

# 000 BINARY NEURAL NETWORK FOR HYPERSPECTRAL 001 002 PAN-SHARPENING 003 004

005 **Anonymous authors**

006 Paper under double-blind review

## 007 008 ABSTRACT 009

010 Hyperspectral pan-sharpening aims to generate high-resolution hyperspectral  
011 (HRHS) images by fusing low-resolution hyperspectral (LRHS) data with high-  
012 resolution panchromatic (PAN) images, enabling applications in mapping, surveil-  
013 lance, and environmental monitoring. While deep learning methods achieve  
014 strong performance, their heavy computational and memory demands limit de-  
015 ployment on resource-constrained satellite platforms. To address this, we explore  
016 binary neural networks (BNNs) for hyperspectral pan-sharpening. Conventional  
017 binarization, however, introduces gradient instability and severe information loss,  
018 compromising spectral spatial fidelity. We propose the Adaptive Tan Identity  
019 Straight-Through Estimator (ATISTE), a soft binarization strategy that decouples  
020 forward approximation from gradient propagation and employs adaptive scaling to  
021 preserve consistency with full-precision features. Building on ATISTE, we design  
022 HS-BiNet, a lightweight binary CNN with residual connections and multi-scale  
023 fusion, to effectively capture spectral-spatial dependencies while avoiding com-  
024 putationally intensive operations such as unfolding inference and non-local self-  
025 attention. This ensures suitability for real-time deployment on edge and satellite  
026 platforms. Extensive experiments show that HS-BiNet consistently outperforms  
027 binary baselines and remains competitive with, and in some cases surpasses, full-  
028 precision models, offering a practical solution for high-fidelity HRHS reconstruc-  
029 tion.

## 030 031 1 INTRODUCTION

032 High-resolution hyperspectral (HRHS) images are vital in numerous remote sensing applications,  
033 including mapping services, military surveillance, and environmental monitoring Ram et al. (2024);  
034 Sharma et al. (2020); Pallas Enguita et al. (2024); Carvalho et al. (2019); Stuart et al. (2019); Bedini  
035 (2017), due to their superior spectral fidelity and fine spatial detail. However, directly acquiring  
036 HRHS images remains a technical challenge, as existing remote sensing sensors are limited in their  
037 ability to capture high spectral and spatial resolutions simultaneously. To mitigate this constraint,  
038 hyperspectral pan-sharpening has emerged as a reliable solution, whereby low-resolution hyperspec-  
039 tral (LRHS) images are fused with high-resolution panchromatic (PAN) images to generate enhanced  
040 HRHS outputs. This fusion process exploits the rich spectral information of HS data and the spatial  
041 granularity of PAN images, producing outputs that approximate true HRHS quality.

042 In recent years, deep learning (DL)-based methods have become the dominant approach for hyper-  
043 spectral pan-sharpening, surpassing traditional techniques Cai & Huang (2020); Yang et al. (2017).  
044 Early models like HyperPNN He et al. (2019) introduced CNNs specifically designed for hyper-  
045 spectral data, with spectral encoders, decoders, and spatial-spectral fusion modules. Building on  
046 this, HSNet Hu et al. (2022) added preprocessing subnets for enhanced PAN feature extraction,  
047 DenseNet-inspired fusion, and spectral-aware loss functions to balance spatial detail and spectral  
048 consistency. Further advances include DHP-DARN Zheng et al. (2020) and DIP-HyperKite Ban-  
049 dara et al. (2021), which use deep image prior (DIP) methods to regularize spectral upsampling  
050 and improve fusion under limited data. Notably, DIP-HyperKite’s inverse U-Net enables spatial  
051 over-expansion, enhancing high-frequency detail beyond PAN resolution. Most recently, HyperD-  
052 SNet Zhuo et al. (2022) combined handcrafted edge detectors (e.g., Sobel, Prewitt) with multiscale  
053 deep-shallow feature extraction and per-band Spectral Attention for precise, adaptive detail injection  
across hyperspectral channels. Since then, a flood of increasingly complex network architectures has

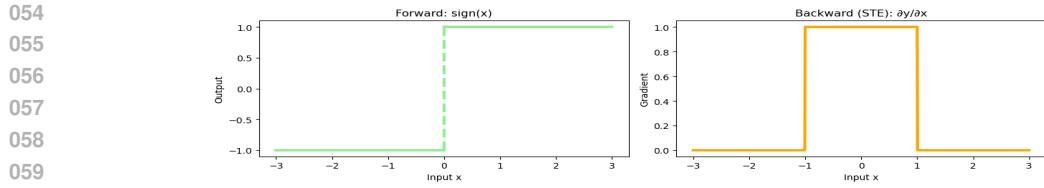


Figure 1: Illustration of Forward and Backward Propagation for the Sign Function.

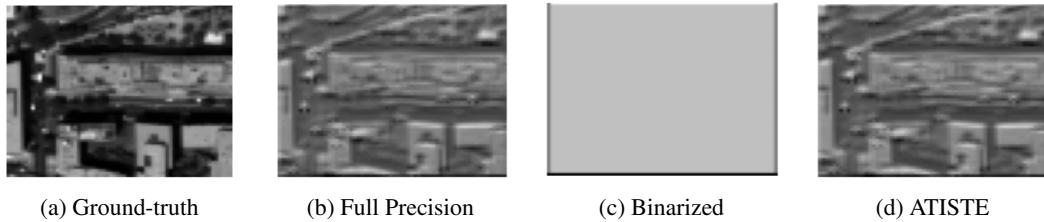


Figure 2: Analysis of binarization: Given the ground-truth (a) and full-precision feature map (b), the corresponding binarized activation (c) exhibits significant degradation, while ATISTE outputs (d) retain more information and better preserve the original features.

073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
been proposed to improve further pan-sharpening outcomes, including multi-scale designs, adaptive frameworks, spatial-frequency representations, and transformer-based models, all contributing to richer spatial-spectral fusion capabilities Bandara & Patel (2022); Dong et al. (2021); Sun et al. (2024); Guan & Lam (2021); Wu et al. (2022); Feng et al. (2024). Despite their strong performance, these state-of-the-art models often demand considerable computational resources, such as high-end GPUs, making them impractical for deployment on resource-constrained satellite platforms. Driven by this challenge, we explore network binarization for hyperspectral pan-sharpening to reduce computational complexity while substantially maintaining high reconstruction fidelity. We begin by analyzing the transformation of real-valued activations under the binarization process. As illustrated in Fig. 2, binarized feature representations suffer from significant information loss compared to their full-precision counterparts. This degradation is especially critical in hyperspectral pan-sharpening, where pixel-level accuracy is essential to preserve spatial detail and spectral integrity. Conventional binarization approaches often fail to maintain this fidelity, leading to suboptimal reconstruction quality. Despite significant progress, prior works on quantization/binarization and surrogate gradient methods for BNNs reveal consistent limitations. Early binarization approaches such as XNOR-Net Rastegari et al. (2016b) enabled efficient binary convolutions but suffered large accuracy drops due to crude sign-based binarization and gradient mismatch. Follow-up studies identified forward-backwards inconsistency, where the forward sign function is discontinuous (as shown in Fig. 1), while backwards estimators (e.g., identity mapping in classic STE) introduce severe bias and vanishing gradients Yin et al. (2019). Attempts to close the accuracy gap in low-bit and binary networks, such as 2-bit quantization methods Choi et al. (2018) and differentiable soft quantization Gong et al. (2019b), improved forward approximation but often relied on rigid functional forms or hand-designed smooth surrogates that could not dynamically adapt, leading to unstable gradients and limited scalability. More advanced strategies, such as information retention Qin et al. (2020c), coupled binary activations Kim et al. (2020), and reviving dead weights Xu et al. (2021b), addressed specific failure cases such as gradient starvation or inactive parameters, but typically introduced additional heuristics or complex mechanisms without fully resolving the fundamental trade-off between gradient stability and binarization fidelity. Even the recent Wu et al. (2023), which reframes STE through an equilibrium perspective, still couples approximation accuracy and gradient stability via a single parameter, causing instability when scaled to deeper architectures or larger datasets. Collectively, these works expose four persistent issues: forward-backwards mismatch, gradient vanishing or explosion, rigid or over-parameterized surrogates with poor adaptability, and unstable convergence in deeper networks.

104  
105  
106  
107  
To overcome these limitations, we introduce Adaptive Tan Identity Straight-Through Estimator (ATISTE), a soft binarization strategy based on a novel straight-through estimator that balances binarization fidelity with gradient stability. Our STE employs a dual-path surrogate formulation: one path leverages a smooth, saturating function (a bounded tanh variant) to stabilize gradients during backpropagation (mitigating gradient vanishing or explosion), and the second path preserves a

sharp approximation for practical forward inference (reducing forward–backwards mismatch). This decoupling enables independent control over the approximation in each pass, ensuring effective gradient flow even in saturated regions and improving convergence stability in deeper networks. Furthermore, STE incorporates an adaptive channel-wise scaling mechanism derived from activation statistics, aligning the magnitude of binary activations with their real-valued equivalents (addressing rigidity of fixed surrogates and reducing information loss). This mechanism further enhances adaptability, making STE particularly suitable for high-precision tasks like hyperspectral fusion.

Building on STE, we develop HS-BiNet, a compact and efficient convolutional architecture specifically tailored for hyperspectral pan-sharpening. HS-BiNet employs a modular structure in which key convolutional layers are binarized using STE-enhanced binary convolutions. We incorporate residual connections and intermediate skip links to strengthen information propagation and preserve contextual awareness. Additionally, the network leverages lightweight local fusion units and multi-scale receptive fields to effectively capture the intricate spectral and spatial dependencies characteristic of hyperspectral imagery. This combination of STE-driven binarization and architectural refinements enables HS-BiNet to maintain high reconstruction fidelity while significantly reducing bitwidth and computation. Extensive experiments confirm that HS-BiNet consistently surpasses conventional binary baselines across all key evaluation metrics and narrows the performance gap with full-precision models. In several cases, it even outperforms full-precision methods, demonstrating the practical viability and strength of the proposed approach for real-world hyperspectral fusion tasks.

The main contributions of this work are summarized as follows:

1. We propose a novel straight-through estimator that introduces a dual-path gradient approximation strategy and an adaptive scaling mechanism, significantly improving the stability and accuracy of binarized neural networks. This work presents the first known application of binary neural networks to hyperspectral pan-sharpening.
2. We design HS-BiNet, a lightweight binary CNN architecture tailored for hyperspectral pan-sharpening, which integrates ATISTE-based binary convolutions with residual and multi-scale modules to ensure effective spectral-spatial feature learning.
3. We demonstrate that our binarized network achieves state-of-the-art performance among binary models and remains competitive with, and even surpasses, many full-precision models, all while drastically reducing memory and computation.

## 2 RELATED WORK

### 2.1 DEEP LEARNING-BASED PAN-SHARPENING

In recent years, deep learning has made remarkable advances in low-level vision tasks, profoundly impacting pan-sharpening, where DL methods now lead He et al. (2023; 2024). HyperPNN He et al. (2019) pioneered the use of compact CNNs for hyperspectral fusion, competing well with traditional optimization-based methods. Hyper-DSNet Zhuo et al. (2022) introduced band-wise processing to handle inter-band variability, while DIP-HyperKite Bandara & Patel (2022) combined Deep Image Prior upsampling with an over-complete HyperKite network to learn residual high-frequency details under spatial and spectral constraints. Subsequently, the field rapidly advanced with architectures like AIDB-Net Sun et al. (2024), employing adaptive information distillation blocks to enhance key spectral and spatial features, and FPFNet Dong et al. (2023), which used progressive feature fusion for improved detail recovery. LPPNet Dong et al. (2021) further improved effectiveness by integrating local and global priors. Moving beyond CNNs, transformer-based models, such as HyperTransformer Bandara & Patel (2022), apply self-attention to capture long-range spectral dependencies, while diffusion models target enhanced fusion quality. Despite their strong performance, these methods incur high computational and memory costs, which restricts their deployment on resource-limited satellite platforms. To tackle this, our work investigates binarized neural networks as a lightweight and effective alternative for pan-sharpening.

### 2.2 BINARY NEURAL NETWORK

Binarization is the most extreme form of model quantization, compressing networks by restricting both weights and activations to binary values of  $-1$  and  $+1$ , which significantly reduces storage

consumption and computational cost. Hubara et al. (2016) introduce the first binarized neural network by directly quantizing parameters during training. Building upon this idea, Rastegari et al. (2016b) incorporate learnable scaling factors in XNOR-Net, achieving up to a  $58\times$  acceleration and a  $32\times$  reduction in memory requirements. Motivated by these efficiency gains, recent works have increasingly explored binary networks for various low-level vision tasks Cai et al. (2023); Chen et al. (2024); Jiang et al. (2023); Song et al. (2023); Xin et al. (2020); Xia et al. (2022). Xin et al. (2020) develop the first fully binarized super-resolution model by binarizing both weights and activations. Jiang et al. (2021) further propose a binary training framework that removes batch normalization layers, while Xia et al. (2022) design a basic binary convolution unit for image restoration through detailed component analysis. Frequency decomposition is separately handled in Jiang et al. (2023), and Chen et al. (2024) introduce a binarized diffusion model for image SR. Remote sensing satellites, which operate under strict constraints on computation, memory, and energy, present a natural need for highly compact models, and binarization provides a promising solution for efficient onboard processing. However, applying conventional binary networks directly to hyperspectral imaging remains challenging. Hyperspectral data consist of many narrow and continuous spectral bands, and the strong quantization in BNNs causes severe information loss that is far more harmful for spectral fidelity than in RGB imagery. Their limited representation capacity also makes it difficult to capture the complex spectral, spatial relationships present in hyperspectral scenes Guerri et al. (2024); Tang et al. (2024). Moreover, the reduced dynamic range of binary activations can distort fine spectral variations and degrade the accuracy of reconstructed spectral curves Hou et al. (2025). Training is further complicated by the non-smooth nature of binarization Qin et al. (2020a). Binary constraints can also weaken band-wise spectral consistency, which is critical for downstream tasks such as material classification and anomaly detection. Due to the high dimensionality and strong inter-band correlations in hyperspectral images, binary networks often lack sufficient expressive power to generalize well. As a result, although binary networks offer significant efficiency benefits, their direct application to hyperspectral imaging faces notable limitations that must be addressed for effective hyperspectral fusion.

### 3 METHODOLOGY

#### 3.1 OVERALL ARCHITECTURE

As illustrated in Fig. 3, the overall framework of our proposed *HS-BiNet* consists of four main components: (1) an Multi-Scale Feature Extractor (MSFE), (2) an Edge Injector Module, (3) a Fusion Module, and (4) a Decoder Module. All components are built on binary convolution (detailed in APPENDIX A.1), implemented with our Straight-Through Estimator (STE) to enable efficient training and inference while preserving representational capacity. Given an input low-resolution hyperspectral image  $\mathbf{HS} \in \mathbb{R}^{H \times W \times B}$  with  $B$  spectral bands and a high-resolution panchromatic image  $\mathbf{P} \in \mathbb{R}^{sH \times sW}$  with  $s$  as the spatial upsampling factor, we first apply a deformable convolution-based Dai et al. (2017) upsampler to the hyperspectral image to match the spatial resolution of  $\mathbf{P}$ , producing the initial feature map  $\mathbf{F}_0 \in \mathbb{R}^{sH \times sW \times B}$ . The upsampled hyperspectral input is then processed by the MSFE module for spectral feature extraction, where multi-scale features are obtained using binary convolutions with kernel sizes  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ . The resulting feature maps are concatenated along the channel dimension to form a multi-scale feature representation  $\mathbf{F}_{ms} \in \mathbb{R}^C$ , where  $C$  represents the hidden states. A subsequent  $1 \times 1$  binary convolution reduces the concatenated features from  $3C$  to  $C$  channels, followed by a residual connection from the input to preserve the original spectral content during multi-scale fusion. This design enriches the representation with multi-scale contextual information while maintaining constant channel size and low computational cost.

Following the multi-scale extraction, the MSFE employs Binary Residual Blocks, each consisting of two binarized  $3 \times 3$  convolutions with stride 1 and padding 1, each followed by batch normalization. A residual connection from the block input to its output preserves spectral information across layers and mitigates degradation. This residual formulation enables the block to learn corrective transformations rather than complete mappings, improving training stability and reconstruction accuracy. At each stage of processing, the Edge Injector applies a binary convolution to the PAN image, producing spatial features  $\mathbf{F}_{sp} \in \mathbb{R}^C$ . This transforms the single-channel PAN input into a multi-channel feature representation that highlights spatial details for integration with the spectral features extracted by the MSFE. The Fusion unit receives  $\mathbf{F}_{sp}$  from the edge injector and  $\mathbf{F}_{spc}$  from

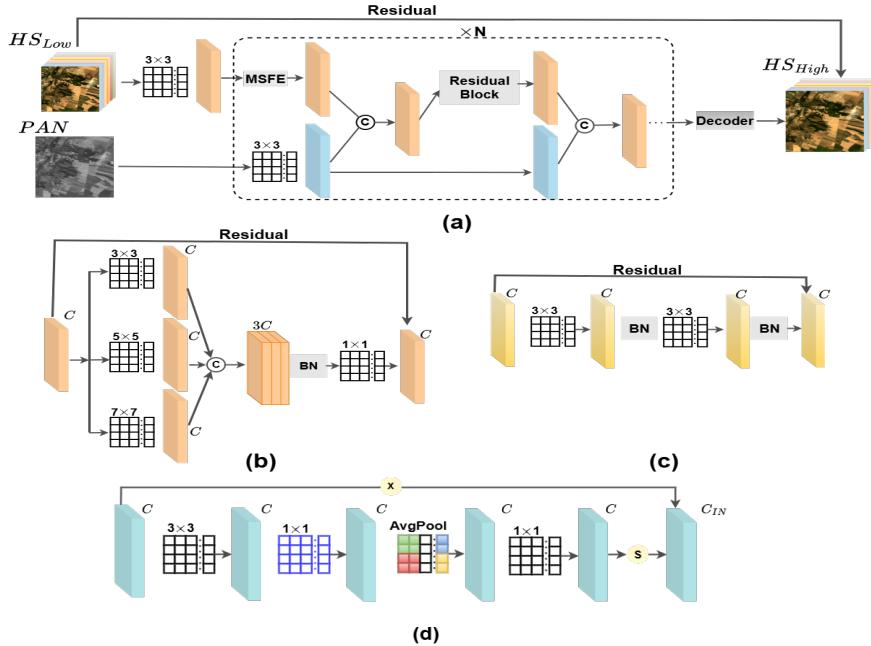


Figure 3: The framework of the proposed HS-BiNet, with hidden dimension  $C$ , comprises four key components: (a) overall network architecture, (b) multi-scale feature extractor, (c) residual block, and (d) decoder block, where  $\odot$  denotes concatenation of feature maps and BN refers to BatchNorm. In the illustration, black-colored kernels represent the binary convolution layers, while blue-colored kernels indicate the full-precision convolution layers.

the MSFE, concatenating them along the channel dimension to form a combined spectral-spatial representation. This representation is then processed by a binary convolution to project it back to the target feature dimension  $C$ , ensuring compatibility with subsequent stages while preserving spectral fidelity and spatial sharpness.

Finally, the Decoder refines the fused spectral-spatial representation to produce the high-resolution hyperspectral output  $\hat{\mathbf{HS}}$ . The input feature map  $\mathbf{F}_{in} \in \mathbb{R}^C$  is first processed by a binarized  $3 \times 3$  convolution to enhance both spatial and channel-level details, followed by a  $1 \times 1$  convolution to adjust channel dimensionality and reduce computational complexity. A lightweight enhancement mechanism then emphasizes the most informative features while suppressing less relevant responses. This mechanism begins with adaptive average pooling to produce a compact descriptor for each channel, followed by a binarized  $1 \times 1$  convolution and a sigmoid activation to compute scaling factors in the range  $[0, 1]$ . These factors are multiplied element-wise with the original feature map to produce the enhanced output  $\mathbf{F}_{out}$ . A global residual connection adds the upsampled hyperspectral input  $\mathbf{HS}$  to the decoder output, resulting in the final high-resolution hyperspectral image  $\hat{\mathbf{HS}}$  that preserves the original spectral signatures while effectively integrating the spatial information from the panchromatic image.

### 3.2 STRAIGHT THROUGH ESTIMATOR

Binarized neural networks face a fundamental challenge in balancing forward fidelity with gradient stability. Naïve binarization Qin et al. (2020a) through the discontinuous sign function introduces severe forward-backward mismatch, where the forward mapping discards critical information while identity-based backward surrogates bias gradients and collapse outside a narrow band. This mismatch is further compounded by gradient starvation: hard or overly sharp surrogates yielding vanishing derivatives in saturation regions or unstable spikes near zero, both of which hinder optimization in deeper architectures. Attempts to address these limitations through smooth approximations such as HardTanh, SoftSign, or Log-Tanh reduce discontinuity but either remain too soft to enforce effective binarization or too rigid to maintain stable gradients. Methods such as ReSTE Wu et al. (2023) struggle when applied to high-dimensional or multi-channel data, where different channels often re-

270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 quire different levels of gradient stability and approximation accuracy. In hyperspectral fusion, this challenge becomes much more severe because the task demands extremely fine spectral precision across hundreds of closely spaced bands. The adaptive parameter in ReSTE, which is intended to balance approximation fidelity and gradient stability, cannot satisfy these conflicting requirements across all spectral channels, leading to suboptimal approximations and unstable gradient flow. Even small residual estimator bias or remaining gradient noise can distort subtle spectral curves, and the progressive schedule of ReSTE, while improving stability, still fails to preserve the delicate spectral details required for accurate hyperspectral reconstruction. As a result, although ReSTE is more stable than standard STE methods, it remains insufficient for maintaining high spectral fidelity and continuous-value accuracy in hyperspectral imaging. The proposed ATISTE addresses the fundamental trade-off between forward approximation fidelity and gradient stability that limits existing surrogate estimators in BNN training. Its formulation integrates a smooth, saturating non-linearity with a linear residual, resulting in a flexible surrogate that guarantees non-zero gradients and bounded variance. Formally, for an input activation  $x \in \mathbb{R}$ , the forward surrogate is defined as:

$$f_{\alpha,\lambda}(x) = (1 - \lambda) \tanh(\alpha x) + \lambda x, \quad (1)$$

where the sharpness parameter  $\alpha > 0$  controls the steepness of the nonlinear path, and the residual weight  $\lambda \in [0, 1]$  adjusts the contribution of the identity path. This design enables smooth interpolation between two established estimation strategies. Specifically, setting  $\lambda = 0$  reduces the estimator to a pure  $\tanh(\alpha x)$ , acting as a smooth approximation of the sign function, while  $\lambda = 1$  recovers the identity mapping, corresponding to the classic STE backward surrogate. By selecting intermediate  $\lambda$  values, ATISTE seamlessly interpolates between these extremes, allowing a controlled transition from soft to hard binarization during training. The backward pass supplies the pseudo-gradient used in optimization, derived as:

$$\frac{\partial f_{\alpha,\lambda}(x)}{\partial x} = (1 - \lambda)\alpha (1 - \tanh^2(\alpha x)) + \lambda. \quad (2)$$

Equation 2 reveals two critical structural terms. The first term, the scaled derivative of  $\tanh$ , is sharply localized near zero and facilitates learning in the binarization-critical region. The second term, a constant gradient floor of magnitude  $\lambda$ , ensures non-zero gradients even when  $|x|$  is large and the  $\tanh$  path saturates. Unlike hard sign surrogates, which suffer from vanishing gradients outside  $[-1, 1]$ , ATISTE maintains informative gradients everywhere. Furthermore, the boundedness of the derivative is expressed as:

$$\lambda \leq f'_{\alpha,\lambda}(x) \leq (1 - \lambda)\alpha + \lambda. \quad (3)$$

This dual bound prevents gradient starvation (via the non-zero floor  $\lambda$ ) and gradient explosion (via the finite ceiling  $(1 - \lambda)\alpha + \lambda$ ). Such properties offer a clear advantage over power-law surrogates, whose derivatives can either diverge near zero or vanish in saturation regions. From the forward perspective, ATISTE is rational: its approximation error is no worse than that of the classic STE. Defining the mean-squared error relative to the ideal sign function as:

$$E(\alpha, \lambda) = \mathbb{E} \left[ (\text{sign}(x) - f_{\alpha,\lambda}(x))^2 \right], \quad (4)$$

we can analyze its asymptotic behavior. As  $\alpha \rightarrow \infty$ , the nonlinear component converges pointwise to the sign function:

$$\lim_{\alpha \rightarrow \infty} f_{\alpha,\lambda}(x) = (1 - \lambda) \text{sign}(x) + \lambda x, \quad (5)$$

yielding:

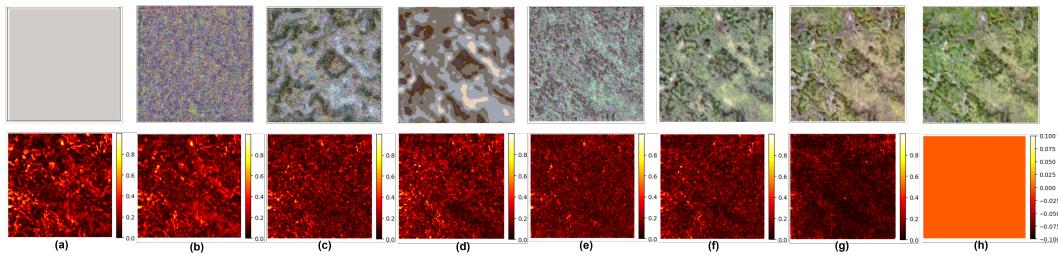
$$\lim_{\alpha \rightarrow \infty} E(\alpha, \lambda) = \lambda^2 \mathbb{E} \left[ (\text{sign}(x) - x)^2 \right] \leq E_{\text{STE}}. \quad (6)$$

Thus, for all  $\lambda < 1$ , ATISTE achieves strictly lower or equal approximation error compared to the identity-based STE.

A key feature of ATISTE lies in its tunable bias-variance trade-off. Increasing  $\alpha$  sharpens the forward mapping, reducing bias by making  $f_{\alpha,\lambda}$  more sign-like, while increasing local gradient variance. Conversely, increasing  $\lambda$  strengthens the residual path, raising the gradient floor and reducing variance, but at the cost of forward bias. The decoupling into two independent parameters provides explicit control over optimization dynamics, enabling schedules where  $\alpha$  is gradually increased and  $\lambda$  decreased over training epochs. This ensures smooth convergence to a strongly binarized model without starving gradients. Because of its dual-path design, ATISTE naturally integrates bounded non-zero gradients (Eq. 3) with a rational forward approximation (Eqs. 4, 5, 6) and

324  
 325 Table 1: Quantitative comparison of our HS-BiNet with representative full-precision and binary  
 326 methods on the Botswana and WDC datasets. Here, **US** denotes unsupervised methods and **S** de-  
 327 notes supervised methods.

328 Category	329 Method	330 Type	331 Botswana					332 Washington DC Mall (WDC)				
			333 PSNR	334 CC	335 SSIM	336 SAM	337 ERGAS	338 PSNR	339 CC	340 SSIM	341 SAM	342 ERGAS
338 Full-Precision	339 GLP	340 US	32.559	0.951	0.837	1.383	1.207	27.946	0.934	0.761	6.546	5.110
	341 GSA	342 US	31.739	0.939	0.828	1.389	1.386	24.462	0.906	0.671	7.846	6.079
	343 CNMF	344 US	30.220	0.917	0.788	1.934	1.718	24.604	0.890	0.678	8.441	6.682
	345 HySure	346 US	30.610	0.928	0.796	1.747	1.595	25.598	0.913	0.718	7.254	5.834
	347 HyperPNN	348 S	33.114	0.961	0.873	1.366	1.195	29.258	0.945	0.860	4.051	5.749
	349 HSpNet1	350 S	31.746	0.942	0.844	1.456	1.663	29.634	0.960	0.870	4.039	4.266
	351 HSpNet2	352 S	32.575	0.953	0.849	1.400	1.348	29.700	0.961	0.872	4.009	4.261
	353 FusionNet	354 S	32.506	0.952	0.850	1.397	1.367	29.696	0.959	0.866	3.917	4.339
	355 Hyper-DSNet	356 S	33.538	0.964	0.876	1.305	1.126	30.232	0.964	0.875	4.102	3.943
	357 FPFNet	358 S	33.451	0.962	0.871	1.369	1.135	30.291	0.957	0.855	4.440	4.250
358 FusionMamba	359 S	360 S	33.943	0.966	0.881	1.277	1.076	31.860	0.965	0.881	3.755	3.882
	361 DM-ZS	362 US	39.280	0.922	0.901	1.384	1.372	38.160	0.955	0.899	4.544	6.562
363 Binary	364 BNN	365 S	23.234	0.119	0.845	3.140	5.044	13.054	0.012	0.665	18.057	17.085
	366 ReActNet	367 S	27.532	0.560	0.862	2.740	3.860	20.256	0.612	0.734	11.110	12.027
	368 BTM	369 S	27.485	0.596	0.881	2.946	2.570	17.311	0.072	0.646	18.111	11.674
	370 FABNet	371 S	23.853	0.175	0.832	3.421	3.463	14.700	0.040	0.646	18.312	12.299
	372 ReSTE	373 S	27.959	0.659	0.887	2.792	2.265	24.208	0.759	0.823	8.195	6.665
	374 Bi-DiffsR	375 S	30.821	0.798	0.883	2.373	2.666	25.989	0.845	0.882	8.007	8.447
	376 <b>HS-BiNet</b>	377 <b>S</b>	<b>37.285</b>	<b>0.900</b>	<b>0.891</b>	<b>1.622</b>	<b>1.852</b>	<b>36.003</b>	<b>0.947</b>	<b>0.885</b>	<b>5.468</b>	<b>4.781</b>
	378 Ideal values		$+\infty$	1	1	0	0	$+\infty$	1	1	0	0



354 Figure 4: Qualitative comparison of our model with representative binary methods on the reduced-  
 355 resolution Botswana dataset. (a) BNN, (b) FabNet, (c) ReactNet, (d) ReSTE, (e) BTM, (f) Bi-DiffsR,  
 356 (g) HS-BiNet, and (h) Ground-truth.

357 explicit control over the error–stability trade-off. These properties collectively explain the empiri-  
 358 cally observed faster convergence and higher accuracy of ATISTE in deep BNNs compared to prior  
 359 single-parameter surrogates. A detailed discussion of ATISTE’s rationality, flexibility, estimation  
 360 error, and gradient instability is provided in Appendix A.2.2.

## 362 4 EXPERIMENT

364 In this section, we present the evaluation on reduced-resolution and full-resolution hyperspectral  
 365 pansharpening. We further provide extensive ablation studies to validate the effectiveness of our  
 366 STE. Due to page limitations, detailed descriptions of the datasets, baselines, and experimental  
 367 settings are included in the Appendix A.2.

### 369 4.1 COMPARISON WITH STATE-OF-THE-ART

370 We first evaluate the similarity between the fused hyperspectral images and the ground-truth data  
 371 using reduced-resolution experiments on the Washington DC Mall (WDC) and Botswana datasets.  
 372 As shown in Table 1, our method clearly outperforms existing binary approaches (BNN Hubara et al.  
 373 (2016), ReActNet Liu et al. (2020), BTM Jiang et al. (2021), FABNet Jiang et al. (2023), ReSTE Wu  
 374 et al. (2023), Bi-DiffsR Chen et al. (2024)) across all evaluation metrics. Our model also achieves  
 375 comparable results to many full-precision baselines (GLP Aiazzi et al. (2006), GSA Aiazzi et al.  
 376 (2007), CNMF Yokoya et al. (2011), HySure Simoes et al. (2014), HyperPNN He et al. (2019),  
 377 HSpNet1/2 Hu et al. (2022), FusionNet Deng et al. (2020), Hyper-DSNet Zhuo et al. (2022),  
 378 FPFNet Dong et al. (2023), FusionMamba Peng et al. (2024), DM-ZS Xiao et al. (2025)), show-

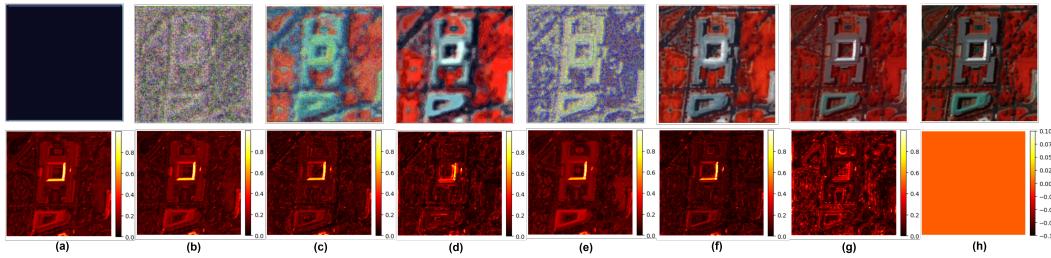


Figure 5: Qualitative comparison of our model against other binary methods on reduced-resolution WDC dataset. (a) BNN, (b) FabNet, (c) ReactNet, (d) ReSTE, (e) BTM, (f) Bi-Diffs, (g) HS-BiNet, and (h) Ground-truth.

Table 2: Full-resolution evaluation on the FR1 dataset with binary methods.

Method	$D_\lambda$	$D_s$	QNR
BNN	0.093	0.062	0.849
ReActNet	0.318	0.009	0.675
BTM	0.117	0.041	0.846
ReSTE	0.405	<b>0.001</b>	0.594
Bi-Diffs	0.068	0.083	0.853
<b>HS-BiNet</b>	<b>0.037</b>	0.104	<b>0.861</b>
Ideal values	0	0	1.00

ing strong capability despite binarization. The qualitative results in Figures 4 and 5 further support these findings. RGB visualizations exhibit clearer structures and more accurate colors, while the MAE maps show fewer bright regions, indicating closer agreement with the ground truth. To test generalization in real-world conditions, we also evaluate full-resolution FR1 data from the PRISMA pansharpening contest. As shown in Table 2, our method achieves state-of-the-art results, producing the highest QNR values and competitive distortion measures ( $D_\lambda$ ,  $D_s$ ). It preserves both spectral and spatial details well and provides the efficiency benefits of binarization. Additional analysis of spectral vectors from a WDC sample is presented in APPENDIX A.2.4, where our method shows noticeably lower spectral distortion than other binary networks.

## 4.2 ABLATION: EXPERIMENTS ON STE

To demonstrate the effectiveness of our method, we conduct a comprehensive performance study on the CIFAR-10 dataset, benchmarking against a broad range of state-of-the-art (SOTA) binary networks. Unlike many existing approaches that rely on auxiliary components such as additional modules or loss functions, our method introduces a learnable straight-through estimator that simultaneously adapts both the forward surrogate and its backwards gradient, without requiring any auxiliary modules or losses. For clarity, we mark in Table 3 the methods that depend on such auxiliaries with an asterisk (\*). We evaluate three widely adopted backbone architectures, ResNet-18, ResNet-20, and VGG-Small, against diverse SOTA binary methods, including LBA Hou et al. (2016), RAD Ding et al. (2019), DSQ Gong et al. (2019a), Xnor-Net Rastegari et al. (2016a), DoReFa-Net Zhou et al. (2016), IR-Net Qin et al. (2020b), LCR-BNN Shang et al. (2022), RBNN Lin et al. (2020), and FDA Xu et al. (2021a).

As shown in Table 3, our approach consistently achieves the highest Top-1 accuracy across all backbones. For example, with ResNet-18, our method outperforms RBNN\* and IR-Net\*, both of which rely on auxiliary designs. On ResNet-20, our method surpasses the strongest baseline RBNN\*. Similarly, with VGG-Small, our approach outperforms RBNN\* and other competing methods. These results highlight the superiority of our method in that, despite eliminating the need for auxiliary modules or loss functions, it consistently outperforms prior SOTA approaches that rely on such components, demonstrating both greater effectiveness and efficiency.

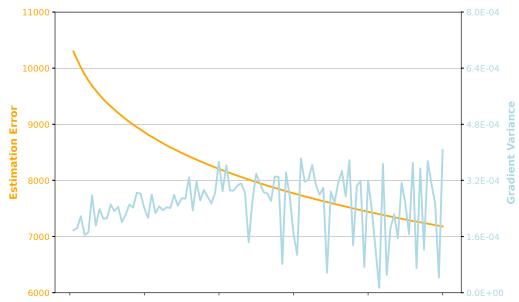


Figure 6: Estimating error (orange curve) and gradient instability (blue curve) indicators on CIFAR-10.

432 Table 3: Performance comparison with state-of-the-art methods on CIFAR-10. Methods marked  
 433 with \* indicate the use of auxiliary components (either additional modules or loss functions).

Method	Top-1 Acc.	Method	Top-1 Acc.	Method	Top-1 Acc.
Full Precision	94.84	Full Precision	91.70	Full Precision	93.33
DSQ	84.11	DoReFa-Net	84.44	LBA	87.70
RAD*	90.50	IR-Net*	85.40	Xnor-Net	89.80
IR-Net*	91.50	SLB*	85.50	BNN	89.90
LCR-BNN*	91.80	LCR-BNN*	86.00	RAD*	90.00
RBNN*	92.20	FDA*	86.20	IR-Net*	90.40
<b>ATISTE</b>	<b>93.00</b>	<b>ATISTE</b>	<b>87.45</b>	<b>RBNN*</b>	<b>91.30</b>

442 ResNet-18 on CIFAR-10

442 ResNet-20 on CIFAR-10

442 VGG-Small on CIFAR-10

444 Table 4: Benchmark comparison of model complexity and inference performance. All parameter  
 445 counts are reported in Millions (M).

Model	Params (M)	FLOPs (G)	Inference (s)
HyperPNN	0.514	7.883	0.619
HspeNet	3.264	52.935	0.620
DHP-Darn	9.067	12.455	1.074
HS-BiNet + ATISTE	1.60	5.599	0.565
HS-BiNet + BNN	1.60	5.599	0.628
HS-BiNet + Bi-DiffSR	1.60	11.010	0.630
HS-BiNet + BTM	1.60	5.599	0.585
HS-BiNet + ReActNet	1.61	5.616	0.624
HS-BiNet + ReSTE	1.60	5.599	0.592

#### 455 4.2.1 ANALYSIS ON ESTIMATING ERROR AND GRADIENT STABILITY

457 Fig. 6 shows a steady decrease in estimation error during training, indicating that the proposed  
 458 estimator progressively aligns with the sign function and improves forward approximation. At the  
 459 same time, the gradient variance remains low with only minor fluctuations, demonstrating stable  
 460 and reliable backward gradients. Together, these trends confirm that the estimator achieves accurate  
 461 forward approximation while preserving training stability, directly supporting the model’s strong  
 462 performance.

#### 463 4.3 ABLATION: COMPUTATIONAL EFFICIENCY

465 Table 4 highlights the efficiency of our approach. Across all integration variants, HS-BiNet main-  
 466 tains a lightweight profile with nearly identical parameter counts and FLOPs, while achieving faster  
 467 or comparable inference times. This demonstrates that HS-BiNet introduces no significant com-  
 468 putational overhead. The HS-BiNet + BNN configuration isolates the effect of our architecture  
 469 within a standard BNN framework, confirming that the observed improvements stem directly from  
 470 HS-BiNet.

## 471 5 CONCLUSION

473 In this paper, we present the first dedicated exploration of binary neural networks for hyperspectral  
 474 pan-sharpening and introduce HS-BiNet, a lightweight architecture driven by the proposed Adaptive  
 475 Tan Identity Straight-Through Estimator (ATISTE). Specifically, ATISTE employs a dual-path sur-  
 476 rogated formulation that decouples forward approximation from gradient propagation, ensuring stable  
 477 optimization while reducing information loss. An adaptive channel-wise scaling mechanism is fur-  
 478 ther incorporated to align binary activations with their real-valued counterparts, thereby preserving  
 479 spectral-spatial fidelity. Building on this estimator, HS-BiNet integrates residual connections, multi-  
 480 scale receptive fields, and local fusion units to capture the complex dependencies of hyperspectral  
 481 data within an efficient binary framework. Unlike many recent models, HS-BiNet avoids compu-  
 482 tationally intensive operations such as unfolding inference and non-local self-attention, making it  
 483 well-suited for real-time deployment on edge and satellite platforms. With these functional designs,  
 484 our approach consistently surpasses conventional binary baselines and narrows, or even closes, the  
 485 gap with full-precision models, demonstrating its practicality for high-fidelity hyperspectral recon-  
 486 struction on resource-constrained platforms. Upon acceptance, we will release the full source code,  
 487 trained models, and data-processing scripts to facilitate reproducibility and further research.

486 REFERENCES  
487

- 488 Bruno Aiazzi, Luciano Alparone, Stefano Baronti, Andrea Garzelli, and Massimo Selva. Mtf-  
489 tailored multiscale fusion of high-resolution ms and pan imagery. *Photogrammetric Engineering  
& Remote Sensing*, 72(5):591–596, 2006.
- 490
- 491 Bruno Aiazzi, Stefano Baronti, and Massimo Selva. Improving component substitution pansharpening  
492 through multivariate regression of ms + pan data. *IEEE Transactions on Geoscience and  
493 Remote Sensing*, 45(10):3230–3239, 2007.
- 494
- 495 Wele Gedara Chaminda Bandara and Vishal M. Patel. Hypertransformer: A textural and spectral  
496 feature fusion transformer for pansharpening. In *Proceedings of the IEEE/CVF Conference on  
497 Computer Vision and Pattern Recognition (CVPR)*, pp. 1767–1777, June 2022.
- 498
- 499 Wele Gedara Chaminda Bandara, Jeya Maria Jose Valanarasu, and Vishal M Patel. Hyperspectral  
500 pansharpening based on improved deep image prior and residual reconstruction. *IEEE Transac-  
501 tions on Geoscience and Remote Sensing*, 60:1–16, 2021.
- 502
- 503 Enton Bedini. The use of hyperspectral remote sensing for mineral exploration: a review. *Journal  
of Hyperspectral Remote Sensing*, 7:189, 12 2017. doi: 10.29150/jhrs.v7.4.p189-211.
- 504
- 505 Jiajun Cai and Bo Huang. Super-resolution-guided progressive pansharpening based on a deep  
506 convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):  
5206–5220, 2020.
- 507
- 508 Yuanhao Cai, Yuxin Zheng, Jing Lin, Xin Yuan, Yulun Zhang, and Haoqian Wang. Binarized  
509 spectral compressive imaging. *Advances in Neural Information Processing Systems*, 36:38335–  
38346, 2023.
- 510
- 511 Thyago P Carvalho, Fabrizzio AAMN Soares, Roberto Vita, Roberto da P Francisco, João P Basto,  
512 and Symone GS Alcalá. A systematic literature review of machine learning methods applied to  
513 predictive maintenance. *Computers & Industrial Engineering*, 137:106024, 2019.
- 514
- 515 Zheng Chen, Haotong Qin, Yong Guo, Xiongfei Su, Xin Yuan, Linghe Kong, and Yulun Zhang. Bi-  
516 narinized diffusion model for image super-resolution. *Advances in Neural Information Processing  
517 Systems*, 37:30651–30669, 2024.
- 518
- 519 Jungwook Choi, Pierce I-Jen Chuang, Zhuo Wang, Swagath Venkataramani, Vijayalakshmi Srin-  
520 ivasan, and Kailash Gopalakrishnan. Bridging the accuracy gap for 2-bit quantized neural net-  
521 works (qnn), 2018. URL <https://arxiv.org/abs/1807.06964>.
- 522
- 523 Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable  
524 convolutional networks. In *Proceedings of the IEEE international conference on computer vision*,  
525 pp. 764–773, 2017.
- 526
- 527 Liang-Jian Deng, Gemine Vivone, Cheng Jin, and Jocelyn Chanussot. Detail injection-based deep  
528 convolutional neural networks for pansharpening. *IEEE Transactions on Geoscience and Remote  
529 Sensing*, 59(8):6995–7010, 2020.
- 530
- 531 Ruizhou Ding, Ting-Wu Chin, Zeye Liu, and Diana Marculescu. Regularizing activation distribution  
532 for training binarized deep networks. In *Proceedings of the IEEE/CVF Conference on Computer  
533 Vision and Pattern Recognition (CVPR)*, June 2019.
- 534
- 535 Wenqian Dong, Tongzhen Zhang, Jiahui Qu, Song Xiao, Jie Liang, and Yunsong Li. Laplacian  
536 pyramid dense network for hyperspectral pansharpening. *IEEE Transactions on Geoscience and  
537 Remote Sensing*, 60:1–13, 2021.
- 538
- 539 Wenqian Dong, Yihan Yang, Jiahui Qu, Yunsong Li, Yufei Yang, and Xiuping Jia. Feature pyramid  
540 fusion network for hyperspectral pansharpening. *IEEE Transactions on Neural Networks and  
541 Learning Systems*, 2023.
- 542
- 543 Yang Feng, Binyu Yan, Seunggil Jeon, and Xiaomin Yang. A hyperspectral pansharpening method  
544 using retrain transformer network for remote sensing images in uav communications system.  
545 *Wireless Networks*, pp. 1–14, 2024.

- 540 Ruihao Gong, Xianglong Liu, Shenghu Jiang, Tianxiang Li, Peng Hu, Jiazen Lin, Fengwei Yu, and  
 541 Junjie Yan. Differentiable soft quantization: Bridging full-precision and low-bit neural networks.  
 542 In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October  
 543 2019a.
- 544 Ruihao Gong, Xianglong Liu, Shenghu Jiang, Tianxiang Li, Peng Hu, Jiazen Lin, Fengwei Yu, and  
 545 Junjie Yan. Differentiable soft quantization: Bridging full-precision and low-bit neural networks.  
 546 In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4852–4861,  
 547 2019b.
- 548 Peiyan Guan and Edmund Y Lam. Multistage dual-attention guided fusion network for hyperspectral  
 549 pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021.
- 550 Mohamed Fadhlallah Guerri, Cosimo Distante, Paolo Spagnolo, Fares Bougourzi, and Abdelmalik  
 551 Taleb-Ahmed. Deep learning techniques for hyperspectral image analysis in agriculture: A re-  
 552 view. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 12:100062, 2024. ISSN  
 553 2667-3932. doi: <https://doi.org/10.1016/j.phphoto.2024.100062>.
- 554 Lin He, Jiawei Zhu, Jun Li, Antonio Plaza, Jocelyn Chanussot, and Bo Li. Hyperpnn: Hyperspectral  
 555 pansharpening via spectrally predictive convolutional neural networks. *IEEE Journal of Selected  
 556 Topics in Applied Earth Observations and Remote Sensing*, 12(8):3092–3100, 2019. doi: 10.  
 557 1109/JSTARS.2019.2917584.
- 558 Lin He, Hanghui Ye, Dahan Xi, Jun Li, Antonio Plaza, and Mei Zhang. Tree-structured neural  
 559 network for hyperspectral pansharpening. *IEEE Journal of Selected Topics in Applied Earth  
 560 Observations and Remote Sensing*, 2023.
- 561 Lin He, Zhou Fang, Jun Li, Jocelyn Chanussot, and Antonio Plaza. Two spectral-spatial implicit  
 562 neural representations for arbitrary-resolution hyperspectral pansharpening. *IEEE Transactions  
 563 on Geoscience and Remote Sensing*, 2024.
- 564 Junming Hou, Xiaoyu Chen, Ran Ran, Xiaofeng Cong, Xinyang Liu, Jian Wei You, and Liang-Jian  
 565 Deng. Binarized neural network for multi-spectral image fusion. In *Proceedings of the IEEE/CVF  
 566 Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2236–2245, June 2025.
- 567 Lu Hou, Quanming Yao, and James T Kwok. Loss-aware binarization of deep networks. *arXiv  
 568 preprint arXiv:1611.01600*, 2016.
- 569 Meiqi Hu, Chen Wu, and Liangpei Zhang. Hypernet: Self-supervised hyperspectral spatial-spectral  
 570 feature understanding network for hyperspectral change detection. *IEEE Transactions on Geo-  
 571 science and Remote Sensing*, 60:1–17, 2022. doi: 10.1109/TGRS.2022.3218795.
- 572 Itay Hubara, Matthieu Courbariaux, Daniel Soudry, Ran El-Yaniv, and Yoshua Bengio. Binarized  
 573 neural networks. *Advances in neural information processing systems*, 29, 2016.
- 574 Xinrui Jiang, Nannan Wang, Jingwei Xin, Keyu Li, Xi Yang, and Xinbo Gao. Training binary  
 575 neural network without batch normalization for image super-resolution. In *Proceedings of the  
 576 AAAI Conference on Artificial Intelligence*, volume 35, pp. 1700–1707, 2021.
- 577 Xinrui Jiang, Nannan Wang, Jingwei Xin, Keyu Li, Xi Yang, Jie Li, Xiaoyu Wang, and Xinbo Gao.  
 578 Fabnet: Frequency-aware binarized network for single image super-resolution. *IEEE Transac-  
 579 tions on Image Processing*, 32:6234–6247, 2023.
- 580 Hyungjun Kim, Kyungsu Kim, Jinseok Kim, and Jae-Joon Kim. Binaryduo: Reducing gra-  
 581 dient mismatch in binary activation network by coupling binary activations. *arXiv preprint  
 582 arXiv:2002.06517*, 2020.
- 583 Mingbao Lin, Rongrong Ji, Zihan Xu, Baochang Zhang, Yan Wang, Yongjian Wu, Feiyue Huang,  
 584 and Chia-Wen Lin. Rotated binary neural network. In H. Larochelle, M. Ranzato, R. Hadsell,  
 585 M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33,  
 586 pp. 7474–7485. Curran Associates, Inc., 2020.

- 594 Zechun Liu, Zhiqiang Shen, Marios Savvides, and Kwang-Ting Cheng. Reactnet: Towards precise  
 595 binary neural network with generalized activation functions. In *European conference on computer*  
 596 *vision*, pp. 143–159. Springer, 2020.
- 597 Sergio Pallas Enguita, Chung-Hao Chen, and Samuel Kovacic. A review of emerging sensor tech-  
 598 nologies for tank inspection: A focus on lidar and hyperspectral imaging and their automation and  
 599 deployment. *Electronics*, 13(23), 2024. ISSN 2079-9292. doi: 10.3390/electronics13234850.
- 600 Siran Peng, Xiangyu Zhu, Haoyu Deng, Liang-Jian Deng, and Zhen Lei. Fusionmamba: Efficient  
 601 remote sensing image fusion with state space model. *IEEE Transactions on Geoscience and*  
 602 *Remote Sensing*, 2024.
- 603 Haotong Qin, Ruihao Gong, Xianglong Liu, Xiao Bai, Jingkuan Song, and Nicu Sebe. Binary neural  
 604 networks: A survey. *Pattern Recognition*, 105:107281, September 2020a. ISSN 0031-3203.  
 605 doi: 10.1016/j.patcog.2020.107281. URL <http://dx.doi.org/10.1016/j.patcog.2020.107281>.
- 606 Haotong Qin, Ruihao Gong, Xianglong Liu, Mingzhu Shen, Ziran Wei, Fengwei Yu, and Jingkuan  
 607 Song. Forward and backward information retention for accurate binary neural networks. In  
 608 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,  
 609 June 2020b.
- 610 Haotong Qin, Ruihao Gong, Xianglong Liu, Mingzhu Shen, Ziran Wei, Fengwei Yu, and Jingkuan  
 611 Song. Forward and backward information retention for accurate binary neural networks. In  
 612 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2250–  
 613 2259, 2020c.
- 614 Billy G. Ram, Peter Oduor, C. Igathinathane, Kirk Howatt, and Xin Sun. A systematic review of  
 615 hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects.  
 616 *Computers and Electronics in Agriculture*, 222:109037, 2024. ISSN 0168-1699. doi: <https://doi.org/10.1016/j.compag.2024.109037>.
- 617 Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, and Ali Farhadi. Xnor-net: Imagenet  
 618 classification using binary convolutional neural networks. In Bastian Leibe, Jiri Matas, Nicu Sebe,  
 619 and Max Welling (eds.), *Computer Vision – ECCV 2016*, pp. 525–542, Cham, 2016a. Springer  
 620 International Publishing. ISBN 978-3-319-46493-0.
- 621 Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, and Ali Farhadi. Xnor-net: Imagenet  
 622 classification using binary convolutional neural networks. In *European conference on computer*  
 623 *vision*, pp. 525–542. Springer, 2016b.
- 624 Yuzhang Shang, Dan Xu, Bin Duan, Ziliang Zong, Liqiang Nie, and Yan Yan. Lipschitz conti-  
 625 nuity retained binary neural network. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Gio-  
 626 vanni Maria Farinella, and Tal Hassner (eds.), *Computer Vision – ECCV 2022*, pp. 603–619,  
 627 Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-20083-0.
- 628 Rohit Sharma, Sachin S Kamble, Angappa Gunasekaran, Vikas Kumar, and Anil Kumar. A system-  
 629 atic literature review on machine learning applications for sustainable agriculture supply chain  
 630 performance. *Computers & Operations Research*, 119:104926, 2020.
- 631 Miguel Simoes, José Bioucas-Dias, Luis B Almeida, and Jocelyn Chanussot. A convex formulation  
 632 for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on*  
 633 *Geoscience and Remote Sensing*, 53(6):3373–3388, 2014.
- 634 Chongxing Song, Zhiqiang Lang, Wei Wei, and Lei Zhang. E2fif: Push the limit of binarized deep  
 635 imagery super-resolution using end-to-end full-precision information flow. *IEEE Transactions on*  
 636 *Image Processing*, 32:5379–5393, 2023.
- 637 Mary B. Stuart, Andrew J. S. McGonigle, and Jon R. Willmott. Hyperspectral imaging in environ-  
 638 mental monitoring: A review of recent developments and technological advances in compact field  
 639 deployable systems. *Sensors*, 19(14), 2019. ISSN 1424-8220. doi: 10.3390/s19143071.
- 640 Qian Sun, Yu Sun, and Chengsheng Pan. Aidb-net: An attention-interactive dual-branch convolu-  
 641 tional neural network for hyperspectral pansharpening. *Remote Sensing*, 16(6):1044, 2024.
- 642

- 648 Xuebin Tang, Ke Zhang, Xiaolei Zhou, Zeng Lingbin, and Shan Huang. Enhancing binary convolutional  
 649 neural networks for hyperspectral image classification. *Remote Sensing*, 16:4398, 11 2024.  
 650 doi: 10.3390/rs16234398.
- 651 Xiande Wu, Jie Feng, Ronghua Shang, Xiangrong Zhang, and Licheng Jiao. Multiobjective guided  
 652 divide-and-conquer network for hyperspectral pansharpening. *IEEE Transactions on Geoscience  
 653 and Remote Sensing*, 60:1–17, 2022.
- 654 Xiao-Ming Wu, Dian Zheng, Zuhao Liu, and Wei-Shi Zheng. Estimator meets equilibrium perspective:  
 655 A rectified straight through estimator for binary neural networks training. In *Proceedings of  
 656 the IEEE/CVF International Conference on Computer Vision*, pp. 17055–17064, 2023.
- 657 Bin Xia, Yulun Zhang, Yitong Wang, Yapeng Tian, Wenming Yang, Radu Timofte, and Luc  
 658 Van Gool. Basic binary convolution unit for binarized image restoration network. *arXiv preprint  
 659 arXiv:2210.00405*, 2022.
- 660 Jin-Liang Xiao, Ting-Zhu Huang, Liang-Jian Deng, Guang Lin, Zihan Cao, Chao Li, and Qibin  
 661 Zhao. Hyperspectral pansharpening via diffusion models with iteratively zero-shot guidance.  
 662 In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12669–12678,  
 663 2025.
- 664 Jingwei Xin, Nannan Wang, Xinrui Jiang, Jie Li, Heng Huang, and Xinbo Gao. Binarized neural  
 665 network for single image super resolution. In *European conference on computer vision*, pp. 91–  
 666 107. Springer, 2020.
- 667 Yixing Xu, Kai Han, Chang Xu, Yehui Tang, Chunjing XU, and Yunhe Wang. Learning frequency  
 668 domain approximation for binary neural networks. In M. Ranzato, A. Beygelzimer, Y. Dauphin,  
 669 P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*,  
 670 volume 34, pp. 25553–25565. Curran Associates, Inc., 2021a.
- 671 Zihan Xu, Mingbao Lin, Jianzhuang Liu, Jie Chen, Ling Shao, Yue Gao, Yonghong Tian, and  
 672 Rongrong Ji. Recu: Reviving the dead weights in binary neural networks. In *Proceedings of  
 673 the IEEE/CVF international conference on computer vision*, pp. 5198–5208, 2021b.
- 674 Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. Pannet: A  
 675 deep network architecture for pan-sharpening. In *Proceedings of the IEEE international confer-  
 676 ence on computer vision*, pp. 5449–5457, 2017.
- 677 Penghang Yin, Jiancheng Lyu, Shuai Zhang, Stanley Osher, Yingyong Qi, and Jack Xin. Un-  
 678 derstanding straight-through estimator in training activation quantized neural nets, 2019. URL  
 679 <https://arxiv.org/abs/1903.05662>.
- 680 Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled non-negative matrix factorization (cnmf)  
 681 for hyperspectral and multispectral data fusion: Application to pasture classification. In *2011  
 682 IEEE international geoscience and remote sensing symposium*, pp. 1779–1782. IEEE, 2011.
- 683 Yuxuan Zheng, Jiaojiao Li, Yunsong Li, Jie Guo, Xianyun Wu, and Jocelyn Chanussot. Hyperspec-  
 684 tral pansharpening using deep prior and dual attention residual network. *IEEE Transactions on  
 685 Geoscience and Remote Sensing*, 58(11):8059–8076, 2020. doi: 10.1109/TGRS.2020.2986313.
- 686 Shuchang Zhou, Yuxin Wu, Zekun Ni, Xinyu Zhou, He Wen, and Yuheng Zou. Dorefa-net: Train-  
 687 ing low bitwidth convolutional neural networks with low bitwidth gradients. *arXiv preprint  
 688 arXiv:1606.06160*, 2016.
- 689 Yu-Wei Zhuo, Tian-Jing Zhang, Jin-Fan Hu, Hong-Xia Dou, Ting-Zhu Huang, and Liang-Jian Deng.  
 690 A deep-shallow fusion network with multidetail extractor and spectral attention for hyperspectral  
 691 pansharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote  
 692 Sensing*, 15:7539–7555, 2022.
- 693
- 694
- 695
- 696
- 697
- 698
- 699
- 700
- 701

702  
703 A APPENDIX704  
705 A.1 BINARY CONVOLUTION706  
707 In model binarization, the surrogate estimator plays a central role in approximating the discontinuous  
708 sign function for both weights and activations. However, conventional binary convolutional layers  
709 rely on fixed approximations that either over-restrict the representation capacity or destabilize the  
710 training process. To address this limitation, we design a customized binary convolutional module,  
711 that incorporates channel-wise learnable sharpness parameters and a residual mixing coefficient to  
712 flexibly control the trade-off between binarization fidelity and gradient stability.713  
714 Specifically, each output channel of the convolutional weights is associated with a learnable sharp-  
715 ness parameter  $\alpha_w$ , while each input channel of the activations has a corresponding sharpness pa-  
716 rameter  $\alpha_a$ . These parameters determine how aggressively the weights and activations are driven  
717 towards their binarized representations. In addition, a residual mixing factor  $\lambda$  is introduced and  
718 made learnable at the layer level. This factor is constrained within  $[0, 1]$  to guarantee stability, and  
719 it governs the convex combination of the nonlinear binarization path and the linear passthrough  
720 path. In this way, the model adaptively balances strict binary quantization against the retention of  
721 full-precision information.722  
723 During the forward pass, both weights and activations are binarized using the proposed ATISTE  
724 estimator, which approximates the hard sign function in the forward computation while supplying  
725 bounded surrogate gradients in the backward pass. By reshaping  $\alpha_w$  and  $\alpha_a$  for proper broadcasting,  
726 the binarization process is applied in a channel-wise manner, ensuring fine-grained control. After  
727 binarization, a per-channel scaling factor is computed as the mean absolute value of the real-valued  
728 weights. This scaling restores the representational power lost through 1-bit quantization and prevents  
729 performance degradation that commonly arises in fully binarized networks.730  
731 Finally, the binarized activations and scaled binarized weights are used in a standard convolution op-  
732 eration, inheriting stride, padding, dilation, and grouping from the base full-precision layer. Through  
733 this design, the BinaryConv2d module ensures that both weights and activations are quantized with  
734 stability-aware surrogates, while maintaining sufficient representational capacity to achieve compet-  
735 itive performance in binarized neural networks.736  
737 A.2 EXPERIMENTAL SETTINGS738  
739 To evaluate the effectiveness of our proposed HS-BiNet for remote sensing pansharpening, we con-  
740 duct extensive experiments on three simulated hyperspectral (HS) datasets<sup>1</sup>, i.e., Washington DC,  
741 and Botswana, as well as one full-resolution dataset, FR1. The key characteristics of these datasets  
742 are summarized in Table 5 for convenient comparison.743  
744 1. Washington DC Mall Dataset745  
746 This dataset was collected by the Hyperspectral Digital Imagery Collection Experiment  
747 (HYDICE) sensor. It originally contains 210 spectral bands covering the range from 0.4  
748 to  $2.4 \mu\text{m}$  (visible light and near-infrared). After removing unusable bands, 191 bands are  
749 retained. The spatial resolution of the data is  $1208 \times 307$  pixels. The dataset covers various  
750 land cover types, including roofs, streets, gravel roads, grass, trees, water, and shadows. In  
751 experimental setups, the Washington DC Mall (WDC) dataset with 191 channels is com-  
752 monly used. The test set is constructed from four  $128 \times 128$  images clipped from the  
753 original image, while the remaining portion is utilized for training the network parameters.  
754 For training, the original PAN and HS images are partitioned into 921 small patch pairs  
755 consisting of  $64 \times 64$  PAN patches and  $16 \times 16$  HS patches. Additionally, 103 patch pairs  
756 are reserved for validation from the simulated patches.757  
758 2. Botswana Dataset759  
760 Captured by the EO-1 Hyperion sensor over Botswana between 2001 and 2004, this dataset  
761 spans a spectral range from 0.4 to  $2.5 \mu\text{m}$  at 10 nm intervals. Originally comprising 242  
762 spectral bands, 145 bands are preserved after noise removal. The spatial size is  $1496 \times 256$   
763 pixels. The dataset includes 14 distinct land cover classes, representing seasonal swamps,764  
765 <sup>1</sup>Dataset source: <https://github.com/liangjiandeng/HyperPanCollection>

756  
757  
Table 5: Hyperspectral Datasets  
758  
759  
760

Dataset	Number of Bands	Spectral Range	Spatial Resolution	Image Size	Land Cover Types
Washington DC	191	0.4–2.4 $\mu\text{m}$	1 m	1208 $\times$ 307	Roofs, Streets
Botswana	145	0.4–2.5 $\mu\text{m}$	30 m	1496 $\times$ 256	Seasonal Swamps
FR1	69	0.4–2.5 $\mu\text{m}$	30 m	2400 $\times$ 2400	Roofs, Streets

761  
762  
763 occasional swamps, and drier woodlands within the distal portion of the Okavango Delta.  
764 In practice, the Botswana dataset is often prepared with 102 effective channels. The test set  
765 is composed of four  $128 \times 128$  image clips from the original data. For training, the PAN  
766 and HS images are divided into 799 overlapping  $64 \times 64$  patches, and 168 patch pairs are  
767 retained for simulation.

### 768 3. FR1 Full-Resolution Dataset

769 Distributed as part of the PRISMA contest, the FR1 dataset is intended for full spatial res-  
770 olution pansharpening. The dataset can be downloaded from the official PRISMA website.  
771 It contains co-registered PAN and HS images, where a  $12 \text{ km} \times 12 \text{ km}$  region (equivalent to  
772  $2400 \times 2400$  pixels for the PAN image and  $400 \times 400 \times 69$  pixels for the HS image) is ex-  
773 tracted from the original  $30 \text{ km} \times 30 \text{ km}$  PRISMA acquisition. In experimental settings, the  
774 FR1 dataset with 69 channels is employed to evaluate pansharpening performance. The test  
775 set consists of two images ( $240 \times 240$  for HS and  $60 \times 60$  for PAN) clipped from the original  
776 image, while the remaining portion is used for training after the downsampling simulation.  
777 The training data are divided into 734 small patch pairs, each comprising  $60 \times 60$  PAN  
778 patches and  $10 \times 10$  HS patches, with 82 patch pairs reserved for validation.

779 Following Wald’s protocol, the original HS images from the three simulated datasets serve as the  
780 reference (REF) images. Low-resolution hyperspectral (LRHS) images are generated by first ap-  
781 plying a Gaussian blur to the reference images, followed by downsampling, where one out of every  
782 four pixels is selected in both horizontal and vertical directions. Simulated PAN images are created  
783 by applying a suitable spectral response vector to the original reference HS images. The resulting  
784 simulated PAN and LRHS images are then used as inputs to various hyperspectral super-resolution  
785 methods, including our proposed HS-BiNet. We implemented and trained our model using the Py-  
786 Torch framework on an NVIDIA GeForce RTX A5000 GPU with 24 GB of memory. The training  
787 employed a batch size of 4, the Adam optimizer, and a cosine annealing warm restarts scheduler.  
788 The model was trained for 1,600 epochs on the Washington DC Mall dataset and for 1,000 epochs  
789 each on the Botswana and FR1 datasets.

790 Finally, the super-resolved HS images are quantitatively compared against the original ground-truth  
791 HS images using standard quality metrics. To assess the quality of the proposed pansharpening  
792 method, we adopt a set of widely accepted spatial and spectral evaluation measures commonly used  
793 in hyperspectral image fusion tasks. Specifically, we evaluate performance using PSNR, SSIM,  
794 Cross-Correlation (CC), Spectral Angle Mapping (SAM), and ERGAS, as these indices collectively  
795 characterize both spatial detail preservation and spectral consistency. The ideal reference values for  
796 these metrics are:  $\text{PSNR} \rightarrow \infty$ ,  $\text{SSIM} \rightarrow 1$ ,  $\text{CC} \rightarrow 1$ ,  $\text{SAM} \rightarrow 0$ , and  $\text{ERGAS} \rightarrow 0$ , indicating  
797 perfect reconstruction fidelity He et al. (2019); Hu et al. (2022); Deng et al. (2020); Zhuo et al.  
798 (2022); Dong et al. (2023). In hyperspectral fusion, such a diverse set of metrics is essential because  
799 a method must simultaneously reproduce high-frequency spatial structures and maintain accurate  
800 spectral signatures across all bands. PSNR measures overall pixel-wise reconstruction accuracy by  
801 evaluating the logarithmic ratio between the signal and the mean squared error; higher values denote  
802 lower distortion, although PSNR alone cannot fully capture perceptual or spectral differences. CC  
803 complements PSNR by quantifying the linear correlation between spatial patterns of the fused im-  
804 age and the ground truth, thereby reflecting how well spatial structures are retained. SSIM provides  
805 a more perceptually grounded spatial assessment by examining luminance, contrast, and structural  
806 similarity between corresponding patches, offering insight into the preservation of edges and tex-  
807 tures, even though it does not inherently capture spectral distortions. To address spectral fidelity,  
808 SAM evaluates the angle between spectral vectors at each pixel, with smaller angles indicating that  
809 the fused image preserves the correct spectral shape relative to the ground truth, an essential re-  
810 quirement for tasks such as material classification and unmixing. Finally, ERGAS offers a global,  
811 scale-adjusted summary of relative reconstruction error across all spectral bands, integrating both  
812 spatial and spectral discrepancies into a single dimensionless score. Together, these complementary

810 metrics ensure a holistic and reliable evaluation of fusion quality, capturing the aspects of spatial  
 811 sharpness, structural coherence, and spectral integrity that are fundamental to effective hyperspectral  
 812 pansharpening.

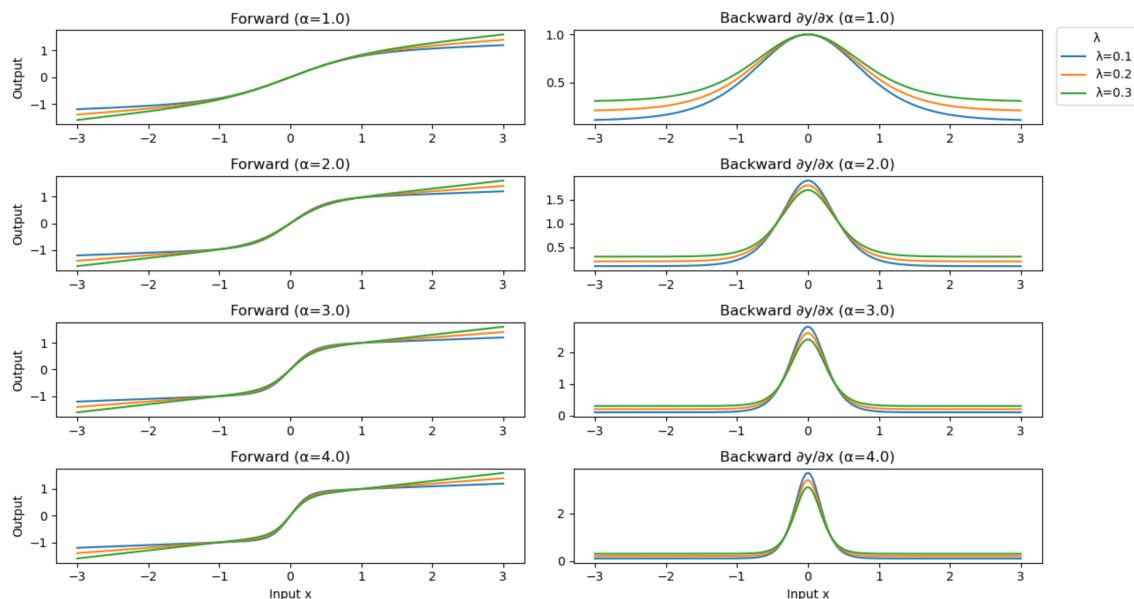
813 **Reconstruction Loss.** We adopt the  $\ell_1$  loss to supervise the reconstruction of high-resolution hy-  
 814 perspectral images. It is defined as

$$816 \quad \mathcal{L}_1(x, x_{\text{gt}}) = \|x - x_{\text{gt}}\|_1, \quad (7)$$

817 where  $x_{\text{gt}}$  denotes the ground-truth HR-HSI and  $x$  is the reconstructed HR-HSI. We choose the  $\ell_1$   
 818 formulation because prior studies have shown that it produces sharper and more faithful reconstruc-  
 819 tions than the  $\ell_2$  loss in hyperspectral pansharpening tasks.

820 The model is optimized using Adam with an initial learning rate of  $1 \times 10^{-3}$  and a weight decay of  
 821  $1 \times 10^{-5}$ , while a cosine-annealing warm-restart schedule ( $T_0 = 50$ ,  $T_{\text{mult}} = 2$ ,  $\eta_{\text{min}} = 1 \times 10^{-5}$ )  
 822 is employed to periodically refresh the learning rate and stabilize training under binary constraints.  
 823 No data augmentation is applied, as spatial or spectral transformations can introduce non-physical  
 824 wavelength distortions that negatively impact hyperspectral fidelity. All binary layers are initialized  
 825 using Kaiming initialization combined with a scaled real-valued weight formulation, which prevents  
 826 early saturation during binarization and promotes stable gradient flow in the initial training stages.  
 827 Furthermore, the entire model is trained from scratch without any pretrained components. Since  
 828 binary networks are particularly sensitive to learning-rate schedules, initialization strategies, and data  
 829 integrity, clearly specifying these design choices ensures consistent and fully reproducible training  
 830 behavior.

831 **A.2.1 ANALYSIS ON  $\alpha$  AND  $\lambda$**



833 Figure 7: Forward and backward processes with  $\alpha = 1.0, 2.0, 3.0, 4.0$  and  $\lambda = 0.1, 0.2, 0.3$ .

834 The results presented in Figure 7 illustrate the influence of the scaling factor  $\alpha$  and the blending co-  
 835 efficient  $\lambda$  on both the forward approximation and the backwards gradient behaviour of the proposed  
 836 ATISTE. In the forward pass,  $\alpha$  regulates the sharpness of the nonlinear surrogate. For smaller values  
 837 of  $\alpha$  (e.g.,  $\alpha = 1$ ), the output exhibits a smooth and gradual transition between  $-1$  and  $+1$ , closely  
 838 resembling a softened activation. As  $\alpha$  increases ( $\alpha = 4$ ), this transition becomes significantly  
 839 steeper, effectively approaching the discontinuous sign function while retaining differentiability. In  
 840 parallel,  $\lambda$  governs the trade-off between the nonlinear surrogate and the identity mapping. Lower  
 841 values of  $\lambda$  emphasize the nonlinear behavior of the tanh-based surrogate. In contrast, higher val-  
 842 ues introduce a stronger linear component, resulting in forward mappings that deviate less from the  
 843 input in the saturation regions and preserve more representational continuity.

The backwards pass reveals a complementary dynamic. The derivative with respect to the input, defined as a convex combination of the tanh gradient and the constant identity gradient, demonstrates that  $\alpha$  primarily controls the localization of the gradient around zero. Specifically, larger values of  $\alpha$  yield sharper, higher-magnitude peaks, enabling stronger learning signals near the decision boundary but diminishing gradient flow elsewhere. In contrast,  $\lambda$  lifts the gradient baseline across the domain, preventing complete gradient vanishing in saturated regions and ensuring stable optimization.

The results demonstrate the dual functionality of ATISTE:  $\alpha$  governs the degree of binarization sharpness, whereas  $\lambda$  regulates the trade-off between expressiveness and trainability. Through the joint effect of these parameters, the estimator achieves a close approximation to binary activations in the forward pass while preserving smooth and reliable gradient propagation in the backwards pass. This property is essential for ensuring convergence and accuracy in binary neural networks.

### A.2.2 PROPERTIES OF ATISTE

**Rationality** A desirable property of any surrogate estimator is *rationality*, meaning that its forward approximation error should be no worse than the baseline identity-based STE. ATISTE satisfies this property by construction. Recall the forward mapping

$$f_{\alpha,\lambda}(x) = (1 - \lambda) \tanh(\alpha x) + \lambda x, \quad (8)$$

and let the target discrete mapping be  $\text{sign}(x)$ . The mean-squared forward estimation error can be written as

$$E(\alpha, \lambda) = \mathbb{E} \left[ (\text{sign}(x) - f_{\alpha,\lambda}(x))^2 \right]. \quad (9)$$

As  $\alpha \rightarrow \infty$ , the nonlinear term converges pointwise to the sign function, yielding

$$\lim_{\alpha \rightarrow \infty} f_{\alpha,\lambda}(x) = (1 - \lambda) \text{sign}(x) + \lambda x. \quad (10)$$

The corresponding asymptotic estimation error becomes

$$\lim_{\alpha \rightarrow \infty} E(\alpha, \lambda) = \lambda^2 \mathbb{E} \left[ (\text{sign}(x) - x)^2 \right] \leq E_{\text{STE}}. \quad (11)$$

Thus, for all  $\lambda < 1$ , ATISTE achieves a strictly lower error than the classic STE, which corresponds to the special case  $\lambda = 1$ . This demonstrates that ATISTE is rational, as it either matches or outperforms identity-based STE in forward approximation fidelity while maintaining stable gradient flow.

**Flexibility** In addition to rationality, ATISTE provides *flexibility* through two independent degrees of freedom: the sharpness parameter  $\alpha$  and the residual weight  $\lambda$ . Unlike prior estimators, such as ReSTE, which have a single parameter that simultaneously controls both forward fidelity and gradient dynamics, ATISTE decouples these roles. Specifically, increasing  $\alpha$  monotonically sharpens the tanh component, driving  $f_{\alpha,\lambda}(x)$  closer to the discrete sign function:

$$\lim_{\alpha \rightarrow \infty} f_{\alpha,\lambda}(x) = (1 - \lambda) \text{sign}(x) + \lambda x. \quad (12)$$

Conversely, adjusting  $\lambda$  modulates the gradient floor:

$$\lambda \leq f'_{\alpha,\lambda}(x) \leq (1 - \lambda)\alpha + \lambda, \quad (13)$$

ensuring that gradients remain non-zero everywhere. A higher  $\lambda$  provides stronger residual gradients and hence greater stability, while a lower  $\lambda$  emphasizes sharper binarization. This independent control over forward fidelity ( $\alpha$ ) and gradient stability ( $\lambda$ ) allows ATISTE to smoothly transition from soft approximations in early epochs to near-hard binarization at convergence, without relying on rigid schedules or heuristic parameter coupling. As a result, ATISTE embodies both rationality, guaranteeing no worse error than STE, and flexibility, enabling adaptive, bounded control of the error-stability trade-off across training.

**Estimation error and Gradient Instability** A binarized neural network (BNN) relies on surrogate estimators to approximate the discontinuous sign function during training. The effectiveness of a surrogate estimator is governed by two fundamental factors: *estimation error* in the forward pass and *gradient stability* in the backward pass.

918 The estimation error quantifies the discrepancy between the ideal binarization (the hard sign function)  
 919 and the chosen surrogate estimator. Formally, for input  $z$ , an estimator  $f(\cdot)$ , and distance metric  
 920  $D(\cdot)$ , the error indicator is defined as

$$921 \quad 922 \quad e = D(\text{sign}(z), f(z)), \quad (14)$$

923 where we adopt the  $L_2$ -norm as the distance metric. Thus, the closer  $f(z)$  is to  $\text{sign}(z)$ , the smaller  
 924 the estimation error.

925 While low estimation error ensures high forward fidelity, it is equally crucial that the surrogate  
 926 produces stable gradients in the backward pass. Instability arises when the variance of gradients  
 927 across parameters is high, which can lead to exploding or vanishing updates. To measure this, we  
 928 define the gradient instability indicator as

$$929 \quad 930 \quad s = \text{var}(|g|), \quad (15)$$

931 where  $g$  denotes the parameter gradients,  $|g|$  is the element-wise absolute magnitude, and  $\text{var}(\cdot)$  de-  
 932 notes the variance operator. This metric captures the divergence in gradient magnitudes, independent  
 933 of direction.

934 A critical aspect of ATISTE lies in its treatment of estimation error and gradient stability, the two  
 935 fundamental axes governing the effectiveness of surrogate estimators in BNNs. From the forward  
 936 perspective, the surrogate is defined as

$$937 \quad 938 \quad f_{\alpha,\lambda}(x) = (1 - \lambda) \tanh(\alpha x) + \lambda x, \quad (16)$$

939 where  $\alpha > 0$  controls the sharpness of the nonlinear path and  $\lambda \in [0, 1]$  introduces a residual linear  
 940 contribution. The estimation error with respect to the ideal sign function can be expressed as

$$942 \quad 943 \quad E(\alpha, \lambda) = \mathbb{E}[(\text{sign}(x) - f_{\alpha,\lambda}(x))^2]. \quad (17)$$

944 As  $\alpha \rightarrow \infty$ , the nonlinear term approaches the hard sign, yielding

$$945 \quad 946 \quad \lim_{\alpha \rightarrow \infty} f_{\alpha,\lambda}(x) = (1 - \lambda) \text{sign}(x) + \lambda x, \quad (18)$$

947 and consequently,

$$948 \quad 949 \quad \lim_{\alpha \rightarrow \infty} E(\alpha, \lambda) = \lambda^2 \mathbb{E}[(\text{sign}(x) - x)^2] \leq E_{\text{STE}}. \quad (19)$$

950 This shows that ATISTE is *rational*: its approximation error is guaranteed to be no worse than that  
 951 of the identity-based STE, and in practice smaller when  $\lambda < 1$ , thereby ensuring that binarization  
 952 fidelity improves as training progresses.

953 From the backward perspective, the pseudo-gradient provided to the optimizer is given by

$$955 \quad 956 \quad f'_{\alpha,\lambda}(x) = (1 - \lambda)\alpha (1 - \tanh^2(\alpha x)) + \lambda. \quad (20)$$

957 Since  $0 \leq 1 - \tanh^2(\alpha x) \leq 1$ , the pseudo-gradient satisfies the uniform bounds

$$958 \quad 959 \quad \lambda \leq f'_{\alpha,\lambda}(x) \leq (1 - \lambda)\alpha + \lambda, \quad (21)$$

960 ensuring both a non-zero gradient floor and a finite ceiling. The lower bound  $\lambda$  prevents gradient  
 961 starvation in saturated regions, while the upper bound  $(1 - \lambda)\alpha + \lambda$  controls variance and avoids ex-  
 962 ploding updates. This boundedness makes ATISTE inherently more stable than classical surrogates  
 963 such as hard sign, identity, or power-based estimators, which often suffer from vanishing gradients  
 964 or uncontrolled variance. By decoupling  $\alpha$  and  $\lambda$ , ATISTE provides independent control of bina-  
 965 rization fidelity and gradient stability, enabling smooth learning dynamics and effective convergence  
 966 in deep BNNs.

### 967 A.2.3 ADDITIONAL QUALITATIVE RESULTS

968 The qualitative results in Figure 8 illustrate that HS-BiNet produces fused hyperspectral images with  
 969 sharper structures and more faithful spectral appearance in the pseudo-color visualizations, while the  
 970 absolute error maps show noticeably lower discrepancies from the ground truth compared to other  
 971 methods.

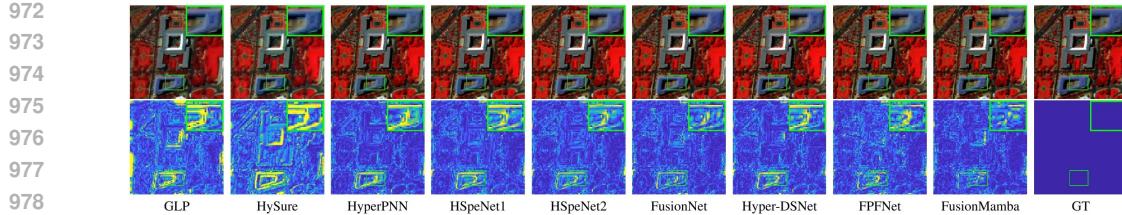


Figure 8: Qualitative evaluation results on the WDC dataset. **Row 1:** Pseudo-color visualizations constructed from spectral bands 20, 50, and 80 of a representative testing sample. **Row 2:** Absolute error maps for spectral band 25 of the same sample, highlighting the difference between the fused result and the ground truth.

Table 6: Component-wise ablation study of HS-BiNet on the test dataset. The impact of removing individual modules on reconstruction performance is reported.

Model Variant	PSNR $\uparrow$	CC $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	ERGAS $\downarrow$
<b>HS-BiNet</b>	<b>37.285</b>	<b>0.900</b>	<b>0.891</b>	<b>1.622</b>	<b>1.852</b>
Without Decoder	36.642	0.893	0.883	1.769	1.917
Without MSFE	36.434	0.885	0.866	1.866	1.912

#### A.2.4 SPECTRAL VECTOR

The spectral vector plots in Figure 9 at positions (30,40), (60,70), and (100,110) indicate that several baseline models produce highly fluctuating and unstable spectral responses. In contrast, the proposed HS-BiNet (red) exhibits smooth, well-scaled reconstructions that remain closely aligned with the ground truth across all spectral bands. Unlike ReActNet and Bi-Diffsr, which produce extreme spikes and unrealistic amplitudes, HS-BiNet avoids such distortions and preserves the natural spectral shape. By effectively zooming, the plots further demonstrate that HS-BiNet captures fine-grained variations within a realistic intensity range, maintaining fidelity to the ground truth without over-amplification. These observations suggest that HS-BiNet achieves more reliable and effective spectral recovery compared to the evaluated methods, effectively balancing reconstruction accuracy and stability.

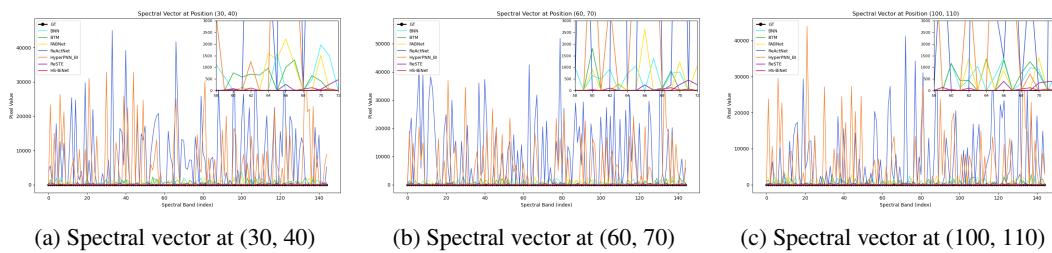
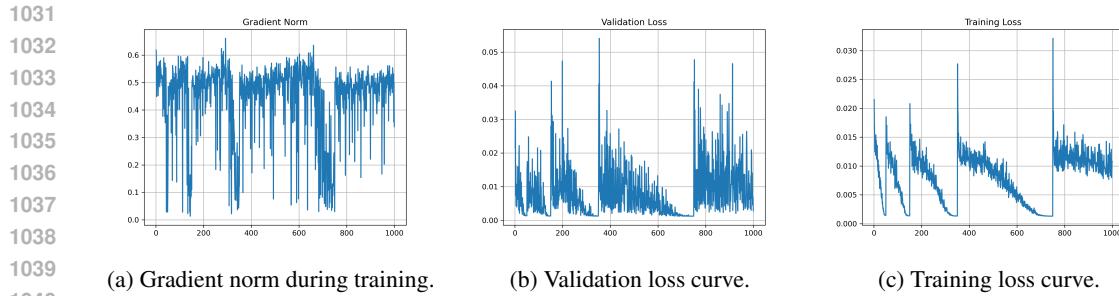


Figure 9: Comparison of three spectral vectors extracted from spatial locations (30, 40), (60, 70), and (100, 110) in a WDC testing sample, illustrating the variability of spectral signatures across the scene..

#### A.2.5 COMPONENT-WISE ABLATION

The results show that every component of HS-BiNet contributes to reconstruction quality. Removing the Decoder causes a significant drop in PSNR, CC, and SSIM, while increasing SAM and ERGAS, indicating reduced spatial and spectral accuracy. Eliminating the MSFE module results in even greater performance degradation, underscoring its crucial role in effective multi-scale feature extraction. The Edge Injector was not removed in the ablation study because directly adding PAN edge information is ineffective: the PAN image has a single channel, whereas hyperspectral features

1026 have multiple channels, resulting in a dimensional mismatch and poor fusion. The Edge Injector, im-  
 1027 plemented using binary convolution, is therefore essential for adjusting the channel dimension and  
 1028 enabling proper edge integration. Overall, the complete HS-BiNet achieves the best performance,  
 1029 confirming the necessity and complementarity of all its components.  
 1030



1031 (a) Gradient norm during training. (b) Validation loss curve. (c) Training loss curve.  
 1032  
 1033  
 1034  
 1035  
 1036  
 1037  
 1038

1039 Figure 10: Training dynamics (left-to-right): the gradient-norm evolution indicating stable updates,  
 1040 the validation-loss trajectory showing cyclic patterns from learning-rate restarts, and the training-  
 1041 loss curve demonstrating consistent within-cycle convergence.  
 1042  
 1043

#### 1044 1045 A.2.6 ANALYSIS OF TRAINING DYNAMICS

1046 As seen in Fig. 10b, the validation loss shows small cycles with short spikes at learning-rate changes,  
 1047 but it quickly returns to low values, meaning the model generalizes well. In Fig. 10c, the training  
 1048 loss follows the same pattern, steadily going down within each phase and only jumping when the  
 1049 learning rate resets, showing that the training is stable. The gradient norm in Fig. 10a stays in a  
 1050 safe range without exploding, which confirms that the updates remain stable throughout training.  
 1051 Overall, the three figures demonstrate that the model trains smoothly, handles learning-rate steps  
 1052 correctly, and maintains control over the gradients.  
 1053  
 1054  
 1055  
 1056  
 1057  
 1058  
 1059  
 1060  
 1061  
 1062  
 1063  
 1064  
 1065  
 1066  
 1067  
 1068  
 1069  
 1070  
 1071  
 1072  
 1073  
 1074  
 1075  
 1076  
 1077  
 1078  
 1079