Super-Resolution of Remote Sensing Images via the Parallel Lattice Attention Network

Allen Patnaik, Arpit Gour, M. K. Bhuyan, Karl F. MacDorman, Sultan Alfarhood, and Mejdl Safran

***Abstract*—Deep learning-based super-resolution models often struggle to balance capturing long-range dependencies with computational efficiency, as convolutional neural networks focus on local features while transformers incur high computational costs. To address this trade-off, we propose a parallel lattice attention network (PLAN) to improve the reconstruction of complex features. Unlike traditional lattice networks that rely on sequential cascading, PLAN introduces a parallel lattice attention block that splits the input feature map into concurrent branches. These branches employ lattice attention units to simultaneously extract fine-grained details and global context, which are subsequently fused via attention-based integration. This parallel design maximizes computational efficiency while ensuring robust reuse of features. Experiments on the AID, UC Merced, and WHU-RS19 datasets at various scales show PLAN's superiority in focusing on critical information to enhance image quality. On the UC Merced dataset with a $2\times$ scaling factor, PLAN achieves a PSNR of 32.12 dB and an SSIM of 0.8807, outperforming state-of-the-art methods such as ESTNet and BMFENet by at least 0.07 dB in PSNR and 0.0210 in SSIM. On the WHU-RS19 dataset with a $4\times$ scaling factor, PLAN achieves a PSNR of 29.16 dB and an SSIM of 0.7597, surpassing previous methods by up to 0.64 dB in PSNR. PLAN also maintains computational efficiency with a runtime of 30.97 ms, faster performance than ESRT (106.34 ms), SwinIR (98.47 ms), FENet (63.52 ms), OmniSR (45.02 ms), BMFENet (51.43 ms), MSCT (57.15 ms), and ESTNet (67.95 ms).**

***Index Terms*—Attention, lattice network, remote-sensing imagery, super-resolution (SR).**

I. Introduction

In computer vision, super-resolution (SR) recovers high-resolution (HR) images from their lower-resolution (LR) counterparts. This technology is essential in remote sensing [1], where precise visual details support environmental monitoring, urban planning, and disaster management applications. Increased image resolution provides finer details and richer information, leading to more accurate analysis and reliable insights.

In remote sensing image super-resolution (RSISR), two common strategies for upsampling are pre-upsampling and post-upsampling. Pre-upsampling uses traditional interpolation methods, such as bicubic or bilinear, to enlarge the low-resolution image to match the desired high-resolution output before feeding it into the network. This approach simplifies network design, as the input is already at the target resolution; however, it can introduce artifacts and blur that the network must correct. Post-upsampling processes the LR image at its original resolution through convolutional layers to extract features and then upsamples it to HR in the final step using learned layers like transposed convolution or PixelShuffle [2]. This method can yield higher-quality images by integrating upsampling into the network. Both strategies highlight the need for advancements in capturing long-range dependencies and global context to improve RSISR performance. Therefore, we adopted a post-upsampling strategy, as it typically offers a better balance of reconstruction accuracy and memory efficiency.

Convolutional neural networks (CNNs) have been instrumental in advancing RSISR, enabling the extraction of richer details from remote sensing data. Early models, such as SRCNN [3] and FSRCNN [4], demonstrated the potential of deep learning for image enhancement. The EDSR [5] model introduced deeper architectures and residual connections, achieving superior results. Other CNN-based models, such as VDSR [6], LapSRN [7], RDN [8], OmniSR [9], and FENet [10], have made notable contributions, each bringing unique architectural innovations that enhance image quality in remote sensing applications. However, these models still struggle with capturing long-range dependencies and global context due to their inherent locality.

Beyond traditional CNN approaches, other methods have emerged to address RSISR challenges. Lightweight convolutional structures like CTN [11] reduce network parameters and operations while maintaining high performance. HSENet [12] leverages similar ground targets within remote sensing images to enhance feature representation. Two-stage networks like TSFNet [13] integrate spatial and frequency features for step-by-step super-resolution refinement. The CSA-FE [14] method effectively extracts features using channel and spatial attention mechanisms.

In recent years, transformer models [15] have demonstrated considerable potential for capturing global contextual information and long-range dependencies. Models like ESRT [16], TransENet [17], SwinIR [18], and MAT [19] demon-

Corresponding authors: Allen Patnaik, Sultan Alfarhood

This research is funded by the Ongoing Research Funding Program (ORF-2026-890), King Saud University, Riyadh, Saudi Arabia.

Allen Patnaik is with the Department of Electronics and Electrical Engineering, IIT Guwahati, Guwahati 781039, India; allen.patnaik@iitg.ac.in.

Arpit Gour is with the Department of Electronics and Electrical Engineering, IIT Guwahati, Guwahati 781039, India; g.arpit@iitg.ac.in.

M. K. Bhuyan is with the Department of Electronics and Electrical Engineering, IIT Guwahati, Guwahati 781039, India; mkb@iitg.ac.in.

Karl F. MacDorman is with the Luddy School of Informatics, Computing, and Engineering, Indiana University, Indiana 46202, USA; kmacdorm@iu.edu.

Sultan Alfarhood is with the Department of Computer Science, College of Computer and Information Sciences, King Saud University, P.O. Box 51178, Riyadh 11543, Saudi Arabia; sultanf@ksu.edu.sa.

Mejdl Safran is with the Department of Computer Science, College of Computer and Information Sciences, King Saud University, P.O. Box 51178, Riyadh 11543, Saudi Arabia; mejdl@ksu.edu.sa.

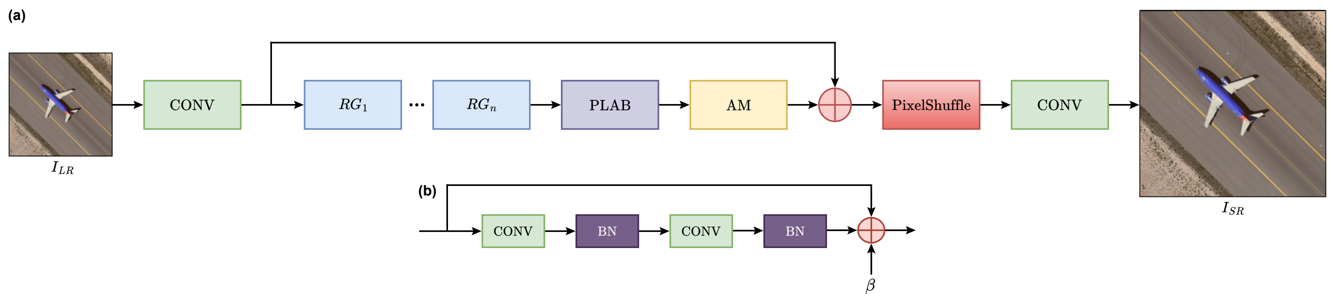


Fig. 1. (a) Overall architecture of PLAN. The model consists of a parallel lattice attention block (PLAB) applied after the residual groups (RGs) to aggregate multi-branch features and enhance global–local interactions. The attention module (AM) further enhances salient regions before reconstruction. (b) The residual block (RB) used in a residual group.

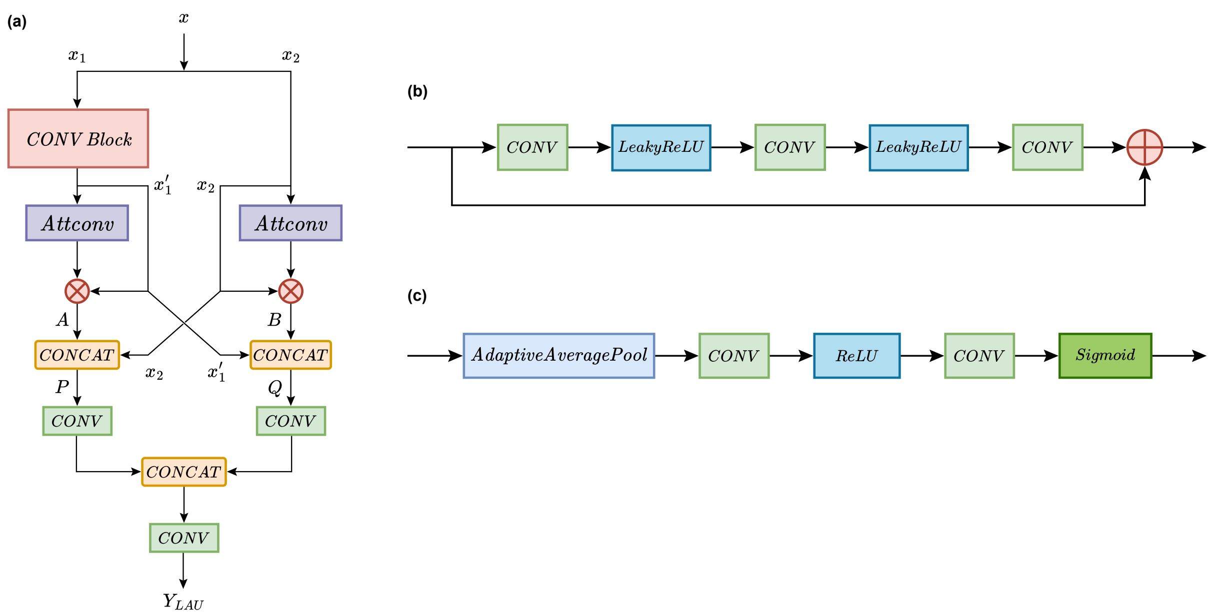


Fig. 2. (a) Lattice attention unit. Illustration of two basic components of LAUnit, including (b) CONV Block and (c) Attconv.

strate marked improvements through self-attention mechanisms. These models enhance feature extraction and reconstruction but are computationally demanding, requiring substantial memory and complex training, especially for high-resolution images. EHC-DMSR [20] and EDiffSR [21] introduce techniques like diffusion models and Fourier high-frequency spatial constraints for precise analysis and interpretation across various applications.

Despite advancements, limitations persist. CNNs excel at capturing local features but often fail to model global context effectively, limiting their ability to reconstruct fine details in complex images. Though effective in capturing global information, transformers can be prone to overfitting when trained on limited datasets due to their high model capacity. Although various attention mechanisms have been proposed to alleviate these issues, they often still focus primarily on either local or global feature interactions rather than jointly enhancing both. These limitations motivate the development of a more efficient attention strategy that can leverage complementary local and global cues while maintaining computational efficiency.

We introduce the parallel lattice attention network (PLAN) to resolve these limitations of CNNs and transformers in RSISR (Fig. 1). PLAN integrates mechanisms to improve feature extraction and detail preservation while maintaining computational efficiency. PLAN employs a lattice to route information between parallel branches. This architecture comprises a lattice attention unit for advanced feature extraction, a parallel lattice attention block (PLAB) for capturing different levels of refined details, and an attention module (AM) for enhancing the essential parts of remote sensing images, leveraging the strengths of CNNs and transformers (Figs. 2–4).

Our model's contributions are listed below:

- The lattice attention unit (LAUnit) accumulates information from different parts of the remote sensing image, captures complex dependencies, and efficiently focuses on important features throughout its processing stages by reusing rich features.
- The parallel lattice attention block configuration enables our model to focus on fine-grained details and their broader context by processing the input feature maps in parallel branches before merging them via a learned fusion layer. The results are then fused with the attention

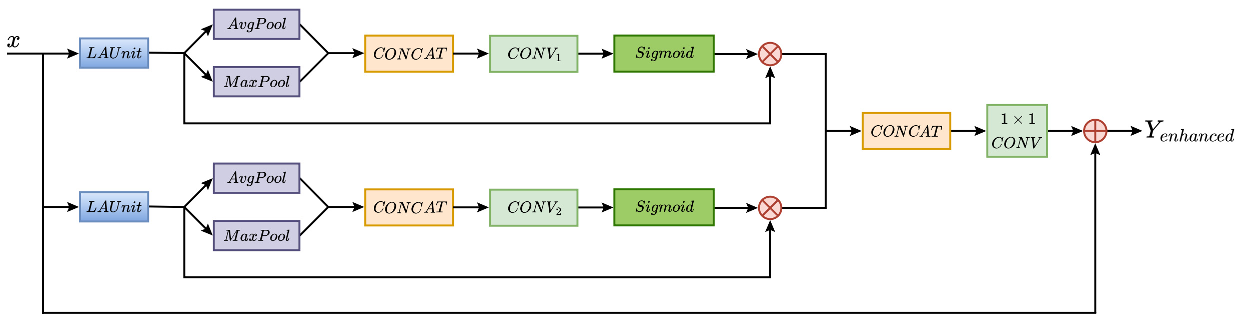


Fig. 3. Architecture of the parallel lattice attention block.

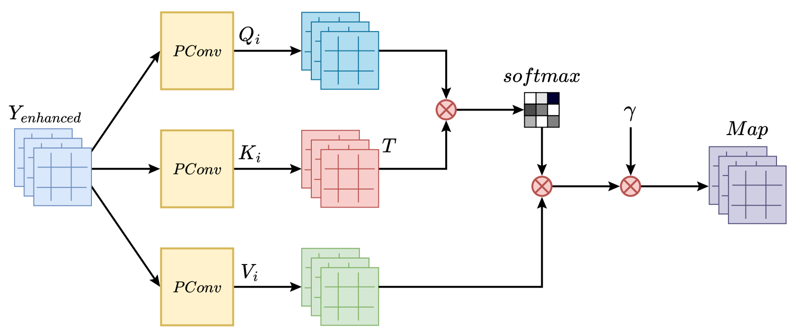


Fig. 4. Architecture of the attention module.

module, which captures essential global context information from remote-sensing images.
- Optimized for remote sensing benchmarks such as the AID, UC Merced, and WHU-RS19 datasets, our model employs parallel lattice attention blocks and residual groups to enhance feature extraction and stability. This design balances accuracy with runtime efficiency, outperforming comparable approaches.

The remaining sections are organized as follows. Section II presents existing research on the RSISR field. Section III discusses the methodology employed for our proposed model. Section IV introduces the datasets, provides a quantitative and qualitative analysis of the data, interprets the results, and presents an ablation study. Finally, Section V gives the conclusion.

II. Related Work

A. CNN-Based Super-Resolution Architectures

Pioneering CNN research has addressed super-resolution through a range of structural innovations. Li et al. [22] developed a recursive fractal network to accelerate execution via recursive mechanisms, while Zhang et al. [23] introduced an unfolding model specifically for denoising and deblurring. To capture higher-level features for detail refinement, Li et al. [24] proposed a feedback structure. Additionally, several methods [25], [26] employ multiscale residual techniques to integrate information across local and global levels. However, complex variations in texture and illumination often still hinder performance in these architectures.

Lattice networks address these limitations by using grid-like topologies that enable nodes to interact over longer distances, thereby refining image details and achieving superior performance in computer vision tasks [27]–[29]. For instance, Luo et al. [30] reuse aggregated information from intermediate layers for image restoration, while Hao et al. [31] employ lattice-gated units to refine fine-grained details via channel interaction.

However, unlike standard lattice networks [27], [28] that typically employ recursive feedback filters or sparse quantization for 3D data, PLAN introduces parallel lattice attention blocks. Instead of recursive sequencing, these blocks split input features into concurrent branches that are processed by dual lattice attention units. This design simultaneously extracts fine-grained details and broader contextual structures, integrating them via attention-based fusion. By replacing sequential processing with this parallel approach, PLAN maximizes GPU parallelism and structural efficiency.

B. Hybrid Approaches for RSISR

Combining CNN and transformer modules has enhanced super-resolution models [32]–[36], leading to improved reconstruction accuracy. For instance, Wang et al. [37] proposed a network utilizing global context information embedded across different scales. Qin et al. [38] integrated channel and spatial attention blocks into a Swin transformer to refine the feature representation, while Tang et al. [39] used a pyramid split attention mechanism alongside enhanced spatial attention to improve extraction capabilities. Similarly, Xudong et al. [40] developed a hybrid model fusing multiscale CNN features with a transformer block to explore multiscale global information. Distinct from these approaches, our model uniquely integrates an attention unit within a parallel lattice structure alongside a self-attention mechanism, effectively capturing both local and global dependencies with greater structural efficiency.

C. Attention Mechanisms for RSISR

Attention mechanisms have been widely incorporated into RSISR models to enhance feature representation and reconstruction quality. Recent studies [41], [42] employ hybrid models that combine channel and spatial attention to better capture both global context and local details. Similarly, self-attention modules have been applied to adaptively explore global structure while preserving fine local details in hierarchical feature spaces [43]–[45].

More recently, transformer-based architectures like ACT-SR [44], SCAT [45], and CSCT [42] have introduced aggregation connections, shifted channel attention, and channel-spatial coherence to enhance global modeling. While these methods achieve state-of-the-art fidelity, they entail high computational costs. Other approaches, such as dual attention enhancement networks [46], focus on efficiency by refining features via contrast-aware channel attention and forward fusion strategies. However, many of these mechanisms still emphasize either local or global features in isolation or rely on separate modules that lack joint optimization. These limitations underscore the need for integrated designs like PLAN. By adopting a parallel lattice framework to simultaneously capture complementary local and global cues, PLAN achieves competitive performance with a significantly lower runtime (30.97 ms), making it highly suitable for resource-constrained on-board processing.

III. Proposed Method

This section introduces the parallel lattice attention network for RSISR. Section III-A presents the overall network framework; Section III-B introduces the lattice attention unit; Section III-C details the parallel lattice attention block structure; and Section III-D explains the attention module and image reconstruction process.

A. Network Architecture

As illustrated in Fig. 1, PLAN comprises four primary components: initial detail extraction, the parallel lattice attention block, the attention module, and final reconstruction. Let I_{LR} and I_{SR} denote the input low-resolution image and the output super-resolved image, respectively. First, a convolutional layer extracts shallow features from I_{LR}, yielding the initial feature map F_0:

$$F_0 = H_c(I_{\text{LR}}) \tag{1}$$

where H_c denotes the initial feature extraction layer consisting of a 3×3 convolution.

F_0 is then processed by residual groups (RG) to learn deep feature representations. These groups use skip connections to mitigate the vanishing gradient problem. Each RG comprises residual blocks (RBs), each consisting of convolutional layers with a 3×3 kernel, followed by batch normalization and ReLU activations. Residual connections maintain network stability by adding the original input to the layer's output, preserving fine details and facilitating effective information flow.

$$F_{\text{RB}} = F_{\text{in}} + \beta H_{\text{CBN}}(F_{\text{in}}) \tag{2}$$

where H_{CBN} represents the residual block transformation (two convolutional layers with batch normalization), β is the scaling parameter, and F_{in} denotes the input to the block. Although some recent SR architectures remove batch normalization to conserve memory [5], [18], the residual groups retain it to stabilize the training of deep features before they enter the parallel lattice attention block.

Next, the features are processed by the parallel lattice attention block, which enhances and refines representations using specialized structural units. These blocks collaborate to produce a more detailed and refined feature map F_{PLAB}:

$$F_{\text{RG}_k} = H_{\text{RG}_k}\left(F_{\text{RG}(k-1)}\right), \quad k = 1, \ldots, n \tag{3}$$

$$F_{\text{PLAB}} = H_p(F_{\text{RG}_n}) \tag{4}$$

where H_p denotes the PLAB operation, and F_{RG_n} is the output of the final residual group.

F_{PLAB} is fed to the attention module to emphasize salient elements and capture long-range dependencies. This module incorporates a self-attention mechanism defined as

$$F_\alpha = \gamma \cdot \text{softmax}(Q_i K_i^T) V_i(F_{\text{PLAB}}) \tag{5}$$

where γ is a learnable scaling parameter for stabilization, and Q_i, K_i, and V_i represent the query, key, and value projections, respectively.

The refined attention features F_α are added back to the initial detail features F_0 to preserve global context. The fused features are then processed by the upsampling module (PixelShuffle), which rearranges elements to increase spatial resolution. Finally, a convolutional layer reconstructs the high-resolution output image I_{SR}:

$$I_{\text{SR}} = H_{\text{rec}}(f_p(F_\alpha + F_0)) \tag{6}$$

where f_p denotes the pixel-shuffle upsampling operation, and H_{rec} represents the final reconstruction convolution.

Algorithm 1 SISR reconstruction of remote sensing images

Require: Low-resolution remote sensing image I_{LR}
Ensure: High-resolution remote sensing image I_{SR}
1: Extract shallow features F_0 from I_{LR} using the initial convolution layer (Eq. 1).
2: Process F_0 through residual groups to learn deep feature representations F_{RG_n} (Eq. 2–3).
3: Process F_{RG_n} through the parallel lattice attention block to refine high-frequency details (textures, edges) and generate F_{PLAB} (Eq. 4).
4: Apply the attention mechanism to F_{PLAB} to capture global context and long-range dependencies, yielding F_α (Eq. 5).
5: Reconstruct the high-resolution output I_{SR} by fusing F_α with F_0 and applying pixel-shuffle upsampling (Eq. 6).
6: **Return** I_{SR}

Algorithm 1 provides the pseudocode of the PLAN network for the reconstruction of high-resolution remote sensing images. Our approach reuses features with contextual information to capture complex relationships within image data, preserving and enhancing fine details.

B. Lattice Attention Unit

The lattice attention unit (LAUnit) in PLAN enhances and refines features from input images, a step crucial for high-quality output. The LAUnit enables cross-channel interaction by reusing rich features that contain high-frequency details, such as textures and edges. This module integrates two components: the CONV Block and Attconv. The CONV Block consists of residually connected pairs of 3×3 convolutional layers followed by LeakyReLU nonlinear activation functions. The Attconv component processes an input feature map x_c of size $H_i \times W_i$ via adaptive average pooling, calculated as

$$Y_c = \frac{1}{H_i \times W_i} \sum_{h=1}^{H_i} \sum_{w=1}^{W_i} x_c(h, w) \tag{7}$$

where H_i and W_i denote the height and width of the feature map, respectively. The attention convolution operation is defined as

$$\mathrm{Attconv}(Y_c) = \phi(H_c(\mathrm{ReLU}(H_c(Y_c)))) \tag{8}$$

where ϕ denotes the sigmoid activation, ReLU is the nonlinear activation function, and H_c represents a convolution layer with a kernel size of 1×1.

The LAUnit efficiently extracts features by dividing the input feature map into two channel groups, which are processed in parallel. As shown in Fig. 2, the cross-stream connections between these branches form a lattice topology, allowing the network to dynamically modulate information flow between local and global feature extractors. Each group is processed individually through a series of CONV Blocks, followed by an Attconv unit that recombines the channels to selectively enhance features. This method ensures efficient feature extraction, a crucial step for high-resolution image reconstruction. Mathematically, for an input $\mathbf{x}$ with channels split into $\mathbf{x}_1$ and $\mathbf{x}_2$:

$$\mathbf{x}_1' = \text{CONV Block}(\mathbf{x}_1) \tag{9}$$

where $\mathbf{x}_1$ comprises half the original channels, and CONV Block denotes a sequence of convolution blocks with LeakyReLU activations.

$$\mathbf{A} = \mathrm{Attconv}(\mathbf{x}_1') \tag{10}$$

where $\mathrm{Attconv}$ is the attention convolution block.

$$\mathbf{P} = \mathrm{concat}(\mathbf{x}_2, \mathbf{A} \odot \mathbf{x}_1') \tag{11}$$

where $\mathbf{x}_2$ comprises the remaining half of the original channels, and $\mathbf{P}$ denotes information being fused from $\mathbf{x}_2$ to $\mathbf{x}_1$.

$$\mathbf{B} = \mathrm{Attconv}(\mathbf{x}_2) \tag{12}$$

$$\mathbf{Q} = \mathrm{concat}(\mathbf{x}_1, \mathbf{B} \odot \mathbf{x}_2) \tag{13}$$

where $\mathbf{Q}$ denotes information being fused from $\mathbf{x}_1$ to $\mathbf{x}_2$.

$$\mathbf{Y}_{\mathrm{LAU}} = w_{out}(\mathrm{concat}(w_p(\mathbf{P}), w_q(\mathbf{Q}))) \tag{14}$$

where w denotes a convolution operation with a kernel size of 3×3, and $\mathbf{Y}_{\mathrm{LAU}}$ represents the combined features from the different branches, the final output of the LAUnit.

The cascaded convolution branch prioritizes fine-grained details, whereas the parallel branch captures broader contextual structures. This dual-path design facilitates the reuse of both local and global features, enhancing information flow throughout the network. To validate this complementary effect, we conducted an ablation study by systematically removing each branch. As shown in Table V, eliminating either branch resulted in a substantial performance degradation, confirming that both components are essential for optimal reconstruction.

C. Parallel Lattice Attention Block

As illustrated in Fig. 3, the image data is processed to adaptively merge features from LAUnits, passing through a parallel lattice attention block (PLAB) to capture both fine details and broader context. This structure enhances key channels by selectively integrating crucial information using convolutional kernels of sizes 3×3 and 5×5. Furthermore, the block selectively aggregates salient information to produce a sharper final image while adaptively reweighting channels.

TABLE I
REMOTE SENSING DATASETS.

Split	Dataset	Images	Classes	Resolution
Train/Val./Test	AID [47]	10,000	30	600×600
Test	UC Merced [48]	2100	21	256×256
Test	WHU-RS19 [49]	1005	19	600×600

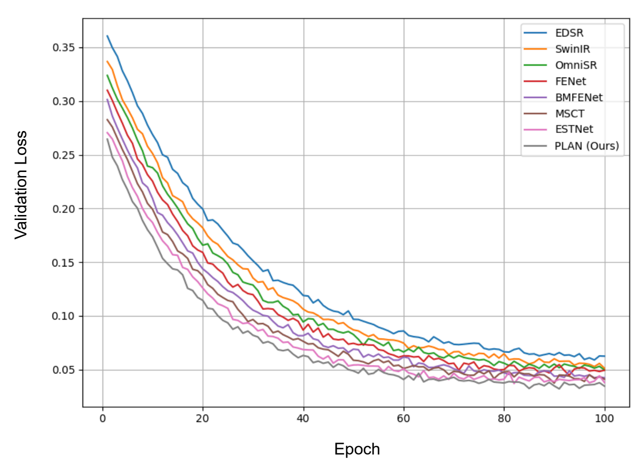


Fig. 5. Validation loss curves comparing the proposed PLAN model with state-of-the-art methods on the AID dataset over 100 epochs.

TABLE II
PSNR AND SSIM COMPARISON OF PLAN WITH OTHERS ON AID, UC MERCED, AND WHU-RS19.

Scale	Method	AID		UC Merced		WHU-RS19	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
2×	Bicubic	29.23	0.8131	24.45	0.7039	24.82	0.7126
	EDSR [5]	33.12	0.8323	30.72	0.7838	31.66	0.8023
	SwinIR [18]	33.45	0.8541	31.43	0.8046	32.12	0.8274
	ESRT [16]	34.27	0.8742	31.52	0.8131	32.45	0.8351
	OmniSR [9]	34.52	0.8951	31.60	0.8367	32.56	0.8467
	FENet [10]	34.72	0.8962	31.89	0.8389	32.60	0.8483
	BMFENet [31]	34.73	0.9021	31.94	0.8425	32.66	0.8501
	MSCT [40]	34.77	0.9141	32.03	0.8589	32.67	0.8688
	ESTNet [43]	34.79	0.9254	32.05	0.8597	32.69	0.8725
	PLAN (ours)	**34.87**	**0.9371**	**32.12**	**0.8807**	**32.71**	**0.8797**
3×	Bicubic	26.34	0.7336	23.15	0.6312	23.66	0.6691
	EDSR [5]	30.66	0.8291	26.44	0.7431	28.60	0.7522
	SwinIR [18]	31.01	0.8302	26.58	0.7530	28.63	0.7772
	ESRT [16]	30.98	0.8341	26.67	0.7535	28.86	0.7788
	OmniSR [9]	31.41	0.8422	27.06	0.7523	29.38	0.7846
	FENet [10]	31.46	0.8443	27.22	0.7684	29.49	0.7869
	BMFENet [31]	31.48	0.8449	27.23	0.76	29.53	0.7925
	MSCT [40]	31.50	0.8452	27.24	0.7878	29.59	0.7975
	ESTNet [43]	31.52	0.8492	27.28	0.7888	29.63	0.7978
	PLAN (ours)	**31.58**	**0.8564**	**27.35**	**0.7902**	**29.74**	**0.7981**
4×	Bicubic	24.51	0.6621	22.74	0.6174	23.83	0.6671
	EDSR [5]	28.65	0.7422	25.19	0.6732	26.96	0.7219
	SwinIR [18]	29.22	0.7521	25.21	0.6750	27.29	0.7245
	ESRT [16]	29.30	0.7641	25.21	0.6760	27.42	0.7332
	OmniSR [9]	29.47	0.7783	25.31	0.6888	28.03	0.7385
	FENet [10]	29.75	0.7792	25.64	0.6941	28.06	0.7450
	BMFENet [31]	29.77	0.7795	25.70	0.6989	28.07	0.7467
	MSCT [40]	29.85	0.7801	25.73	0.7026	28.10	0.7488
	ESTNet [43]	29.88	0.7853	25.76	0.7083	28.52	0.7497
	PLAN (ours)	**29.98**	**0.7923**	**26.59**	**0.7156**	**29.16**	**0.7597**

A residual connection within the PLAB maintains network stability. The generation of the reused feature $\mathbf{Y}$ is

$$\mathbf{Y}_1 = Y_{\text{LAU1}}(\mathbf{x}) \tag{15}$$

$$\mathbf{Y}_2 = Y_{\text{LAU2}}(\mathbf{x}) \tag{16}$$

where LAU_i represents successive LAUnit operations.

The top-branch features are calculated as

$$Y_{\text{top}} = \phi H_{c1}[\text{concat}\{P_{\text{avg}}(Y_1), P_{\text{max}}(Y_1)\}] \odot Y_1 \tag{17}$$

where H_{c1} has a kernel size of 3×3; P_{avg} and P_{max} denote average pooling and max pooling, respectively. The bottom-branch features are derived as

$$Y_{\text{bottom}} = \phi H_{c2}[\text{concat}\{P_{\text{avg}}(Y_2), P_{\text{max}}(Y_2)\}] \odot Y_2 \tag{18}$$

where H_{c2} has a kernel size of 5×5, and $\odot$ denotes element-wise multiplication. Finally, the enhanced features are fused via

$$\mathbf{Y}_{\text{enhanced}} = H_{1\times1}[\text{concat}(Y_{\text{top}}, Y_{\text{bottom}})] + \mathbf{x} \tag{19}$$

where $H_{1\times1}$ denotes a convolution layer with a kernel size of 1×1.

The PLAB architecture ensures the effective capture of complex details at varying levels. We analyzed the contributions of the top and bottom branches (Table VI). The results confirm that integrating both branches is essential for robust feature enhancement, yielding superior image quality.

D. Attention Module

Following the PLAB, the attention module (AM) emphasizes salient regions within the feature maps. This mechanism captures global context and long-range dependencies, resulting in coherent and detailed reconstructions.

As illustrated in Fig. 4, the module employs a self-attention mechanism. In our implementation, we employ a multi-head attention mechanism with 2 heads to capture contextual dependencies. For simplicity, the following formulation describes the process for a single head.

We compute three projections of the input feature map: query (Q_i), key (K_i), and value (V_i). These projections are generated via convolutional layers. The Q_i and K_i projections generate an attention map that identifies salient spatial locations. Mathematically, for an input feature map $\mathbf{Y}_{\text{enhanced}}$, the process is defined as

$$Q_i = H_q(\mathbf{Y}_{\text{enhanced}}) \tag{20}$$

$$K_i = H_k(\mathbf{Y}_{\text{enhanced}}) \tag{21}$$

$$V_i = H_v(\mathbf{Y}_{\text{enhanced}}) \tag{22}$$

where H_q, H_k, and H_v denote pointwise (1×1) convolutional layers for the query, key, and value projections, respectively.

The attention map $\mathbf{A}$ is computed as the softmax of the dot product between the query Q_i and the transposed key K_i:

$$\mathbf{A} = \mathrm{softmax}(Q_i \cdot K_i^\top) \tag{23}$$

Subsequently, this attention map weights the value projection V_i to generate the modulated features F_α:

$$F_\alpha = \gamma \cdot \mathbf{A} \cdot V_i \tag{24}$$

where γ is a learnable scaling parameter that stabilizes the feature refinement. Finally, the output of the attention module is fused with the initial feature map F_0 via a residual connection:

$$F_{\text{AM}} = F_0 + F_\alpha \tag{25}$$

The fused feature map F_{AM} is subsequently processed by the reconstruction module, which uses upsampling layers to increase spatial resolution. This process generates the final super-resolved output I_{SR}, preserving fine structural details.

We adopt the $\mathcal{L}_1$ loss as the optimization objective to supervise the reconstruction of high-resolution images:

$$\mathcal{L}_1 = \|I_{\text{SR}} - I_{\text{HR}}\|_1 \tag{26}$$

IV. Experiments

A. Datasets and Metrics

We trained the PLAN and benchmark models on the AID dataset [47], which contains 10,000 images in 30 classes, including beaches and baseball fields, each sized at 600×600 pixels. The dataset was divided into training (70%), validation (20%), and testing (10%) sets.

We evaluated model performance on remote sensing image super-resolution at multiple scales using three benchmark datasets: AID, UC Merced [48], and WHU-RS19 [49] (Table I). To verify the quality of super-resolved images, we used peak signal-to-noise ratio (PSNR) [50] and structural similarity index measurement (SSIM) [51], focusing on the Y channel (luminance) in the YCbCr color space for visual perception.

During training on AID, the $\mathcal{L}_1$ loss on both the training and validation sets decreased steadily, with no substantial gap between them, indicating stable optimization and limited overfitting. Fig. 5 shows the validation loss curves for all models, indicating convergence after approximately 100 epochs. For each network, the training checkpoint with the lowest validation loss was selected for testing.

TABLE III
PAIRED TWO-TAILED t-TESTS (p-VALUE, COHEN'S d) COMPARING PLAN AND COMPETING METHODS ON UC MERCED ($2\times$ SCALE).

Method	t	p	d
Bicubic	157.32	9.79×10^{-9}	6.87
EDSR [5]	28.85	8.70×10^{-6}	1.26
SwinIR [18]	19.02	4.50×10^{-5}	0.83
ESRT [16]	23.74	1.87×10^{-5}	1.04
OmniSR [9]	25.31	1.45×10^{-5}	1.10
FENet [10]	6.46	2.96×10^{-3}	0.28
BMFENet [31]	7.10	2.10×10^{-3}	0.31
MSCT [40]	6.02	3.84×10^{-3}	0.26
ESTNet [43]	7.85	1.65×10^{-3}	0.34

TABLE IV
RUNTIME, PARAMETER COUNT, GFLOPS, PSNR, AND SSIM COMPARISON ON WHU-RS19 ($4\times$ SCALING).

Method	**Runtime (ms)**	**Params (M)**	**GFLOPs**	**PSNR**	**SSIM**
EDSR [5]	21.68	21.84	1427.15	26.96	0.7219
SwinIR [18]	98.47	2.11	129.90	27.29	0.7245
ESRT [16]	106.34	3.16	72.62	27.42	0.7332
FENet [10]	63.52	0.37	1.45	28.03	0.7385
OmniSR [9]	45.02	0.81	3.12	28.06	0.7450
BMFENet [31]	51.43	0.48	29.40	28.07	0.7467
MSCT [40]	57.15	1.39	16.76	28.10	0.7488
ESTNet [43]	67.95	3.28	89.32	28.52	0.7497
PLAN (ours)	30.97	1.20	52.97	**29.16**	**0.7597**

B. Implementation Settings

We implemented all models using PyTorch [52] on an NVIDIA GeForce RTX 3080 GPU with FP32 precision. The PLAN architecture comprises 8 residual groups, each containing 5 residual blocks, and uses 2 heads in the attention module.

To ensure a fair comparison, all models were trained for 100 epochs with a batch size of 16 under identical optimization settings. Network weights were initialized using Kaiming normal initialization [53]. Input data consisted of 192×192 patches randomly cropped from remote sensing images, augmented via rotation, color jittering, and Gaussian blurring. We optimized the models using the L1 loss function and the Adam optimizer with a learning rate of 10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The code is available at https://github.com/allenptnk/PLAN.

C. Quantitative Results

We evaluated remote-sensing image SR on the AID, UC Merced, and WHU-RS19 datasets at scales of $2\times$, $3\times$, and $4\times$, generating low-resolution inputs via bicubic downsampling. We benchmarked PLAN against bicubic, EDSR [5], SwinIR [18], ESRT [16], OmniSR [9], FENet [10], BMFENet [31], MSCT [40] and ESTNet [43] using PSNR and SSIM (Table II). The quantitative results confirm the model's superior retention of rich edges and spatial details. As provided in Table III, statistical tests on the improvements in PSNR confirm

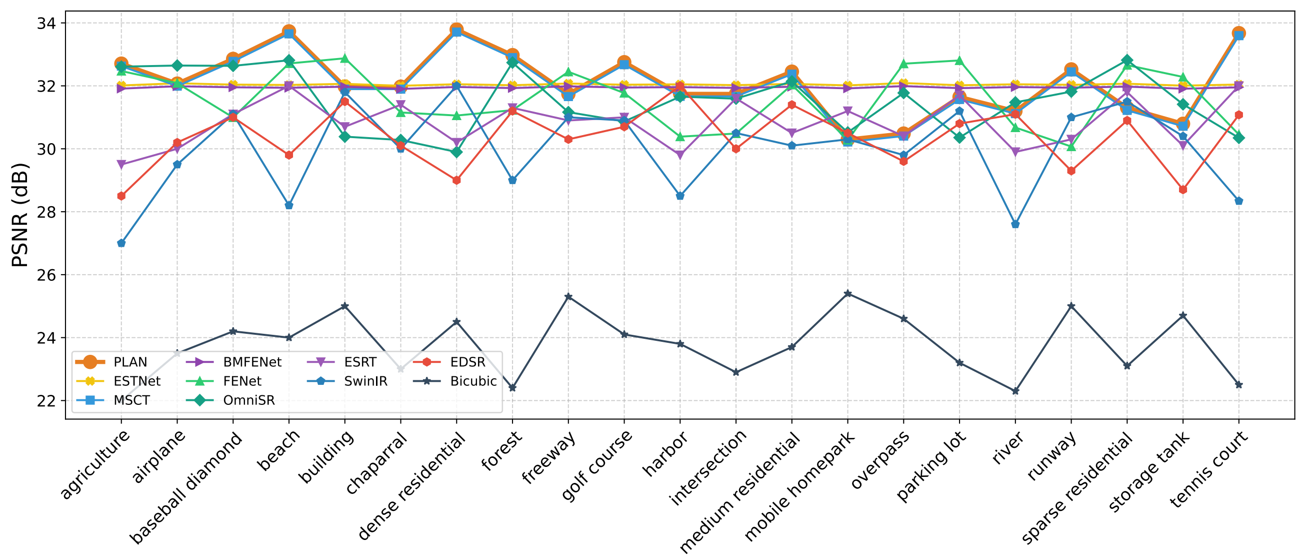


Fig. 6. PSNR comparison of methods by class on UC Merced (2× scaling).

TABLE V
ABLATION STUDY ON THE LATTICE ATTENTION UNIT ON UC MERCED (2× SCALING).

Variant	PSNR (dB)	SSIM
Full model (P + Q)	**32.12**	**0.8807**
w/o Branch P (only Q)	30.78	0.8721
w/o Branch Q (only P)	31.02	0.8759
w/o both branches (baseline)	29.10	0.8651

TABLE VI
ABLATION STUDY ON THE PARALLEL LATTICE ATTENTION BLOCK ON WHU-RS19 (4× SCALING). HERE k DENOTES THE KERNEL SIZE OF THE CONVOLUTION BLOCK.

Variant	PSNR	SSIM
top branch: $k = 5$; bottom branch: $k = 5$	27.46	0.7390
top branch: $k = 3$; bottom branch: $k = 3$	27.86	0.7447
top branch: $k = 3$; bottom branch: $k = 5$	**29.16**	**0.7597**
without top branch	26.35	0.7383
without bottom branch	26.42	0.7388

TABLE VII
ABLATION STUDY OF NETWORK COMPONENTS ON WHU-RS19 (4× SCALING). AM: ATTENTION MODULE, PLAB: PARALLEL LATTICE ATTENTION BLOCK, RG: RESIDUAL GROUP.

RG	PLAB	AM	Params (M)	GFLOPs	PSNR	SSIM
✓	×	✓	0.96	45.55	24.22	0.6998
✓	✓	×	1.19	53.25	26.42	0.7002
×	✓	✓	0.87	43.01	27.44	0.7230
✓	✓	✓	1.20	52.97	**29.16**	**0.7597**

TABLE VIII
IMPACT OF LAUNIT QUANTITY AND KERNEL SIZES ON WHU-RS19 (4× SCALING).

No. of LAUnits	Kernel Sizes	Parameters	GFLOPs	PSNR	SSIM
1	3	1.20 M	52.70	27.90	0.7493
2	3, 5	1.20 M	52.97	**29.16**	**0.7597**
3	3, 5, 7	1.21 M	53.25	28.51	0.7497

that PLAN performs significantly better than the comparison methods. Fig. 6 details performance across individual classes.

D. Model Complexity Analysis

We compare our model with several state-of-the-art models on the WHU-RS19 dataset, focusing on runtime, parameter count, and GFLOPs. Table IV shows results on the WHU-RS19 dataset with a 4× scaling factor. Our model balances performance and efficiency, achieving superior super-resolution results with lower complexity and higher PSNR and SSIM values than other models. FENet and OmniSR models have fewer parameters and consume fewer GFLOPs than ours, but our model achieves higher accuracy and faster runtime. This highlights our model's ability to deliver high-quality output with greater computational efficiency.

Although FENet [10] and OmniSR [9] possess fewer parameters and GFLOPs, they exhibit higher inference latency than PLAN. While high GFLOPs typically correlate with increased energy consumption, they do not directly dictate inference speed if the architecture can be effectively parallelized. FENet relies on sequential feature fusion, while OmniSR employs cascaded aggregators; both impose sequential dependencies that hinder GPU throughput. In contrast, PLAN's parallel lattice framework processes split feature maps via concurrent branches within the PLAB. This design maximizes GPU parallelism and minimizes sequential overhead, yielding faster runtime despite a higher operation count. Consequently, PLAN prioritizes low-latency performance, making it advantageous for time-critical on-board processing where response speed is paramount.

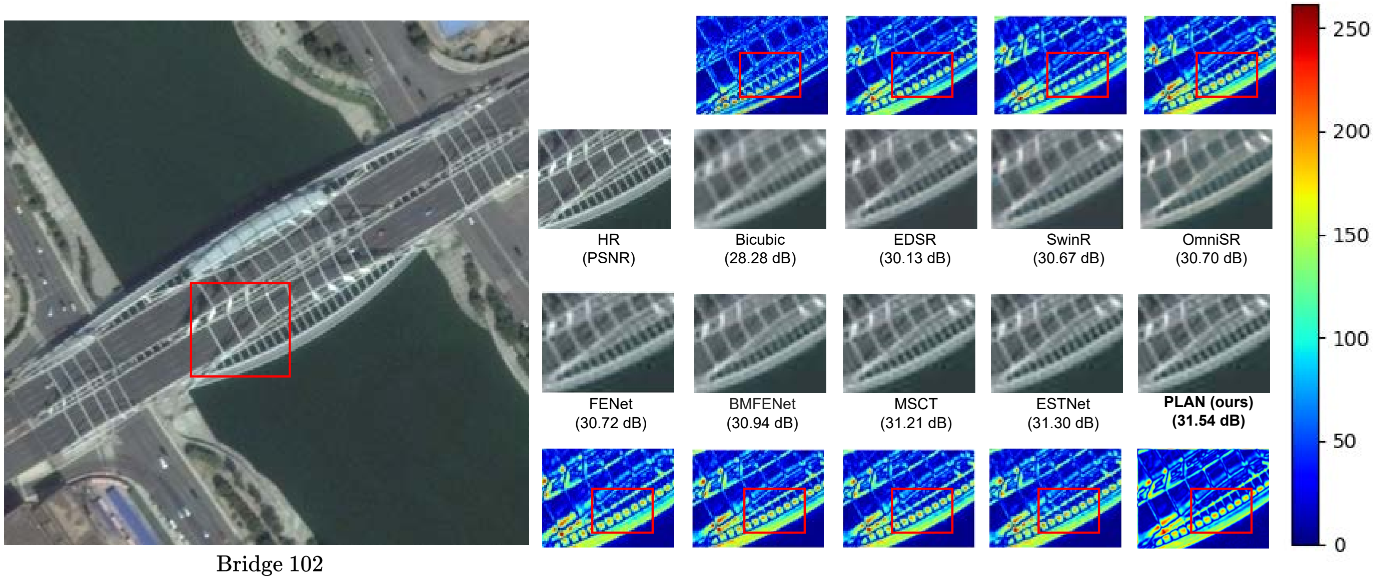


Fig. 7. Comparison of SR methods on AID Bridge 102 with color error maps (4× scaling). The red box indicates the zoomed-in region for visual comparison. Color error maps are shown to visualize pixel-wise reconstruction errors, where warmer colors indicate larger errors.

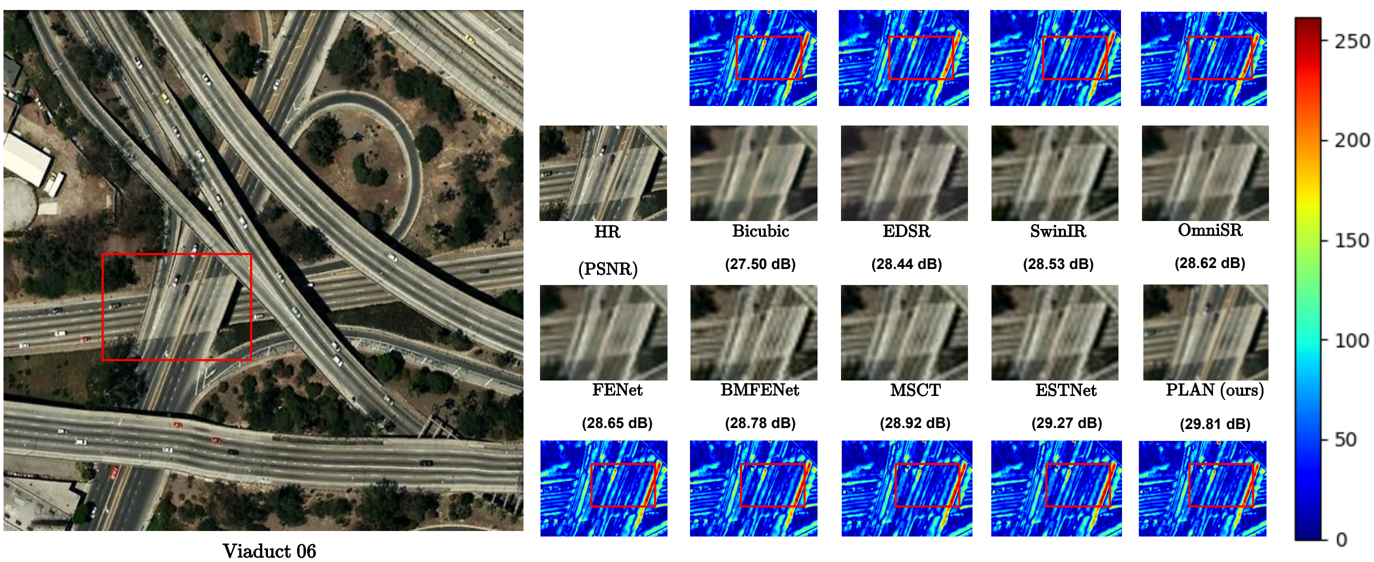


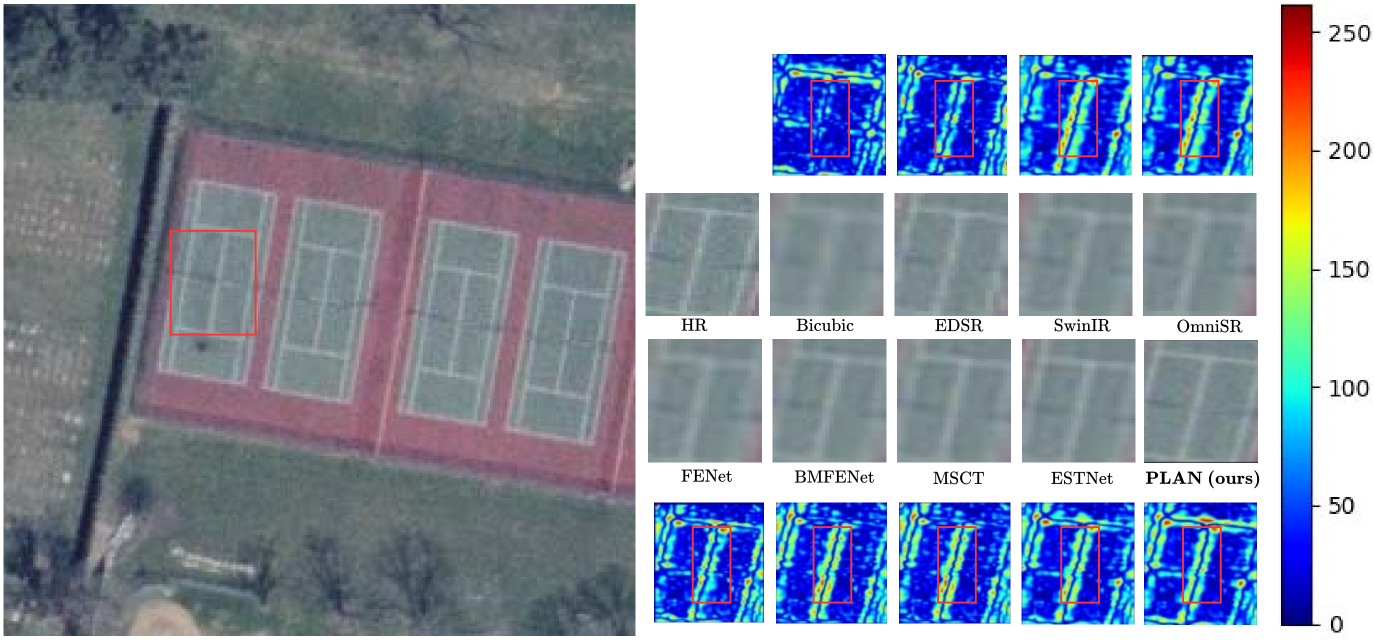
Fig. 8. Comparison of SR methods on WHU-RS19 Viaduct 06 with color error maps (4× scaling). The red box indicates the zoomed-in region for visual comparison. Color error maps visualize pixel-wise reconstruction errors, where warmer colors denote larger errors.

TABLE IX
PERFORMANCE COMPARISON ON UC MERCED (4× SCALING).

Method	AG ↑	NIQE ↓
Bicubic	0.52	33.24
EDSR [5]	0.53	32.72
SwinIR [18]	0.53	32.64
ESRT [16]	0.54	32.16
FENet [10]	0.54	31.52
OmniSR [9]	0.54	31.99
BMFENet [31]	0.55	31.65
MSCT [40]	0.55	31.40
ESTNet [43]	0.55	31.29
PLAN (ours)	**0.56**	**29.86**

E. Ablation Study

This section presents results from experiments on the WHU-RS19 dataset to examine the contribution of each component in our method. To assess the specific contributions of architectural components, including residual groups, the parallel lattice attention block, and the attention module, we conducted ablation studies on the WHU-RS19 dataset with a ×4 scaling factor. The results indicate that incorporating PLAB and residual blocks yields substantial improvements in PSNR and SSIM, while the AM further enhances reconstruction quality, with the full integration of these components achieving the highest performance (Table VII). We further investigated kernel size configurations and found that a hybrid combination of sizes 3 and 5 is most effective (Table VI). Finally, we



Tennis court 88

Fig. 9. Comparison of SR methods on UC Merced Tennis Court 88 with color error maps (4× scaling) under real-world degradation conditions. The red box indicates the zoomed-in region for visual comparison. Color error maps visualize pixel-wise reconstruction errors, where warmer colors represent larger errors.

examined the impact of LAUnit quantity and determined that a configuration of 2 LAUnits offers the optimal balance between computational efficiency and parameter count (Table VIII). The statistical analysis confirms that removing either branch, using the same kernel size, or more or fewer LAUnits leads to statistically significant performance degradation, validating the effectiveness of the proposed parallel lattice attention design.

F. Qualitative Results

We compared our model's results with those of several state-of-the-art models, with the findings presented in Figs. 7 and 8 for a $4\times$ scaling factor. The comparison highlights our model's superior performance, particularly in delineating bridge edges and viaduct markings, demonstrating its structural clarity and effectiveness in preserving fine details at higher magnification levels. We validated our model against state-of-the-art models using color error map analysis, as shown in Figs. 7 and 8. Visual comparisons, as shown in the rectangular box, indicate that our model achieves superior results. The proposed method better preserves fine structural details and produces lower color reconstruction errors in the highlighted regions compared to existing SR methods.

The experiments demonstrate that the classical bicubic method attenuates high-frequency details. While EDSR [5] improves performance by increasing depth, it underutilizes low-level features. Similarly, attention-based and lattice frameworks, including SwinIR [18], FENet [10], OmniSR [9], and BMFENet [31], struggle to capture fine details or suffer from limited receptive fields. Hybrid models, such as ESRT [16], MSCT [40], and ESTNet [43], combine CNNs with transformer architectures; however, they lack the interaction between low-level and high-level features that is vital for effective remote-sensing image reconstruction. In contrast, the proposed PLAN model preserves both fine-grained and coarse details, providing sharper edges and richer textures than competing methods.

G. Experiments on Nonsynthetic Dataset

To evaluate the effectiveness of the proposed PLAN model under real-world conditions, we conducted experiments on the UC Merced dataset at $4\times$ scaling without applying simulated degradation. We assessed performance using the average gradient (AG) and Natural Image Quality Evaluator (NIQE) [54]. The AG metric employs the Sobel operator to measure edge sharpness (where higher values are better), while NIQE uses a multivariate Gaussian model to assess perceptual quality (where lower values are better). Table IX presents a quantitative comparison with benchmark methods. To ensure robustness, we conducted paired two-tailed t-tests on per-image AG and NIQE scores, confirming that PLAN's improvements are statistically significant. The lower NIQE scores demonstrate that our method restores perceptual quality consistent with human vision, while the higher AG scores highlight its superior ability to reconstruct fine edges and textures. These findings are visually corroborated in Fig. 9, where the proposed network resolves clearer lines and textures in the tennis court scene.

V. Conclusion

In this article, we present PLAN, an RSISR network that achieves superior image quality with efficient computation. PLAN uses LAUnit to enhance features, extracting features with rich edges, textures, and contours. After that, parallel lattice units merge information separately to process complex features, while an attention mechanism captures significant context from remote-sensing images. Extensive experiments validate the effectiveness and robustness of PLAN for high-resolution remote sensing applications, significantly enhancing visual quality. Our approach performs well on images of size 600×600 pixels; however, real-world remote sensing applications may require processing significantly larger regions to avoid boundary artifacts introduced by patch-based processing. Future work will explore strategies like model optimization to enhance scalability.

Acknowledgment

We extend our appreciation to King Saud University for funding this research through the Ongoing Research Funding Program (ORF-2026-890).

References

[1] J. Shermeyer and A. Van Etten, “The effects of super-resolution on object detection performance in satellite imagery,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 1432–1441.

[2] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[3] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[4] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 391–407.

[5] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144.

[6] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016, pp. 1646–1654.

[7] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Fast and accurate image super-resolution with deep laplacian pyramid networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2599–2613, 2018.

[8] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.

[9] H. Wang, X. Chen, B. Ni, Y. Liu, and J. Liu, “Omni aggregation networks for lightweight image super-resolution,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 378–22 387.

[10] Z. Wang, L. Li, Y. Xue, C. Jiang, J. Wang, K. Sun, and H. Ma, “FeNet: Feature enhancement network for lightweight remote-sensing image super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[11] S. Wang, T. Zhou, Y. Lu, and H. Di, “Contextual transformation network for lightweight remote-sensing image super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[12] S. Lei and Z. Shi, “Hybrid-scale self-similarity exploitation for remote sensing image super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2022.

[13] J. Wang, Y. Lu, S. Wang, B. Wang, X. Wang, and T. Long, “Two-stage spatial-frequency joint learning for large-factor remote sensing image super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[14] N. Sultan, A. Hajian, and S. Aramvith, “An advanced features extraction module for remote sensing image super-resolution,” in *2024 21st International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, 2024, pp. 1–6.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[16] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, and T. Zeng, “Transformer for single image super-resolution,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 457–466.

[17] S. Lei, Z. Shi, and W. Mo, “Transformer-based multistage enhancement for remote sensing image super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[18] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “SwinIR: Image restoration using swin transformer,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1833–1844.

[19] C. Xie, X. Zhang, L. Li, Y. Fu, B. Gong, T. Li, and K. Zhang, “MAT: Multi-range attention transformer for efficient image super-resolution,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[20] L. Han, Y. Zhao, H. Lv, Y. Zhang, H. Liu, G. Bi, and Q. Han, “Enhancing remote sensing image super-resolution with efficient hybrid conditional diffusion model,” *Remote Sensing*, vol. 15, no. 13, p. 3452, 2023.

[21] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin, and L. Zhang, “EDiffSR: An efficient diffusion probabilistic model for remote sensing image super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[22] J. Li, Y. Yuan, K. Mei, and F. Fang, “Lightweight and accurate recursive fractal network for image super-resolution,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 3814–3823.

[23] K. Zhang, L. V. Gool, and R. Timofte, “Deep unfolding network for image super-resolution,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3217–3226.

[24] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, “Feedback network for image super-resolution,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3867–3876.

[25] D. Kong, L. Gu, X. Li, and F. Gao, “Multiscale residual dense network for the super-resolution of remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–12, 2024.

[26] W. Ye, B. Lin, J. Lao, Y. Liu, and Z. Lin, “Mra-idn: A lightweight super-resolution framework of remote sensing images based on multiscale residual attention fusion mechanism,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 7781–7800, 2024.

[27] X. Luo, Y. Xie, Y. Zhang, Y. Qu, C. Li, and Y. Fu, “Latticenet: Towards lightweight image super-resolution with lattice block,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*. Springer, 2020, pp. 272–289.

[28] M. Nikzad, Y. Gao, and J. Zhou, “An attention-based lattice network for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[29] W.-Y. Hsu and Y.-Y. Hsu, “Multi-scale and multi-layer lattice transformer for underwater image enhancement,” *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 11, pp. 1–24, 2024.

[30] X. Luo, Y. Qu, Y. Xie, Y. Zhang, C. Li, and Y. Fu, “Lattice network for lightweight image restoration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4826–4842, 2023.

[31] T. Wu, R. Zhao, M. Lv, Z. Jia, L. Li, Z. Wang, and H. Ma, “Lightweight remote sensing image super-resolution via background-based multiscale feature enhancement network,” *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.

[32] M. Cheng, H. Ma, Q. Ma, X. Sun, W. Li, Z. Zhang, X. Sheng, S. Zhao, J. Li, and L. Zhang, “Hybrid transformer and CNN attention network for stereo image super-resolution,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1702–1711.

[33] J. Yoo, T. Kim, S. Lee, S. H. Kim, H. Lee, and T. H. Kim, “Enriched CNN-transformer feature aggregation networks for super-resolution,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4956–4965.

[34] J. Talreja, S. Aramvith, and T. Onoye, "DHTCUN: Deep hybrid transformer CNN U network for single-image super-resolution," *IEEE Access*, pp. 122 624–122 641, 2024.

[35] C. Lin, X. Mao, C. Qiu, and L. Zou, "DTCNet: Transformer-CNN distillation for super-resolution of remote sensing image," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[36] C. Zhang, J. Wang, Y. Shi, B. Yin, and N. Ling, "A cnn-transformer hybrid network with selective fusion and dual attention for image super-resolution," *Multimedia Systems*, vol. 31, no. 2, pp. 1–17, 2025.

[37] J. Wang, B. Wang, X. Wang, Y. Zhao, and T. Long, "Hybrid attention-based U-shaped network for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[38] Y. Qin, J. Wang, S. Cao, M. Zhu, J. Sun, Z. Hao, and X. Jiang, "SRBPSwin: Single-image super-resolution for remote sensing images using a global residual multi-attention hybrid back-projection network based on the swin transformer," *Remote Sensing*, vol. 16, no. 12, p. 2252, 2024.

[39] Y. Tang, T. Wang, and D. Liu, "MFFAGAN: Generative adversarial network with multilevel feature fusion attention mechanism for remote sensing image super-resolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 6860–6874, 2024.

[40] X. Yao, H. Zhang, S. Wen, Z. Shi, and Z. Jiang, "Single-image super resolution for rgb remote sensing imagery via multi-scale CNN-transformer feature fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–16, 2024.

[41] A. Patnaik, M. K. Bhuyan, and K. F. MacDorman, "A two-branch multiscale residual attention network for single image super-resolution in remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 6003–6013, 2024.

[42] K. Zhang, L. Li, L. Jiao, X. Liu, W. Ma, F. Liu, and S. Yang, "Csct: Channel–spatial coherent transformer for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–14, 2025.

[43] X. Kang, P. Duan, J. Li, and S. Li, "Efficient swin transformer for remote sensing image super-resolution," *IEEE Transactions on Image Processing*, vol. 33, pp. 6367–6379, 2024.

[44] Y. Kang, X. Wang, X. Zhang, S. Wang, and G. Jin, "ACT-SR: Aggregation connection transformer for remote sensing image super-resolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 8953–8964, 2025.

[45] Y. Kang, X. Zhang, S. Wang, and G. Jin, "SCAT: Shift Channel Attention Transformer for Remote Sensing Image Super-Resolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[46] H. Li, W. Deng, Q. Zhu, Q. Guan, and J. Luo, "Local-global context-aware generative dual-region adversarial networks for remote sensing scene image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[47] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[48] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.

[49] D. Dai and W. Yang, "Satellite image classification via two-layer sparse coding with biased image representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 8, no. 1, pp. 173–176, 2011.

[50] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.

[51] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[52] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*, 2017.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

[54] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

Allen Patnaik received the B.Tech. degree in Electronics and Communication Engineering from the Gandhi Institute for Technology, Odisha, India, in 2014, and the M.Tech. degree in Communication Systems Engineering from the Veer Surendra Sai University of Technology, Odisha, India, in 2018. He completed his Ph.D. in Signal Processing and Machine Learning with the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, India. His research interests include computer vision, deep learning, and remote sensing. He is currently working as a Research Staff Member at the AI Center of Excellence, Indian Institute of Science, Bangalore, India.

Arpit Gour received an M.Tech degree in signal processing and machine learning from IIT Guwahati in 2024. His research interests include deep learning and machine learning. He is currently working in the AI and data science field.

M.K. Bhuyan (M'2010–SM'2013) received his Ph.D. in electronics and communication engineering from the Indian Institute of Technology (IIT) Guwahati, India. He is currently a Professor with the Department of Electronics and Electrical Engineering, IIT Guwahati. Additionally, he is a Visiting Professor at Chubu University, Japan, and an associate faculty member of the Mehta Family School of Data Science and Artificial Intelligence. Previously, he conducted postdoctoral research at the University of Queensland and NICTA, Australia. He served as an Assistant Professor at IIT Roorkee and Jorhat Engineering College, and worked in Indian Engineering Services. In 2014–2015, he was a Visiting Professor at Indiana University and Purdue University, USA.

Dr. Bhuyan has over 30 years of experience and more than 300 publications, including the textbook *Computer Vision and Image Processing* (CRC Press). His research spans machine learning, AI, computer vision, HCI, VR/AR, and biomedical signal processing. A Fellow of IETE and IEI, he is a recipient of the National Award for Best Applied Research (2012), presented by the President of India, the Fulbright-Nehru Academic and Professional Excellence Fellowship, and the BOYSCAST Fellowship.

Karl F. MacDorman (M'1999–SM'2007) received the B.A. degree in computer science from the University of California, Berkeley, USA, in 1988, and the Ph.D. degree in computer science from the University of Cambridge, U.K., in 1997. Since 2005, he has been an Associate Professor in the Luddy School of Informatics, Computing, and Engineering at Indiana University Indianapolis, USA, where he has served as Associate Dean of Academic Affairs since 2013. He has been an Honorary Professor at the Indian Institute of Technology Guwahati, India, since 2020. He previously served as Assistant Professor and later Associate Professor at Osaka University, Japan. Dr. MacDorman has published over 130 articles on human–robot interaction, machine learning, and cognitive science. In 2005, he co-organized founding workshops on the uncanny valley and conducted the first empirical study on the topic. Stanford University ranks him among the top 2% of scientists globally.

Sultan Alfarhood received his Ph.D. degree in computer science from the University of Arkansas. He is currently an Associate Professor with the Department of Computer Science, King Saud University (KSU). Since joining KSU in 2007, he has made contributions to the field through his research in machine learning, recommender systems, linked open data, text mining, and ML-based IoT systems. His recent work focuses on applying deep learning techniques to enhance the accuracy and effectiveness of these systems, and he has published in several high-impact journals and conferences.

Mejdl Safran received his B.S. degree in computer science from King Saud University, Riyadh, Saudi Arabia, in 2007, and his M.S. and Ph.D. degrees in computer science from Southern Illinois University Carbondale, Carbondale, IL, USA, in 2013 and 2018, respectively. He is currently an Associate Professor of Computer Science with King Saud University, where he has been a faculty member since 2008. Since 2018, he has served as an AI consultant for several national and international agencies. He has made contributions to the field through leading grant projects in AI in medical imaging and smart farming. He has published about 100 articles in peer-reviewed journals and conference proceedings. His current research interests include developing novel deep learning methods for image processing, pattern recognition, natural language processing, predictive analytics, and modeling user behavior in online platforms.