baseline without the proposed re-

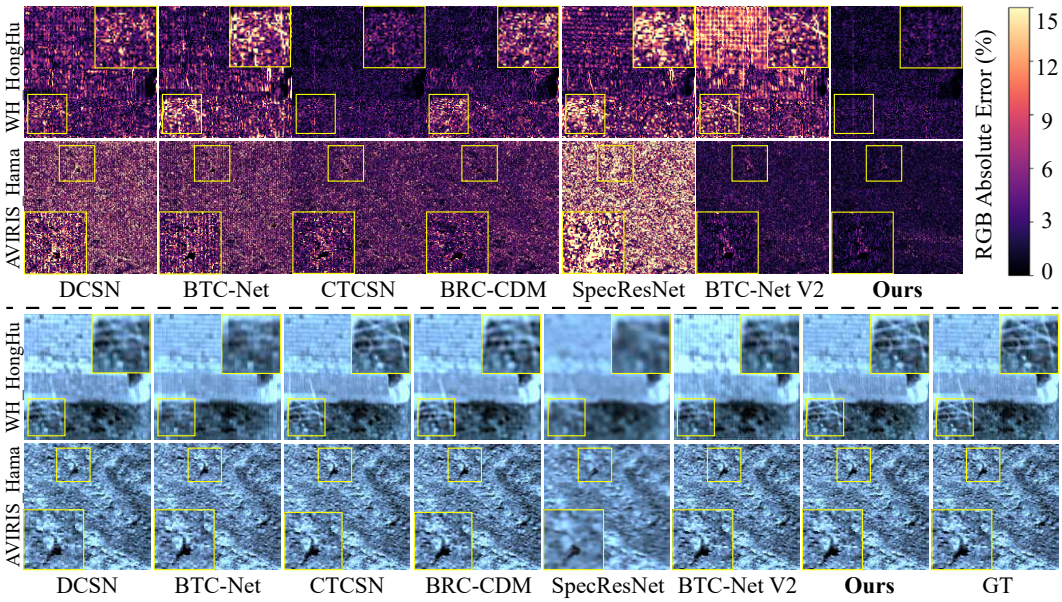

Figure 5: Qualitative comparison of reconstruction results and corresponding error maps on WHU-Hi HongHu and the geographically unseen AVIRIS Hama region.

Table 3: Ablation study of different components under 8-bit quantization on the AVIRIS dataset. Starting from the basic compression-backprojection baseline, key components are progressively introduced to construct the full QuHIC model.

ID	Configuration	PSNR (dB)↑	RMSE↓	SAM (°)↓
A0	Basic Compression-Backprojection Baseline	28.35	180.53	5.163
A1	+ Quantization-Consistency Deep Unrolling	37.02	68.93	2.249
A2	+ RBSP w/o Spectral Curvature	38.01	63.88	1.988
A3	+ RBSP	38.53	59.03	1.926
A4	+ MSDP	40.60	49.15	1.609
A5	+ ASSCP only w/ Dual-Branch Local Modeling	40.88	48.69	1.577
A6	+ ASSCP w/o Adaptive Fusion	41.46	46.59	1.509
A7	+ ASSCP	41.71	45.61	1.478
A8	**+ GSF (Full QuHIC)**	**41.97**	**43.98**	**1.423**

construction modules. A1 adds quantization-consistent deep unrolling for iterative reconstruction under explicit quantization constraints. A2 incorporates RBSP without spectral-curvature modeling, whereas A3 uses the complete RBSP to exploit raw-band spectral continuity and curvature. A4 introduces MSDP to enhance multi-scale spatial details. Within ASSCP, A5 uses only dual-branch local modeling, A6 further adds cross-branch interaction without adaptive fusion, and A7 employs the complete ASSCP. Finally, A8 incorporates GSF to selectively fuse high-resolution skip and decoder-refined features, forming the full QuHIC model. The progressive gains from A0 to A8 confirm the complementary contributions of quantization consistency, spectral and spatial priors, spectral-spatial context modeling, and gated feature fusion.

5 Conclusion

In this paper, we proposed QuHIC, an HSI compression and reconstruction framework that combined lightweight learnable quantized encoding with quantization-consistent multi-prior unrolling. QuHIC employed QCM to correct deviations from feasible quantization intervals and MPRN to integrate raw-band spectral modeling, multi-scale spatial refinement, low-resolution spectral–spatial interaction, and gated feature fusion. Extensive experiments demonstrated superior reconstruction quality and spectral fidelity at low bit rates, together with strong cross-region generalization.

AI USE STATEMENT

In this work, we used generative AI tools for grammatical correction and language refinement of author-written manuscript text. We have not used generative AI tools for retrieval or discovery of related work, research ideation or execution, drafting substantive sections of the paper, generating synthetic datasets, or proving mathematical claims, and all other substantive research tasks are not applicable to this work. Additionally, we used generative AI tools for improving the clarity and readability of author-written text. We have reviewed all AI-assisted work. All AI-assisted edits were reviewed and verified by the authors to ensure technical accuracy and consistency with the implemented methods and experimental results. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work focuses on hyperspectral image compression and reconstruction using publicly available remote-sensing benchmark datasets. It does not involve human or animal subjects, personally identifiable information, private data, or the collection of new sensitive data. The proposed method is intended to improve the efficiency and fidelity of hyperspectral data storage and transmission for scientific remote-sensing applications. Although hyperspectral imaging technologies may potentially be applied in broader sensing or surveillance scenarios, our study does not develop or evaluate applications involving individual identification, behavioral monitoring, or other privacy-sensitive tasks. We follow standard research practices in dataset usage, experimental evaluation, and result reporting, and we are not aware of additional ethical concerns arising directly from the methodology presented in this work.

REPRODUCIBILITY STATEMENT

We have made efforts to ensure the reproducibility of this work by providing a detailed description of the QuHIC framework, including the quantization-consistent unrolling formulation, stage-wise reconstruction process, and the architectures of QCM and MPRN in the main paper. The datasets, evaluation protocols, training settings, and implementation details are described in the experimental section, and additional derivations, architectural configurations, bitrate computation, and dataset details are provided in the appendix. The reported experiments are conducted on publicly available AVIRIS and WHU-Hi datasets following the stated preprocessing and evaluation protocols. To further facilitate reproducibility, the complete source code, training and evaluation scripts, and pretrained model will be made publicly available upon acceptance.

REFERENCES

Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2019.

Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc V Gool. Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. volume 35, pp. 37749–37761, 2022.

Wudi Chen, Zhiyuan Zha, Xin Yuan, Shigang Wang, Bihan Wen, Jiantao Zhou, Gang Yan, Zipei Fan, and Ce Zhu. Phy-cosf: Physics-guided continuous spectral fields reconstruction and spectral super-resolution for snapshot compressive imaging. In *Forty-third International Conference on Machine Learning*, pp. 1–15, 2026.

Wenqian Dong, Junying Ren, Song Xiao, Leyuan Fang, Jiahui Qu, and Yunsong Li. Cycle translation-based collaborative training for hyperspectral-rgb multimodal change detection. *IEEE Transactions on Image Processing*, 34:6347–6360, 2025.

Qian Du, Nam Hoai Ly, and James E. Fowler. An operational approach to PCA+JPEG2000 compression of hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2237–2245, 2014.

Robert O. Green, Michael L. Eastwood, Charles M. Sarture, Thomas G. Chrien, Mikael Aronsson, Bruce J. Chippendale, Jessica A. Faust, Betina E. Pavri, Christopher J. Chovit, Manuel Solis, Martin R. Olah, and Orlesa Williams. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, 65(3):227–248, 1998.

Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Twenty-Seven International Conference on Machine Learning*, pp. 399–406, 2010.

Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2021.

Danfeng Hong, Bing Zhang, Xuyang Li, Yuxuan Li, Chenyu Li, Jing Yao, Naoto Yokoya, Hao Li, Pedram Ghamisi, Xiuping Jia, Antonio Plaza, Paolo Gamba, Jon Atli Benediktsson, and Jocelyn Chanussot. Spectralgpt: Spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5227–5244, 2024.

Chih-Chung Hsu, Chia-Hsiang Lin, Chi-Hung Kao, and Yen-Cheng Lin. DCSN: Deep compressed sensing network for efficient hyperspectral data transmission of miniaturized satellite. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7773–7789, 2021.

Jingtao Li, Yingyi Liu, Xinyu Wang, Yunning Peng, Chen Sun, Shaoyu Wang, Zhendong Sun, Tian Ke, Xiao Jiang, Tangwei Lu, Anran Zhao, and Yanfei Zhong. Hyperfree: A channel-adaptive and tuning-free foundation model for hyperspectral remote sensing imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23048–23058, 2025.

Miaoyu Li, Ying Fu, Ji Liu, and Yulun Zhang. Pixel adaptive deep unfolding transformer for hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12959–12968, 2023.

Jie Lian, Lizhi Wang, He Sun, and Hua Huang. Gt-had: Gated transformer for hyperspectral anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):3631–3645, 2025.

Jiahui Liu, Lili Zhang, Jingang Wang, and Lele Qu. Biresidual compression network with conditional diffusion model for hyperspectral image compression. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025.

Jiawei Ma, Xiao-Yang Liu, Zheng Shou, and Xin Yuan. Deep tensor admm-net for snapshot compressive imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10222–10231, 2019.

Jiangjun Peng, Qi Xie, Qian Zhao, Yao Wang, Yee Leung, and Deyu Meng. Enhanced 3DTV regularization and its applications on HSI denoising and compressed sensing. *IEEE Transactions on Image Processing*, 29:7889–7903, 2020.

Fahad Saeed, Shumin Liu, and Jie Chen. SpecResNet: Hyperspectral image compression via hybrid residual learning and spectral calibration. *Remote Sensing*, 18(7):1074–1091, 2026.

Di Wang, Meiqi Hu, Yao Jin, Yuchun Miao, Jiaqi Yang, Yichu Xu, Xiaolei Qin, Jiaqi Ma, Lingyu Sun, Chenxing Li, Chuan Fu, Hongruixuan Chen, Chengxi Han, Naoto Yokoya, Jing Zhang, Minqiang Xu, Lin Liu, Lefei Zhang, Chen Wu, Bo Du, Dacheng Tao, and Liangpei Zhang. Hypersigma: Hyperspectral intelligence comprehension foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8):6427–6444, 2025.

Yao Wang, Lin Lin, Qian Zhao, Tianwei Yue, Deyu Meng, and Yee Leung. Compressive sensing of hyperspectral images via joint tensor tucker decomposition and weighted total variation regularization. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2457–2461, 2017.

Shuyuan Yang, Min Wang, Peng Li, Li Jin, Bin Wu, and Licheng Jiao. Compressive hyperspectral imaging via sparse tensor and nonlinear compressed sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 53(11):5943–5957, 2015.

Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Deep admm-net for compressive sensing mri. In *Advances in Neural Information Processing Systems*, volume 29, pp. 1–9, 2016.

Linlin Yu, Yifei Lou, and Feng Chen. Uncertainty-aware graph-based hyperspectral image classification. In *International Conference on Learning Representations*, pp. 20503–20534, 2024.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022.

Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, 2018.

Lei Zhang, Longsheng Zhang, Chengpeng Song, and Peng Zhang. Hyperspectral image compression sensing network with cnn–transformer mixture architectures. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

Xuanyu Zhang, Yongbing Zhang, Ruiqin Xiong, Qilin Sun, and Jian Zhang. Herosnet: Hyperspectral explicable reconstruction and optimal sampling deep network for snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17532–17541, 2022.

Yanfei Zhong, Xin Hu, Chang Luo, Xinyu Wang, Ji Zhao, and Liangpei Zhang. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (h^2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sensing of Environment*, 250:112012–112017, 2020.

Fangqin Zhou, Mert Kilickaya, Joaquin Vanschoren, and Ran Piao. Hytas: A hyperspectral image transformer architecture search benchmark and analysis. In *European Conference on Computer Vision*, pp. 236–252, 2025.

Xichuan Zhou, Xuan Zou, Xiangfei Shen, Wenjia Wei, Xia Zhu, and Haijun Liu. BTC-Net: Efficient bit-level tensor data compression network for hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024.

Xichuan Zhou, Mingyang Wan, Hang Yu, Tao Chen, Rulong He, Xiangfei Shen, Lihui Chen, and Haijun Liu. BTC-Net V2: Spatial-priority hierarchical reconstruction for hyperspectral image compression. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–17, 2026.

A Discussion, Limitations and Future Work

The proposed QuHIC establishes a quantization-consistent multi-prior deep unrolling framework for HSI compression. Through the synergistic design of lightweight learnable quantized encoding and multi-stage unrolled reconstruction, QuHIC explicitly incorporates quantization-consistency constraints and spectral-spatial priors into the reconstruction process. Although its asymmetric architecture shifts most of the computational burden to the ground-side decoder, thereby maintaining a lightweight onboard encoder, the repeated QCM and MPRN operations during deep unrolling still incur relatively high decoding FLOPs, which may affect the processing throughput and near-real-time reconstruction of large-scale remote-sensing data. Future work will explore cross-stage parameter sharing, adaptive stage scheduling, and lightweight spectral-spatial interaction modules to reduce model storage and computational overhead while maintaining reconstruction quality. Furthermore, QuHIC can be extended to unified variable-bit compression under dynamic transmission bandwidths and other high-dimensional imaging tasks involving multitemporal or spectral video data.

B Additional Related Work

B.1 Deep Unrolling Network

Deep algorithm unrolling establishes an interpretable connection between iterative optimization and learnable neural architectures. Early works such as LISTA (Gregor & LeCun, 2010) demonstrated that iterative sparse inference can be represented by a fixed-depth learnable network. Subsequent approaches, including ADMM-Net (Yang et al., 2016), ISTA-Net (Zhang & Ghanem, 2018), and MoDL (Aggarwal et al., 2019), further incorporated explicit physical data-consistency steps and learnable regularization modules into iterative inverse-problem solvers. Deep Tensor ADMM-Net (Ma et al., 2019) extended this idea to snapshot compressive imaging by unfolding tensor-based ADMM iterations.

Deep unfolding has also been extensively investigated for hyperspectral reconstruction. Heros-Net (Zhang et al., 2022) unfolds an ISTA-style optimization process for snapshot spectral imaging, DAUHST (Cai et al., 2022) introduces degradation-aware parameter estimation into spectral unfolding, PADUT (Li et al., 2023) employs pixel-adaptive data updates together with a learned spectral prior, and Phy-CoSF (Chen et al., 2026) incorporates physics-guided continuous spectral fields to bridge discrete-wavelength training with arbitrary continuous spectral rendering.

Different from these methods, QuHIC addresses the inverse problem induced by a learned quantized representation rather than a conventional continuous linear measurement. Consequently, the transmitted representation does not specify a single measurement target. Instead, each quantized value defines a feasible interval, and QCM explicitly projects the re-encoded estimate onto this interval before back-projecting only the quantization-violation residual. MPRN subsequently serves as the learned proximal component for spectral-spatial regularization. Therefore, QuHIC extends the standard data-consistency/prior decomposition of deep unfolding to bit-level quantized HSI compression.

C Additional Method

C.1 Problem Formulation and Unrolling Strategy

This section provides additional derivations for the quantization-consistent multi-prior hyperspectral image (HSI) unrolling strategy introduced in Section 3.1. The learnable compression and b-bit quantization processes are defined in Eq. 1, and the corresponding quantization-feasible interval is given in Eq. 2. We therefore focus here on deriving the regularized reconstruction model and its stage-wise unrolling process.

From Quantized Observation to Constrained Reconstruction. For a candidate HSI $\boldsymbol{x}$, the parameter-shared encoder produces its continuous representation $\mathcal{F}_\theta(\boldsymbol{x})$. According to Eq. 2, a reconstruction is quantization-consistent as long as its re-encoded representation lies inside the feasible set associated with the transmitted signal $\widetilde{\boldsymbol{y}}$. Hence, all admissible reconstructions form the

feasible set $\mathcal{S}$:

$$\mathcal{S}\left(\widetilde{\boldsymbol{y}}\right)=\left\{\boldsymbol{x}:\mathcal{F}_{\theta}(\boldsymbol{x})\in\mathcal{C}_{b}\left(\widetilde{\boldsymbol{y}}\right)\right\}. \tag{13}$$

Because the compression mapping substantially reduces both spectral and spatial dimensions, the same quantized representation can generally correspond to multiple candidate HSIs. Quantization consistency alone is therefore insufficient to select a unique reconstruction. Introducing an HSI prior $\mathcal{R}(\boldsymbol{x})$, the reconstruction process can first be formulated as the following constrained optimization problem:

$$\widehat{\boldsymbol{x}}=\arg\min_{\boldsymbol{x}}\mathcal{R}(\boldsymbol{x})\quad\text{s.t.}\quad\mathcal{F}_{\theta}(\boldsymbol{x})\in\mathcal{C}_{b}\left(\widetilde{\boldsymbol{y}}\right). \tag{14}$$

Soft Relaxation of Quantization Consistency. To obtain a continuous data fidelity measure, we introduce an auxiliary feasible representation $\boldsymbol{z}\in\mathcal{C}_{b}(\widetilde{\boldsymbol{y}})$ and relax the equality constraint between the re-encoded representation and $\boldsymbol{z}$ using a quadratic penalty:

$$\min_{\boldsymbol{x},\,\boldsymbol{z}\in\mathcal{C}_{b}(\widetilde{\boldsymbol{y}})}\frac{\mu}{2}\left\|\mathcal{F}_{\theta}(\boldsymbol{x})-\boldsymbol{z}\right\|_{2}^{2}+\lambda\mathcal{R}(\boldsymbol{x}), \tag{15}$$

where $\mu>0$ controls the strength of the consistency penalty. For a given $\boldsymbol{x}$, the optimal auxiliary variable is the nearest point in the feasible set:

$$\boldsymbol{z}^{\star}=\arg\min_{\boldsymbol{z}\in\mathcal{C}_{b}(\widetilde{\boldsymbol{y}})}\left\|\mathcal{F}_{\theta}(\boldsymbol{x})-\boldsymbol{z}\right\|_{2}^{2}=\Pi_{\mathcal{C}_{b}(\widetilde{\boldsymbol{y}})}\left(\mathcal{F}_{\theta}(\boldsymbol{x})\right). \tag{16}$$

Substituting Eq. 16 into Eq. 15 converts the quadratic penalty term into the squared Euclidean distance from the predicted representation to the quantization-feasible set. After absorbing μ into the relative regularization weight and the learnable stage-wise step sizes, this yields exactly the quantization-consistency fidelity term defined in Eq. 3, and consequently the regularized reconstruction objective in Eq. 4.

Gradient Interpretation of QCM. We derive the update formula for the QCM from the fidelity term in Eq. 3. At stage k, let $\boldsymbol{x}_{k}$ denote the current estimate, and define its re-encoded predicted representation as $\boldsymbol{f}_{\text{pred}}=\mathcal{F}_{\theta}(\boldsymbol{x}_{k})$. According to the projection operation defined in Eq. 5, its nearest feasible representation is given by:

$$\boldsymbol{z}_{k}=\Pi_{\mathcal{C}_{b}(\widetilde{\boldsymbol{y}})}\left(\boldsymbol{f}_{\text{pred}}\right),\qquad\boldsymbol{e}_{k}=\boldsymbol{f}_{\text{pred}}-\boldsymbol{z}_{k}, \tag{17}$$

where $\boldsymbol{e}_{k}$ represents the quantization-violation residual. For a closed convex set, the gradient of the squared distance function satisfies:

$$\nabla_{\boldsymbol{f}_{\text{pred}}}\mathcal{D}=\boldsymbol{f}_{\text{pred}}-\Pi_{\mathcal{C}_{b}(\widetilde{\boldsymbol{y}})}\left(\boldsymbol{f}_{\text{pred}}\right)=\boldsymbol{e}_{k}. \tag{18}$$

This result explains why only components falling outside the feasible interval produce QCM corrections: components already satisfying the quantization constraints yield a zero projection residual. Writing the learnable encoder as $\mathcal{F}_{\theta}(\boldsymbol{x})=\sigma(\mathcal{H}_{\theta}(\boldsymbol{x})+\boldsymbol{b}_{\theta})$, applying the chain rule yields the exact gradient with respect to the HSI:

$$\nabla_{\boldsymbol{x}_{k}}\mathcal{D}=\mathcal{H}_{\theta}^{\top}\left[\sigma'\left(\mathcal{H}_{\theta}(\boldsymbol{x}_{k})+\boldsymbol{b}_{\theta}\right)\odot\boldsymbol{e}_{k}\right]. \tag{19}$$

Here, $\mathcal{H}_{\theta}^{\top}$ denotes the adjoint operator of the sensing operator, implemented via transposed convolution sharing the sensing kernel. The practical implementation of QCM adopts the Straight-Through Estimator (STE) for local Sigmoid derivatives and directly back-projects $\boldsymbol{e}_{k}$. Under this approximation, Eq. 19 simplifies to the QCM update defined in Eq. 6.

Proximal Interpretation of MPRN. Following the QCM data-consistency correction, solving the regularization term can be formulated as a classic proximal subproblem. This subproblem encourages the solution to conform to spatial and spectral priors while preventing it from deviating excessively from the QCM output. Since complex HSI priors lack an explicit analytical proximal mapping, QuHIC employs a stage-specific MPRN to implicitly learn and substitute the proximal operator. Each stage features an independently parameterized MPRN, allowing the learned mapping to adapt flexibly to different reconstruction states. Finally, the refined output of the MPRN is incorporated into the current estimate via a stage-adaptive relaxation update. Starting from the learned initialization $\boldsymbol{x}_{0}$, alternating between QCM and MPRN for K iterations yields the final reconstruction result $\widehat{\boldsymbol{x}}$.

Table 4: Test scenes and regions used in the experiments.

Dataset	Scenes / Regions
AVIRIS (Green et al., 1998)	HyspIRI17 (F), HyspIRI40 (M), California (U), Ft. Myers (L), Moffett
unseen AVIRIS (Green et al., 1998)	Amikougami. Fort Myers (U). Hama. Dewar. Campbell River.
WHU-Hi (Zhong et al., 2020)	LongKou. HongHu.

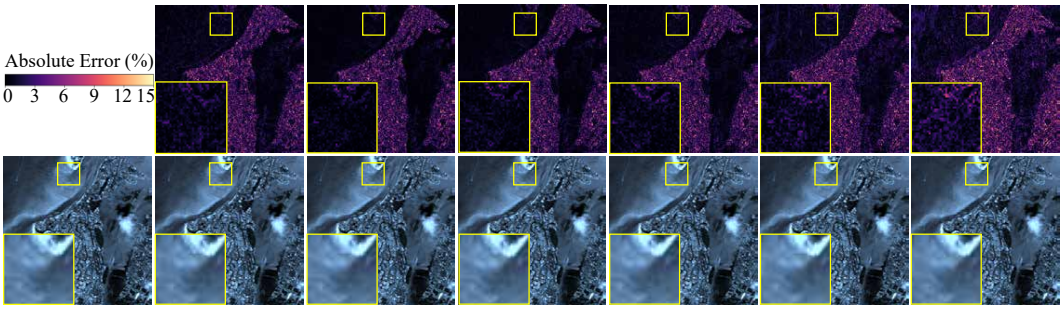

Figure 6: Qualitative comparison of reconstruction results under different quantization precisions for the Ft. Myers (L) scene in the AVIRIS dataset. The lower row shows the false-color RGB reconstructions, and the upper row presents the corresponding RGB absolute error maps. QuHIC results under the 8-, 7-, 6-, 5-, and 4-bit settings are compared with the 8-bit BTC-Net V2 (Zhou et al., 2026) result. The bit depth and corresponding PSNR are reported below each reconstruction. The yellow boxes mark representative regions, with zoomed-in views provided for detailed comparison.

C.2 Second-Order Differences

Although both RBSP and ASSCP employ a second-order difference operation, they characterize different types of variation at different representation levels. In RBSP, $\mathcal{D}^2(\bar{\boldsymbol{x}}_k)$ is directly computed along the original physical spectral dimension of $\bar{\boldsymbol{x}}_k \in \mathbb{R}^{C\times H\times W}$ before deep feature transformation. It therefore explicitly characterizes the local curvature of neighboring physical spectral bands and supplements the raw-band spectral continuity prior with higher-order wavelength variations.

ASSCP operates on the low-resolution latent representation $\boldsymbol{f}_n \in \mathbb{R}^{2C_k\times H/2\times W/2}$, where $\mathcal{D}^2(\boldsymbol{f}_n)$ is computed along the latent channel dimension rather than the original wavelength axis. Its purpose is to enhance local variation patterns among learned latent features and complement the neighboring-channel dependencies captured by the latent spectral mixer. Therefore, the second-order difference in RBSP serves as an explicit physical spectral-curvature prior, whereas that in ASSCP provides a feature-level latent variation prior for subsequent spectral-spatial interaction.

D Additional Experimental Results

D.1 Implementation Details

The input HSI contains $C = 172$ spectral bands for AVIRIS and $C = 270$ spectral bands for WHU-Hi. Let C_k denote the number of latent feature channels at the k-th stage. The latent channel dimensions of the nine stages are set to $[80, 80, 96, 112, 112, 112, 96, 80, 80]$, respectively. Accordingly, the corresponding bottleneck channel dimensions are $[160, 160, 192, 224, 224, 224, 192, 160, 160]$.

For qualitative visualization, false-color RGB images are generated from three spectral bands nearest to 650, 550, and 450 nm. For each scene, the three selected ground-truth bands are normalized using the 1st and 99th percentile values, and the same GT-derived normalization ranges are applied to all reconstructed results to ensure a fair visual comparison. The normalized channels are then transformed by a fixed linear color-mixing matrix followed by a gamma correction with 1.4, producing a consistent false-color representation for all methods.

Table 5: Rate-distortion comparison between QuHIC and BTC-Net V2 at different quantization precisions.

Bit	BTC-Net V2 (Zhou et al., 2026)			QuHIC (Ours)		
	PSNR↑	RMSE↓	SAM↓	PSNR↑	RMSE↓	SAM↓
8	40.93	48.66	1.614	**41.97**	**43.98**	**1.423**
7	40.78	51.21	1.663	**41.71**	**44.70**	**1.463**
6	40.12	54.18	1.752	**41.21**	**46.45**	**1.519**
5	38.76	61.57	1.947	**40.08**	**51.62**	**1.696**
4	36.41	78.95	2.383	**38.22**	**63.11**	**2.007**

D.2 Dataset Details

To comprehensively evaluate the reconstruction performance and generalization capability of the proposed QuHIC, we conduct extensive experiments on the widely used AVIRIS and WHU-Hi datasets. The scenes/regions selected for the test set are summarized in Table 4. Following the BTC-Net V2 preprocessing protocol, 172 effective spectral bands are retained from the AVIRIS dataset for network input, and five representative scenes are used for the main evaluation. In addition, five geographically unseen AVIRIS scene groups are employed to assess cross-region generalization. The WHU-Hi dataset, which contains approximately 270 spectral bands acquired from two UAV-borne agricultural scenes, is further used to evaluate the robustness of QuHIC under a different imaging platform and spectral distribution.

D.3 Bitrate Computation

In Table 1, the bpppb values under different bit depths are calculated directly from the low-dimensional latent representation output by the encoder. For an input HSI $\boldsymbol{X} \in \mathbb{R}^{172\times H\times W}$, the encoder (comprising sensing convolutional layers and Sigmoid activation functions) maps it into a latent representation $\boldsymbol{Y} \in \mathbb{R}^{27\times(H/4)\times(W/4)}$.

Under the 32-bit floating-point representation, the fixed bitrate without entropy coding is:

$$R_{32} = \frac{32 \times 27 \times (H/4) \times (W/4)}{172 \times H \times W} = \frac{32 \times 27}{172 \times 16} \approx 0.314 \text{ bpppb}. \tag{20}$$

Similarly, when using b-bit uniform quantization, the theoretical bitrate is uniformly expressed as:

$$R_b = \frac{27b}{172 \times 16} \text{ bpppb}. \tag{21}$$

D.4 Reconstruction under Extremely Low Bit Rates

Since most competing methods do not support configurable low-bit quantization, the broad comparison is conducted under the 32-bit setting for compatibility, while additional low-bit experiments compare QuHIC with BTC-Net V2 at matched quantization precisions. As shown in Figure 6, QuHIC maintains high reconstruction quality as the quantization precision is progressively reduced from 32 to 4 bits. Table 5 presents the corresponding rate–distortion comparison, where the 7-, 6-, 5-, and 4-bit results of BTC-Net V2 are cited from its original publication. QuHIC achieves consistently higher PSNR and lower RMSE and SAM than BTC-Net V2 across all evaluated bit depths. Although reducing the bit depth inevitably increases information loss, the reconstructed images preserve the overall scene structure and most local details, with relatively low RGB errors in major structural regions. In particular, under 8-bit quantization, QuHIC produces lower RGB reconstruction errors, sharper local structures, and fewer visible distortions than BTC-Net V2. These qualitative observations agree with the quantitative rate–distortion results, demonstrating the robustness of QuHIC to aggressive quantization and its ability to preserve spatial fidelity at low bit rates. Additional qualitative results are provided in the supplementary material.

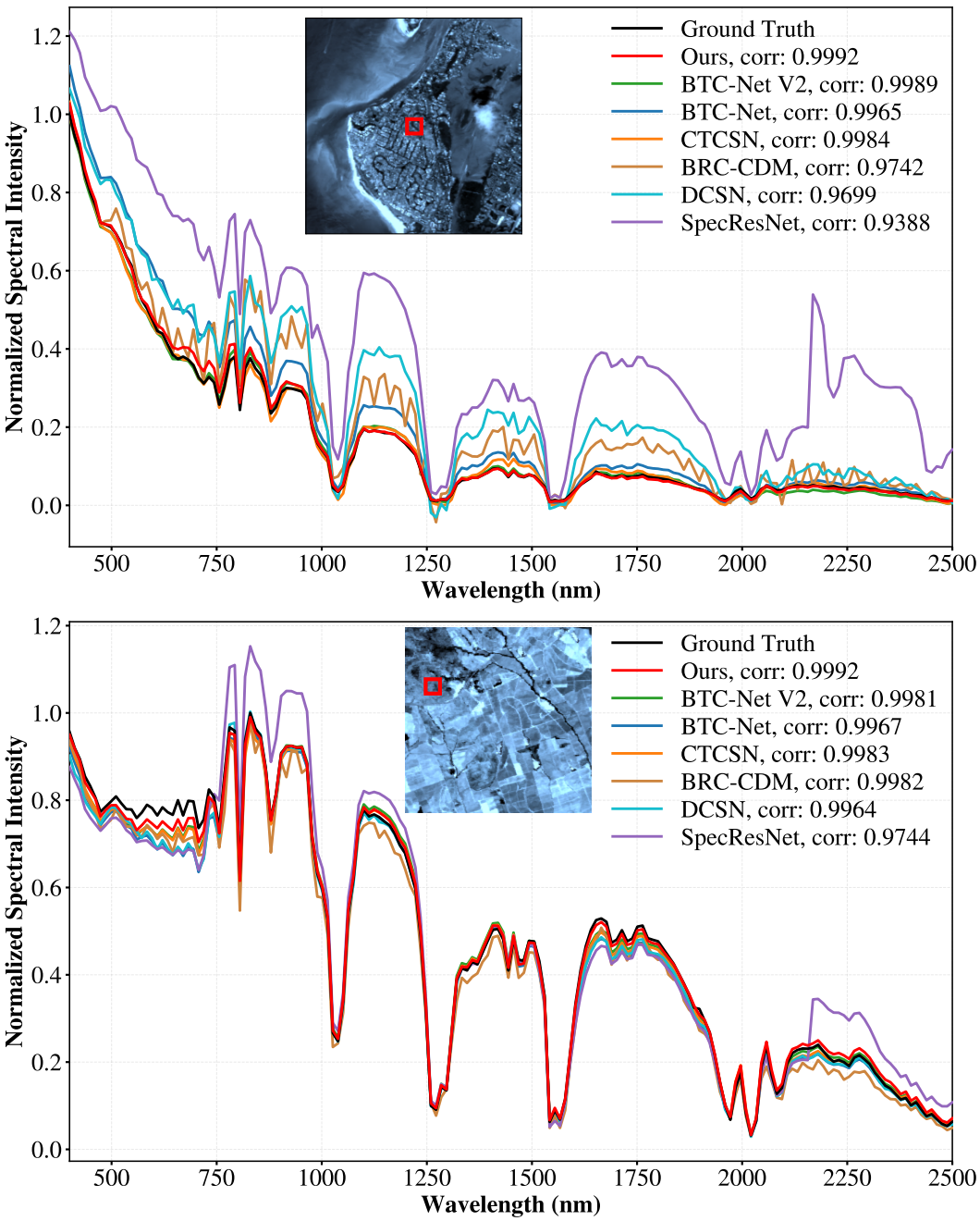

Figure 7: Spectral density curve comparison on the Ft. Myers (L) and HyspIRI17 (F) scenes of AVIRIS. The inset false-color images indicate the spatial regions used for spectral extraction, and the curves are computed from all hyperspectral bands within the corresponding regions. The correlation coefficient with the ground truth is reported for each method.

D.5 Spectral Density Curve

As shown in Figure 7, QuHIC closely follows the ground-truth spectral responses over the full wavelength range in both representative AVIRIS scenes. The high correlation coefficients further indicate that QuHIC effectively preserves the spectral characteristics of the original HSI, and reduces the noticeable spectral deviations observed in several competing methods.

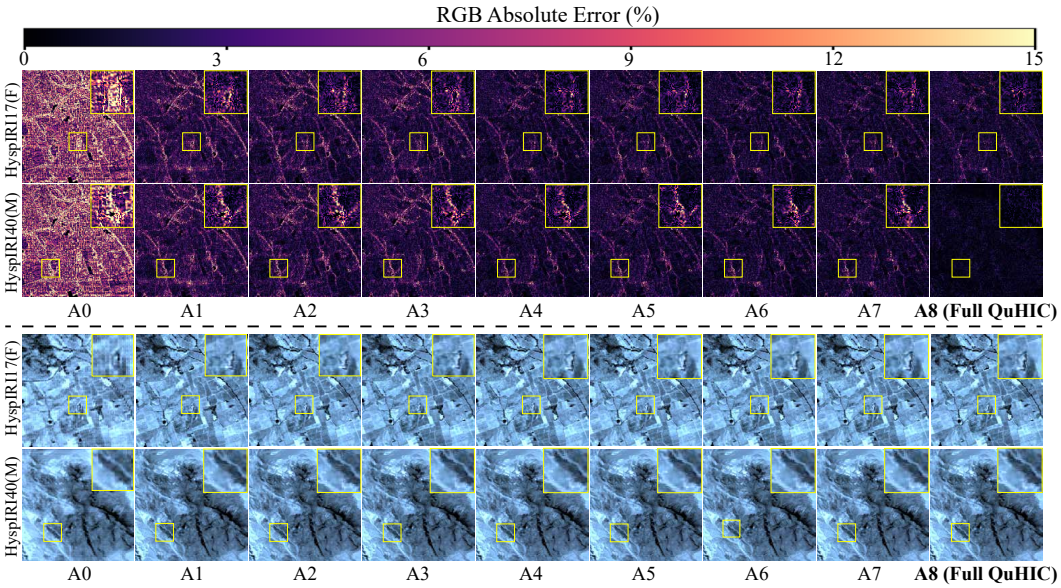

Figure 8: Ablation studies on the HyspIRI17 (F) and HyspIRI40 (M) scenes of the AVIRIS dataset under 8-bit quantization. The lower part presents the false-color RGB reconstructions together with the ground truth, and the upper part shows the corresponding RGB absolute error maps for variants A0-A8.

D.6 Ablation Studies

To provide a clear evaluation of the proposed components, Figure 8 presents a visual comparison across variants under 8-bit quantization, which exposes quantization effects while preserving stable reconstruction. As key modules are integrated from A0 to A8, reconstruction quality improves progressively. Specifically, baseline A0 suffers from severe structural blurring due to direct backprojection after quantization. A1 substantially mitigates unrolling residuals and quantization noise, restoring overall scene structures. A2 and A3 highlight the impact of RBSP: A2 stabilizes inter-band representations, while A3 models first- and second-order spectral continuity, eliminating spectral jump artifacts and recovering fine spectral variations. These spectral priors guarantee high fidelity across steep wavelength transitions without localized distortion. The addition of MSDP in A4 enriches multi-scale spatial features, sharpening object boundaries. Within ASSCP, A5 to A7 progressively introduce dual-branch local modeling, cross-branch communication, and adaptive context modeling, suppressing ringing artifacts via dynamic spectral-spatial recalibration. This joint processing enables seamless interaction between spatial texture refinement and spectral consistency. Finally, A8 incorporates GSF to selectively fuse decoder-refined features with high-resolution skip connections via gated routing. This progressive design demonstrates the complementary strength of each component, achieving sharp spatial structures and optimal spectral-spatial fidelity.

D.7 Generalization Capability Evaluation

As shown in Figure 9, QuHIC consistently exhibits lower RGB reconstruction errors across AVIRIS, WHU-Hi, and unseen AVIRIS scenes. The corresponding false-color reconstructions also preserve local structures and spatial details more faithfully in the highlighted regions, demonstrating the strong reconstruction fidelity and generalization capability of QuHIC.

D.8 Band-wise qualitative comparison

As shown in Figure 10, QuHIC consistently produces lower reconstruction errors across representative spectral bands from AVIRIS, WHU-Hi, and unseen AVIRIS scenes. Compared with competing methods, QuHIC better preserves local structures and fine spatial details in the highlighted regions, and its reconstructed band images remain visually closer to the ground truth. These qualitative results further demonstrate the spectral reconstruction fidelity and generalization capability of QuHIC.

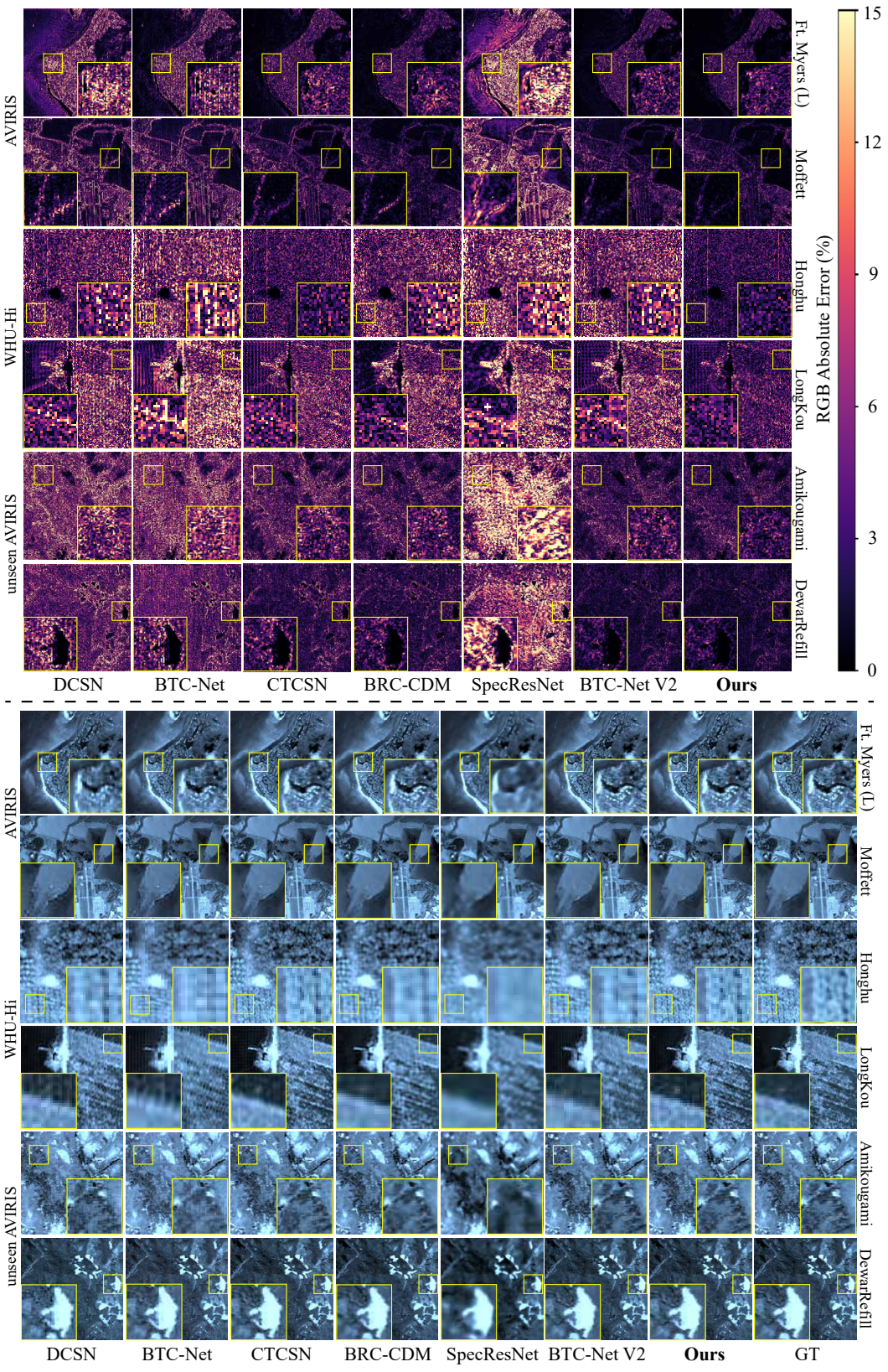

Figure 9: Qualitative comparison on AVIRIS, WHU-Hi, and unseen AVIRIS scenes. The lower part presents the false-color RGB reconstructions, and the upper part shows the corresponding RGB absolute error maps, where brighter colors indicate larger reconstruction errors.

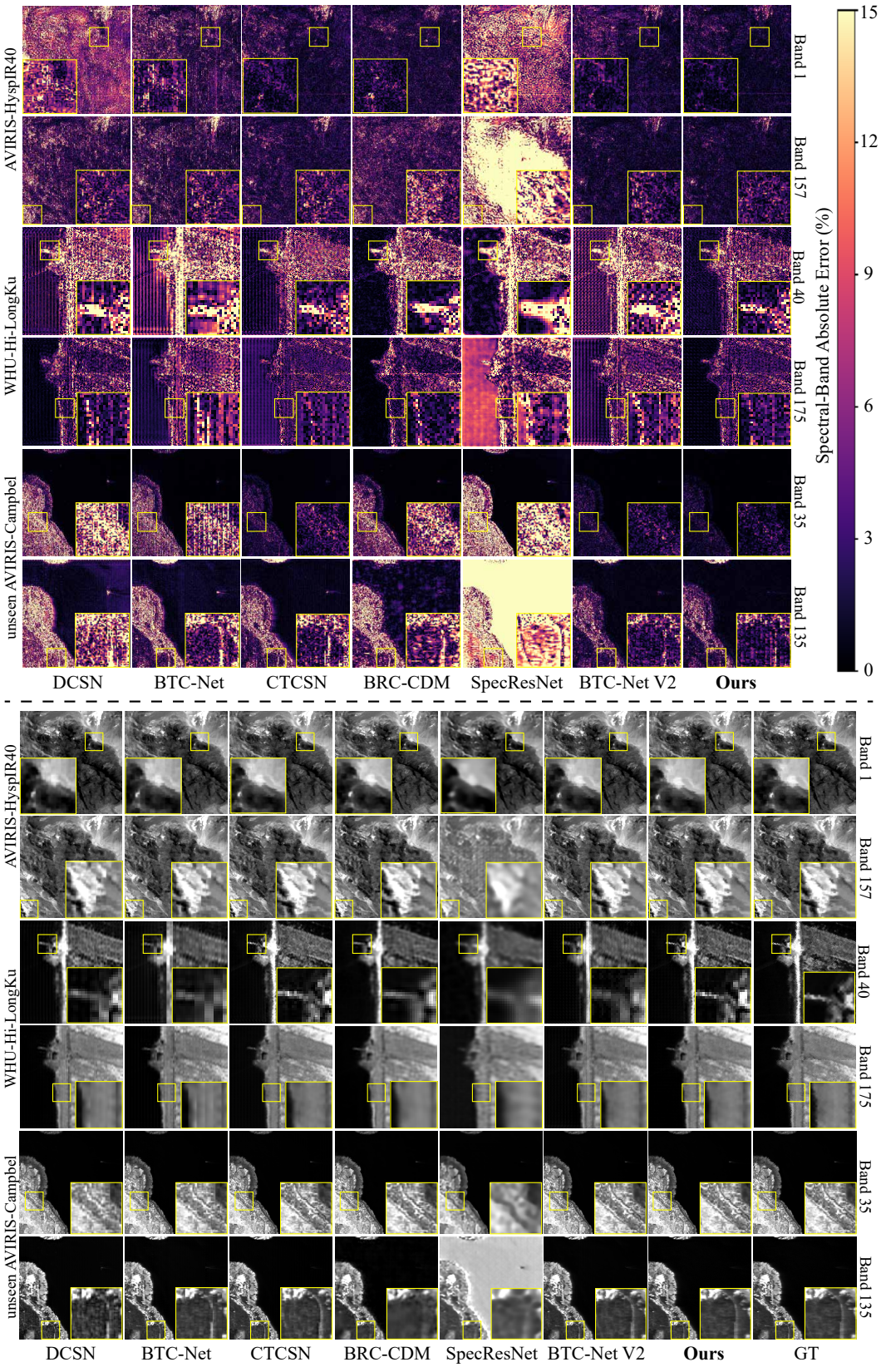

Figure 10: Band-wise qualitative comparison on AVIRIS, WHU-Hi, and unseen AVIRIS scenes. The lower part presents the reconstructed spectral-band images together with the ground truth, and the upper part shows the corresponding spectral-band absolute-error maps.