Edge-Focused Super-Resolution for Omnidirectional Images with Spherical Geometric Augmentation

Shaolin Wang[1*] Yuying Li[1*] Lei Zhong[2] Shigang Li[3] Jianfeng Li[1†]
[1]College of Electronic and Information Engineering, Southwest University, China
[2]University of Edinburgh, United Kingdom
[3]Graduate School of Information sciences, Hiroshima City University, Japan

Abstract

Omnidirectional image super-resolution (ODISR) remains challenging due to extreme magnification factors (e.g., 8×, 16×) and projection-specific distortions, which degrade edge integrity and limit model performance. This paper proposes an edge-focused framework combined with spherical geometric augmentation to address these issues. Our approach includes an Edge Focused Block (EFB) that integrates spatial-channel attention via Edge Enhanced and Refined Blocks, strengthening edge feature capture and optimization. We also design an Edge-Aware Multi-Scale (EAM) pipeline, leveraging shallow convolutions for initial feature extraction, local modules for deep mining, and a Global Integration Block for multi-scale aggregation, ensuring coherent edge reconstruction in distorted regions. To mitigate data scarcity, we introduce a rotation-translation augmentation strategy based on spherical projections, expanding datasets while preserving scene continuity. Extensive experiments show our method outperforms state-of-the-art approaches on public datasets.

1. Introduction

Limited by storage, transmission and bandwidth, current ODIs are often low-resolution, while users require high-definition details for head-mounted displays. Therefore, super-resolution of omnidirectional images is essential.

In recent years, deep learning (DL) has significantly advanced single image super-resolution (SISR). Researchers have continuously improved super-resolution (SR) performance through various approaches, including convolutional neural networks (CNNs) [14, 16, 23, 37], generative adversarial networks (GANs) [24, 25, 28, 36, 38], Vision Transformers (ViTs) [1, 5, 9, 13, 29, 31, 32, 34], and diffusion models [17, 33]. Meanwhile, for the task of super-resolving

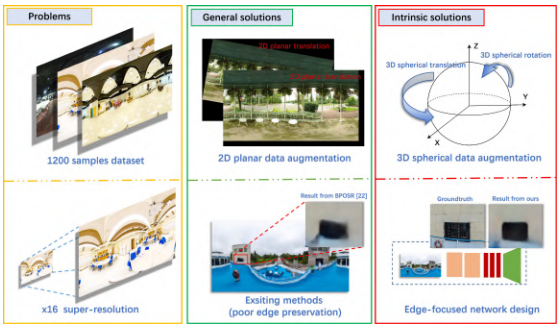

Figure 1. Challenges of ODISR and our proposed solutions.

omnidirectional images with special imaging projections, various approaches have been explored. Some works introduce multi-projection fusion [2, 3, 6, 7, 10, 11, 15, 18, 21, 22, 27, 30–32, 35], combining planar and spherical projections to enhance the information volume for network learning. Others focus on the uneven projection characteristics, employing deformable convolutions or spatial partitioning [2, 7, 19, 20, 22, 29] to achieve adaptive region-wise processing and improve reconstruction quality.

As shown in Fig.1, for extreme magnifications of 8× and 16×, omnidirectional image super-resolution (ODISR) confronts two distinct key challenges: First, scarcity of public dataset samples: The task relies on a public dataset with only 1200 samples. Second, poor edge preservation of existing methods: Existing methods lack edge-focused designs. And due to their complex network architectures, they adopt a patch or region-based input and output stitching design, resulting in poor edge presevation. To address these, we propose a geometry-correct 3D spherical data augmentation paradigm (rotation + translation) to expand data diversity, and an end-to-end lightweight Edge-Aware Multi-scale (EAM) pipeline to enhance edge preservation.

Specifically, we employ a spherical projection model to map 2D images to 3D space, and then perform 3D translation and rotation on entire scenes to augment the original

*Equal contribution
†Corresponding author

dataset. We use spherical coordinate transformation and rotation matrices around the X, Y, Z axes for spatial transformation. Translation is performed along the Z-axis, while rotation is only applied around the X and Y axes. This approach conforms to spherical geometric properties, effectively preserving scene continuity and integrity, and provides richer samples for model training. In summary, our contributions are threefold:

- We propose a rotation-translation augmentation strategy based on spherical properties. By simulating geometric transformations of omnidirectional images in spherical space, it enhances data diversity and improves the model's adaptability to projection distortions.
- We design an Edge Focused Block (EFB), comprising an Edge Enhanced Block (EEB) and an Edge Refined Block (ERB). It fuses spatial and channel attention to strengthen the capture and optimization of edge features in omnidirectional images, improving edge detail reconstruction quality.
- We propose an Edge-Aware Multi-Scale (EAM) pipeline for edge optimization via local modules. The Global Integration Block aggregates multi-scale features, and progressive upsampling generates high-resolution omnidirectional images. We further employ multi-loss optimization to boost super-resolution performance.

2. Related Work

Single Image Super-Resolution (SISR). With deep learning advances, significant progress has been made in single image super-resolution (SISR), outperforming traditional algorithms. In convolutional neural networks (CNNs), SRCNN [8] pioneered deep learning for end-to-end SISR, though its simplicity limits context utilization. Then EDSR [14] boosted performance via complex residual blocks and removed batch normalization. RCAN [37] further improved results by integrating channel attention. Beyond CNNs, generative adversarial networks (GANs) have also played a prominent role in this field. SRGAN [12] first applied GANs to SISR, using VGG loss instead of traditional MSE to enhance perceptual quality, inspiring works like ESRGAN [25]. In recent years, Vision Transformers (ViTs) offered new SISR approaches: IPT [4] (ViT-like) learned stronger features via pre-training on large datasets and low-level tasks. SwinIR [13], based on Swin Transformer, achieved higher performance with fewer parameters by using shifted window mechanisms to handle long-range dependencies. Diffusion models have also emerged: Resshift [33] enabled efficient super-resolution via residual shifting, with good real-scene robustness, providing a new way to address complex degradation. These methods continuously advance SISR performance from various angles, driving the field forward and laying foundations for future research.

Omnidirectional Image Super-Resolution (ODISR). Initially, ODISR research focused on adapting to different projection modes, exploring super-resolution paths matching the geometric features of low-resolution ODIs by integrating features from spherical, planar and other projection spaces, with emphasis on solving information loss during projection conversion. LAU-Net [7] divides ERP images into patches by latitude to learn ERP distortions in different latitude ranges for the non-uniformity of ODIs. However, non-overlapping patch processing causes discontinuous image information and obvious faults between patches, limiting super-resolution improvement. SphereSR [31] breaks through the limitation that traditional methods only target fixed projection types, constructs continuous spherical image representations, and realizes super-resolution reconstruction under any 360° image projection type through feature extraction modules and spherical local implicit image functions (SLIF). Although it improves the information consistency between different projection types and performs well, its computational complexity is high. OSRT [32] innovatively proposes a fisheye downsampling method for ODISR to generate low-resolution samples closer to real scenes. Its distortion-aware Transformer learns offsets through latitude conditional learning for adaptive modulation of ERP distortion, and synthesizes pseudo-ERP images from ordinary images to expand data and alleviate overfitting, achieving excellent results on multiple datasets. But it still lacks in complex edge details and edge continuity in polar regions.

Current ODISR methods have advanced in multi-projection fusion and region-adaptive learning, yet still suffer from poor global continuity and insufficient edge preservation in large-scale super-resolution. Leveraging the spherical geometry of omnidirectional images and the significance of edges, we optimize data augmentation to fit spherical structures, enhance edge perception and reconstruction, and improve large-scale super-resolution performance and visual coherence in highly distorted regions. Experiments demonstrate that this method yields excellent results, offering a robust solution for ODISR.

3. Method

3.1. Dataset augmentation

In ODIs research, existing datasets such as ODI-SR and SUN360 are constrained by limited sample sizes and the absence of data augmentation techniques that properly preserve spherical geometry. When conventional 2D image augmentation operations like rotation and translation are directly applied to omnidirectional images, they cause distortions including edge warping and disruption of spherical topology. These artifacts, such as unnatural stretching near the equator and compression at the poles, fail to main-

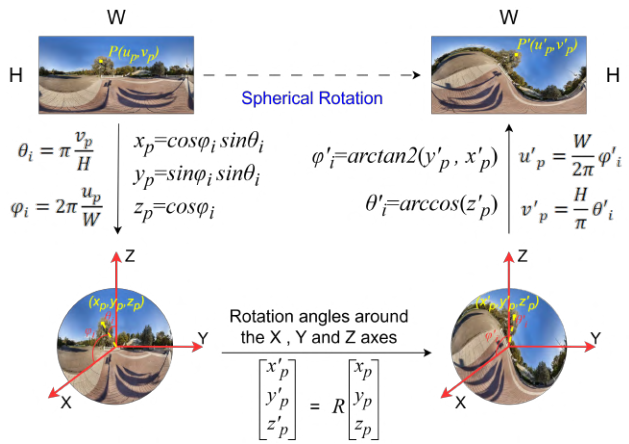

Figure 2. Mathematical calculation of data augmentation based on spherical model.

tain realistic perspective transformations in omnidirectional scenes.

Spherical omnidirectional images are 2D projections of 3D spherical scenes, with pixels following spherical geometry. Thus, data augmentation should use spherical coordinate transformations to keep scenes consistent and rational, helping super-resolution models better recover edges and textures. We propose a spherical coordinate-based method for this, improving model generalization by simulating various views. It uses equirectangular projection properties of these images and achieves diverse augmentation via 3D rotations. As in Fig.2, the mapping between 2D coordinates (u, v) of input omnidirectional image $I_i(u, v)$ and their 3D spherical coordinates (x_p, y_p, z_p) is defined as:

$$\varphi_i = 2\pi \frac{u}{W}, \quad \theta_i = \pi \frac{v}{H} \tag{1}$$

Where W and H represent the width and height of the omnidirectional image respectively, φ_i denotes the azimuth angle, and θ_i is the polar angle.

$$x_p = \cos\varphi_i \sin\theta_i, \;\; y_p = \sin\varphi_i \sin\theta_i, \;\; z_p = \cos\theta_i \tag{2}$$

Rotation and translation operations for omnidirectional images are achieved by performing 3D coordinate rotations about the X, Y or Z axes.

The rotation matrices $R_x(\alpha)$, $R_y(\beta)$, and $R_z(\gamma)$ for rotations about the X, Y and Z axes are respectively given by:

$$\begin{aligned} R_x(\alpha) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \\ R_y(\beta) &= \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \\ R_z(\gamma) &= \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{3}$$

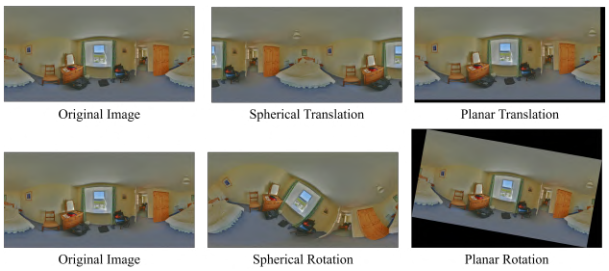

Figure 3. Comparison of omnidirectional image translation and rotation operations.

Where $\alpha, \beta \in \left[-\frac{\pi}{12}, \frac{\pi}{12}\right], \gamma \in [0, 2\pi]$.

The rotated coordinates (x'_p, y'_p, z'_p) can be obtained by matrix multiplication of the rotation matrix $\boldsymbol{R}$ with the original coordinates (x_p, y_p, z_p). When $\boldsymbol{R}$ takes the rotation matrices $R_x(\alpha)$, $R_y(\beta)$, and $R_z(\gamma)$ corresponding to rotations around the X, Y, and Z axes by angles α, β, and γ respectively, we have:

$$\begin{bmatrix} x'_p \\ y'_p \\ z'_p \end{bmatrix} = R \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} \tag{4}$$

At this point, the changes in the azimuth and polar angle coordinates are as follows:

$$\varphi'_i = \arctan 2(y'_p, x'_p), \quad \theta'_i = \arccos(z'_p) \tag{5}$$

The final coordinates mapped back to the 2D image space are:

$$u'_p = \frac{W}{2\pi}\varphi'_i, \quad v'_p = \frac{H}{\pi}\theta'_i \tag{6}$$

Both translation and rotation of panoramic images are implemented by applying rotation transformations to the original 3D spherical coordinates (x_p, y_p, z_p): translation is rotation around the Z-axis by $\gamma \in [0, 2\pi]$ using matrix $R_z(\gamma)$, while rotation is performed around either the X-axis by $\alpha \in \left[-\frac{\pi}{12}, \frac{\pi}{12}\right]$ with matrix $R_x(\alpha)$ or the Y-axis by $\beta \in \left[-\frac{\pi}{12}, \frac{\pi}{12}\right]$ with matrix $R_y(\beta)$. The new azimuth φ'_i and polar θ'_i angles are computed via Eq.(5), and finally mapped to 2D image space using Eq.(6).

Compared to traditional flat-image translation and rotation, as shown in Fig.3, spherical translation preserves the 360° ring-shaped structure without edge truncation or padding, ensuring natural transitions. Spherical rotation follows spherical geometry, maintaining geometric integrity and panoramic continuity. Even under large-angle rotations, key features remain undistorted. Overall, it outperforms traditional methods, which often suffer from truncation, padding, edge distortion, and disruption of the ring-shaped topology.

3.2. Network Architecture

In super-resolution, image boundaries and contours are core to visual semantics; their integrity and smoothness di-

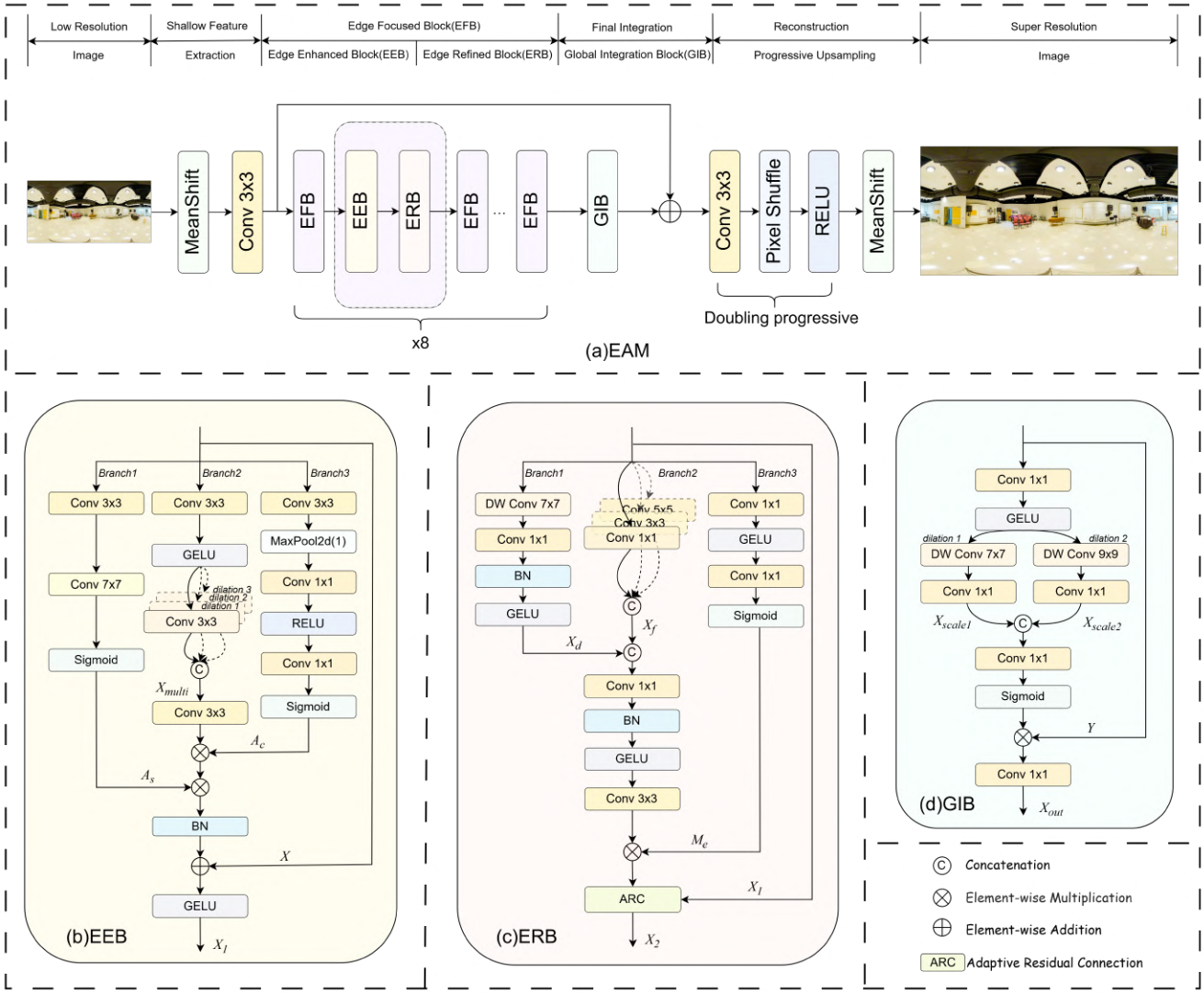

Figure 4. (a) The network architecture of the proposed EAM; (b) Illustration of the proposed Edge Enhanced Block(EEB); (c) Illustration of the proposed Edge Refined Block(ERB); (d) Illustration of the proposed Global Integration Block(GIB).

rectly determine result quality. Thus, network design focuses on "boundary preservation and refinement", leading to our Edge-Aware Multi-Scale super-resolution network (EAM).

As shown in Fig.4. EAM comprises four parts: shallow feature extraction, Edge-Focused Block (EFB), Global Integration Block (GIB), and super-resolution image reconstruction. In super-resolution, existing methods usually avoid feature downsampling, since super-resolution essentially recovers high-frequency details from low-resolution images; downsampling loses key spatial and edge features, making subsequent reconstruction fail to accurately restore boundaries and contours. For input low-resolution image $I_{\mathrm{LR}} \in \mathbb{R}^{3\times H\times W}$, the shallow module first removes brightness bias via Meanshift, then uses 3×3 convolution for initial feature mapping to retain the original spatial structure and basic edge information, extracting shallow features F_p for subsequent processing, as follows:

$$F_{\mathrm{p}} = \mathrm{Conv}_{3\times3}(\mathrm{MeanShift}(I_{\mathrm{LR}})) \tag{7}$$

Natural image structures (e.g., edges, textures) have scale links; local details depend on global contours. Thus we propose EFB, which enhances and refines edges to optimize from local boundary repair to global feature consistency, preserving same-scale features. EFB has two parts: EEB uses multi-scale dilated convolutions to capture cross-range edges and strengthens boundaries via attention; ERB optimizes enhanced edges, adjusts feature maps with spatial attention, boosting refinement. F_p enters cascaded EFB for

extraction, yielding F_r.

$$F_r = \mathrm{EFB}(F_p) \tag{8}$$

Multiple cascaded EFBs may over-focus on local features and lack global correlation, so the Global Integration Block (GIB) is designed: it expands the receptive field via large-kernel depthwise convolution to capture long-range feature dependencies, introduces an attention mechanism to enhance regions crucial to the global structure, ensures coordination between local edges and overall contours, and outputs integrated same-scale features:

$$F_l = \mathrm{GIB}(F_r) \tag{9}$$

Here, F_l is the final integrated deep feature from GIB. The reconstruction adopts a progressive upsampling strategy: the target magnification is decomposed into multiple 2× upsampling steps. Each step uses PixelShuffle to increase resolution and combines 3×3 convolutions to supplement high-frequency details at that scale, avoiding edge blurring and detail loss caused by single-step large-scale upsampling. Finally, a 1×1 convolution maps the features back to 3 channels, and inverse MeanShift is applied to restore the original image brightness range, outputting the high-resolution image $I_{\mathrm{SR}} \in \mathbb{R}^{3\times\alpha H\times\alpha W}$ (where α is the scaling factor). This work uses two scaling factors: $\alpha = 8$ and $\alpha = 16$.

3.3. Edge Enhanced Block(EEB)

In super-resolution tasks, low-resolution images have blurred edge information, and traditional convolutions struggle to effectively distinguish real edges. To address this, we design the Edge Enhanced Block (EEB), which combines multi-scale feature extraction with edge-aware attention mechanisms to achieve targeted edge enhancement. For input features $X \in \mathbb{R}^{C\times H\times W}$ from the shallow feature extraction module, Fig.4 (b) illustrates the computational flow of EEB, comprising four key components: multi-scale feature extraction, edge-aware channel attention, edge-aware spatial attention, and feature fusion with residual connections.

Multi-scale feature extraction. To extract effective features, A multi-branch parallel structure is employed (Branch 2 in Fig.4 (b)). Initially, basic features are extracted via standard 3×3 convolution and processed with GELU activation. The base features are then processed using dilated convolutions with dilation rates of 1, 2 and 3 to capture multi-scale features. This design enables the network to perceive edge information at varying scales: small dilation rates focus on local fine edges, medium rates extract object contours, while large rates capture global structural features. The outputs of the three branches are concatenated along the channel dimension, forming fused features X_{multi} enriched with multi-scale information.

Edge-aware channel attention. To focus on the contributions of different channels, with particular emphasis on enhancing edge-related channels, we have designed a channel attention branch (Branch 3 in Fig.4 (b)). First, a 3×3 convolution is used to refine the features, then adaptive max pooling collapses the spatial dimensions to 1×1, thus eliminating spatial information. This step effectively shifts the focus entirely to channel dimensions. Subsequently, channel attention weights A_c are computed via two 1×1 convolutions and Sigmoid activation. The process is:

$$A_c = \sigma\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Conv}_{1\times1}(\mathrm{Pool}(\mathrm{Conv}_{3\times3}(X)))\right)\right) \tag{10}$$

Here, σ represents the Sigmoid activation function. and the resulting A_c effectively strengthens edge-aware channels.

Edge-aware spatial attention. To focus on the contributions of different spatial regions, with particular emphasis on edge continuity, we have designed a spatial attention branch via two-stage large-kernel convolution (Branch 1 in Fig.4 (b)). First, 3×3 convolution generates edge-informed feature maps, then 7×7 large kernel captures long-range spatial dependencies to address the limited receptive field of traditional small kernels. In this process, the number of channels will gradually decrease to 1, thus shifts the focus entirely to space dimensions. This design can model the edge continuity across larger spatial regions. Finally, Sigmoid activation produces spatial attention weights A_s, with the expression:

$$A_s = \sigma\left(\mathrm{Conv}_{7\times7}\left(\mathrm{Conv}_{3\times3}(X)\right)\right) \tag{11}$$

Feature fusion with residual connections. Finally, A gated residual connection strategy is adopted to integrate the attentions above. First, the multi-scale feature X_{multi} undergoes channel integration via 3×3 convolution, then is element-wise multiplied with channel attention weights A_c and spatial attention weights A_s, followed by batch normalization to obtain:

$$X_1 = X + \mathrm{BN}\left(\mathrm{Conv}_{3\times3}(X_{\mathrm{multi}}) \otimes A_c \otimes A_s\right) \tag{12}$$

where $\otimes$ denotes element-wise multiplication and BN stands for batch normalization. This dual attention mechanism adaptively enhances key region responses.

3.4. Edge Refined Block(ERB)

EEB effectively enhances edge information but requires further refinement for more accurate and delicate feature representation. Thus, we designed the Edge Refinement Block (ERB), which takes EEB-processed feature $X_1 \in \mathbb{R}^{C\times H\times W}$ as input. The specific architecture is shown in Fig.4 (c), and the computational workflow proceeds as follows:

The first branch (Branch 1 in Fig.4 (c)) enhances features with a larger receptive field: 7×7 depthwise convolution (groups=C) expands the receptive field while preserving channel independence to capture long-range spatial dependencies. It fuses cross-channel information via 1×1 convolution, combined with batch normalization and GELU activation, outputting large field features $X_d \in \mathbb{R}^{C\times H\times W}$.

The second branch (Branch 2 in Fig.4 (c)) is a multi-path convolution structure: 1×1, 3×3, and 5×5 convolutions capture point-level, local texture, and broad structural features respectively. Channel-wise concatenation reconstructs $X_f \in \mathbb{R}^{C\times H\times W}$, covering multi-level details to enrich representation.

Another branch (Branch 3 in Fig.4 (c)) generates weight features for edge reliability: 1×1 convolution compresses channels to half, enhanced by GELU for non-linearity, then converted to a single-channel map via 1×1 convolution. Output via Sigmoid as $M_e \in [0,1]^{1\times H\times W}$ to mark edge reliability.

For comprehensive feature processing, we designed a feature fusion and adaptive weight mechanism: concatenate X_d and X_f along the channel, compress via 1×1 convolution, apply batch normalization and GELU activation, then fuse through 3×3 convolution to obtain refined features $X_r \in \mathbb{R}^{C\times H\times W}$. An adaptive weight $\alpha \in [0,1]$ balances the original input and refined features, with the final output as their weighted sum, expressed as follows:

$$X_2 = (1-\alpha)\cdot X_1 + \alpha\cdot(X_r \otimes M_e) \tag{13}$$

Here, $\otimes$ is element-wise multiplication, and α is a learnable parameter, which dynamically adjusts weights to optimize edges while preserving non-edge features.

3.5. Global Integration Block(GIB)

In super-resolution reconstruction, image global semantics and local details are strongly correlated. Traditional convolutions, limited by fixed receptive fields, fail to balance fine details and long-range dependencies, easily causing local distortions or inconsistent global structures in results. Thus, we propose the Global Integration Block (GIB), which adaptively weights and integrates global features via multi-scale large-kernel convolutions and attention fusion to enhance feature representation robustness. For input features $Y \in \mathbb{R}^{C\times H\times W}$ processed by multiple EFB modules, its pipeline (Fig.4 (d)) has two core components: multi-scale context extraction and attention fusion.

Multi-scale context extraction. A dual-branch parallel structure captures cross-range contextual information. 1×1 convolution enhances input feature channel interaction, activated by GELU to get X_{init}, laying the foundation for subsequent channel representation. Branch 1 uses 7×7 depthwise convolution (groups=C) to capture contextual information within a moderate range, aggregated via 1×1 pointwise convolution to obtain X_{scale1}. Branch 2 employs 9×9 depthwise convolution with dilation rate=2 (groups=C) to establish a larger receptive field, following 1×1 pointwise convolution to get X_{scale2}. This design combines depthwise and pointwise convolutions, significantly expanding the receptive field while maintaining computational efficiency, enabling the network to focus on global structures.

Attention fusion. We adopt a channel-adaptive weight allocation mechanism to dynamically integrate multi-scale features. First, the feature X_{scale1} and the feature X_{scale2} are concatenated in the channel dimension. Then, 1×1 convolution is used to compress the dual-channel features to the original number of channels, and the attention weight A is generated after Sigmoid activation, which is expressed as:

$$A = \sigma\left(\text{Conv}_{1\times1}\left(\text{Concat}(X_{\text{scale1}}, X_{\text{scale2}})\right)\right) \tag{14}$$

Finally, the attention weight A is multiplied element-wise with the initial feature Y to achieve feature weighting. To further enhance the feature representation capability, a 1×1 convolution is used to refine the channel information of the weighted features, and the final output is X_{out}:

$$X_{\text{out}} = \text{Conv}_{1\times1}\left(Y \otimes A\right) \tag{15}$$

3.6. Training Loss

To address core issues like detail loss, edge blurring, and texture distortion in super-resolution, this paper adopts a multi-objective joint loss function. It enhances numerical accuracy, visual quality, and structural consistency through collaborative optimization at pixel, feature, and structural levels. The total loss function can be expressed as:

$$L_{\text{Total}} = L_{\text{L1}} + 0.01\times L_{\text{Perceptual}} + 0.1\times(1-L_{\text{SSIM}}) \tag{16}$$

L1 Loss ensures pixel mapping foundation and accelerates convergence. Perceptual Loss (weight 0.01) captures multi-scale high-level features via a pre-trained VGG network to optimize visual perception. SSIM Loss, derived from transforming the SSIM metric with a weight of 0.1, drives the model to optimize structural consistency and reduce misalignment. The selection of weight values adopts the common values for each loss function.

4. Experiments

4.1. Dataset and Implementation

Experiments use the ODI-SR and SUN360 datasets. For training, the ODI-SR dataset with images of 1024×2048 resolution is used and augmented based on the spherical model. Low-resolution images are generated by directly resizing high-resolution images with scaling factors of ×8 and ×16. The training process employs a combined loss function comprising L1 Loss, Perceptual Loss, and SSIM Loss, uses the Adam optimizer with an initial learning rate of 0.001, and a batch size of 4. For testing, we evaluated on ODI-SR and SUN360 datasets with identical preprocessing, using the official ODI-SR database metric code for WS-PSNR and WS-SSIM as evaluation metrics.

4.2. Quantitative Results

Tab.1 presents the quantitative comparison between our EAM network and other state-of-the-art methods on both datasets for ×8 and ×16 SR tasks. Our method demonstrates superior performance across all evaluation scenarios. On the ODI-SR dataset, our method achieves performance improvements of 1.15dB and 1.13dB in WS-PSNR metrics respectively compared to the representative FATO [1] method. Furthermore, our approach also delivers the best performance on the SUN360 dataset. These results strongly demonstrate our network's advantages in edge awareness and detail restoration, exhibiting superior reconstruction accuracy and visual quality.

Tab.2 presents the Floating-Point Operations (FLOPs), the number of parameters (Params), and the running time (Time) when our method (EAM) is compared with state-of-the-art methods on the ODI-SR dataset with a scale factor of ×16. It can be found that our method achieves a decent performance with lower computational complexity and faster running speed.

Table 1. Quantitative results (WS-PSNR/WS-SSIM) on ODI-SR and SUN360 datasets.

Dataset	ODI-SR				SUN360			
Scale	×8		×16		×8		×16	
Method	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM
Bicubic	19.64	0.5908	17.12	0.4332	19.72	0.5403	17.56	0.4638
EDSR[14]	23.97	0.6483	22.24	0.6090	23.79	0.6472	21.83	0.5974
RCAN[37]	24.26	0.6554	22.49	0.6176	23.88	0.6542	21.86	0.5938
360-SS[16]	21.65	0.6417	19.65	0.5431	21.48	0.6352	19.62	0.5308
LAU-Net[7]	24.36	0.6602	22.52	0.6284	24.24	0.6708	22.05	0.6058
SphereSR[31]	24.37	0.6777	22.51	0.6370	24.17	0.6820	21.95	0.6342
OSRT[32]	24.53	0.6780	22.69	0.6261	24.38	0.7072	22.13	0.6388
BPOSR[22]	24.61	0.6782	22.72	0.6285	24.47	0.7084	22.16	0.6433
FATO[1]	24.54	0.6784	22.73	0.6314	24.42	0.7120	22.18	0.6449
LAPR[2]	24.72	0.6886	22.90	0.6480	24.53	0.6885	22.37	0.6475
GDGT-OSR[29]	24.60	0.6687	22.78	0.6087	25.00	0.7068	22.60	0.6303
MambaOSR[26]	24.62	0.6792	22.66	0.6293	24.49	0.7119	22.12	0.6452
EAM(Ours)	25.69	0.6839	23.86	0.6290	25.81	0.7102	23.49	0.6415

For fairness, we ran our WS-PSNR and WS-SSIM using the ODI-SR database official metric code, which is consistent with that of LAU-Net, OSRT.

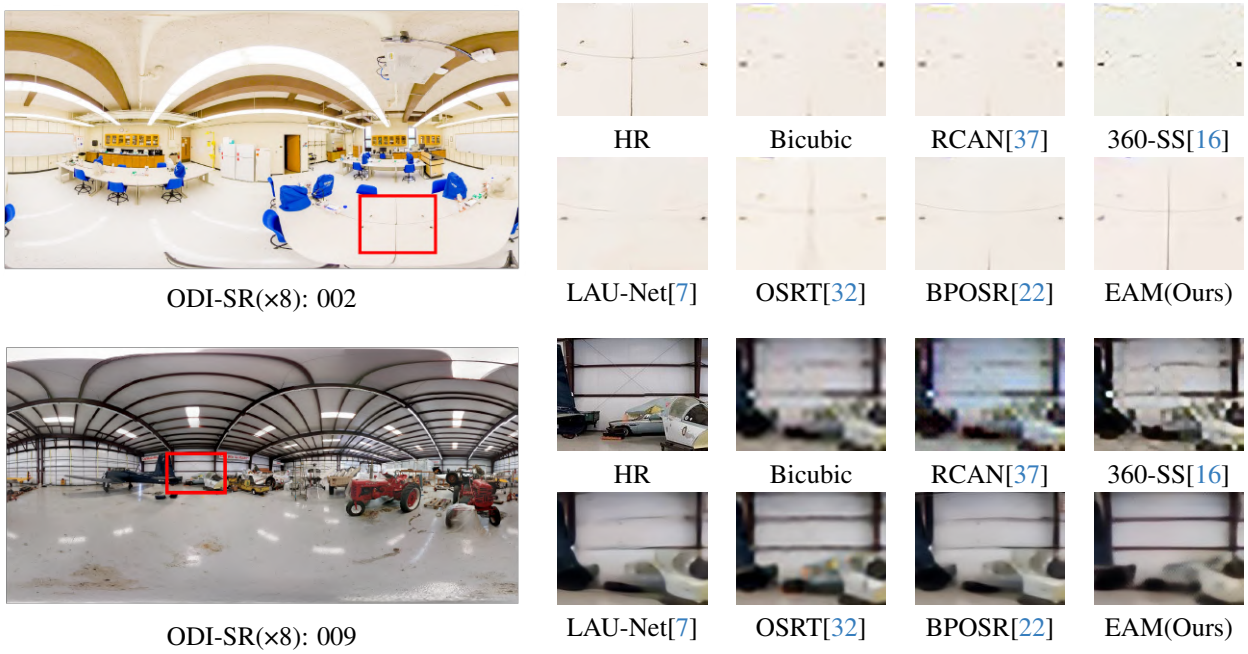

Figure 5. Visual comparisons of ×8 SR results of different methods on ODI-SR dataset.

4.3. Qualitative Results

Fig.5 shows the visual comparison results from the ODI-SR dataset at ×8 SR task, presenting both the complete HR image and cropped local regions for detailed analysis. It should be noted that some comparative methods are included in this visualization, namely RCAN [37], 360-SS [16], LAU-Net [7], OSRT [32], and BPOSR [22], while others could not be reproduced due to unavailable code or pretrained models. The results reveal that RCAN, 360-SS and LAU-Net exhibit noticeable edge distortion, whereas OSRT and BPOSR show relatively better performance but still contain minor artifacts like tilting or edge fractures. In comparison, our EAM network demonstrates superior reconstruction quality by effectively preserving edge continuity, rendering finer texture details, and achieving more natural restoration of complex structures, which collectively validate its advanced detail recovery capability.

4.4. Ablation Study and Analysis

We perform ablation experiments under consistent settings to analyze the model. Only one component is varied at a time, with

Table 2. Computational efficiency comparison across methods for ×16 super-resolution on the ODI-SR dataset.

Model	FLOPs	Params	Time
SwinIR[13]	900 G	11.5 M	0.982 s
360-SS[16]	**15 G**	**1.6 M**	0.025 s
LAU-Net[7]	685 G	9.4 M	0.443 s
SphereSR[31]	587 G	8.7 M	0.401 s
LAPR[2]	372 G	7.8 M	0.312 s
EAM (Ours)	38 G	2.0 M	**0.022 s**

Figure 6. Visual effects of data augmentation. For ODI-SR(×8): 022 and ODI-SR(×16): 027, (a) shows the effect of the augmented ODI-SR dataset, and (b) shows the effect of the original ODI-SR dataset.

evaluation on the ODI-SR dataset.

Table 3. Ablation study on Data augmentation. All models are trained on ×8 and ×16 SR task on ODI-SR dataset.

ODI-SR	Scale	WS-PSNR	WS-SSIM
Original	×8	24.70	0.6529
Augmented		**25.69**	**0.6839**
Original	×16	23.10	0.6131
Augmented		**23.86**	**0.6290**

Data augmentation. As shown in Fig.6 and Tab.3, the proposed data augmentation method based on spherical coordinate transformation achieves significant improvements. Visually, the augmented images exhibit higher clarity, richer details, sharper edges, and more natural colors, while the original results suffer from blurred details and rigid edges. Quantitatively, for the ×8 task, WS-PSNR increases from 24.70 dB to 25.69 dB and WS-SSIM from 0.6529 to 0.6839, and similar gains are obtained for the ×16 task. This method improves generalization, alleviates overfitting, and enhances feature robustness, providing new insights for omnidirectional image super-resolution.

EAM components. As shown in Tab.4 , each module in the EAM network effectively improves super-resolution performance. Removing a single module causes only minor drops in WS-PSNR and WS-SSIM, within one decimal place. This aligns with the common pattern in ODISR where performance gains often manifest as such subtle improvements. It indicates each module plays an indispensable role in enhancing edge features and integrating global information, collectively boosting model performance and validating the rationality of the EAM network's overall design and the necessity of each component.

Table 4. Ablation study on EAM components.All models are trained on ×8 SR task on ODI-SR dataset.

EEB	ERB	GIB	ODI-SR	
			WS-PSNR	WS-SSIM
×	✓	✓	25.17	0.6691
✓	×	✓	25.17	0.6699
✓	✓	×	24.95	0.6539
✓	✓	✓	**25.69**	**0.6839**

Table 5. Ablation study on Loss of EAM. All models are trained on ×8 SR task on ODI-SR dataset.

L1 Loss	SSIM Loss	Perceptual Loss	ODI-SR	
			WS-PSNR	WS-SSIM
✓	×	✓	25.29	0.6665
✓	✓	×	25.27	0.6690
✓	✓	✓	**25.69**	**0.6839**

Loss in EAM. As shown in Tab.5, in the design of EAM loss functions, combinations of different loss functions have complementary effects. The combination of L1 and perceptual loss highlights the necessity of SSIM. Since perceptual loss has weak constraints on geometric details, SSIM can achieve precise optimization. The combination of L1 and SSIM loss can improve the perceptual quality of reconstructed images, and perceptual loss can make up for the deficiency of L1 in structural coherence. In super-resolution tasks, perceptual loss and SSIM loss are complementary. The former enhances geometric fidelity, while the latter ensures semantic rationality, working together to achieve the unity of perceptual quality and structural accuracy.

5. Conclusion

In conclusion, to tackle ODISR challenges from extreme magnification and projection distortions, we propose an edge-focused framework with spherical geometric augmentation. It incorporates EFB, EAM pipeline, and rotation-translation augmentation, achieving superior performance over state-of-the-art methods as shown in experiments.

Acknowledgements

This work was supported in part by the Natural Science Foundation of Chongqing, China (CSTB2024NSCQ-MSX0437).

References

[1] Hongyu An, Xinfeng Zhang, Shijie Zhao, and Li Zhang. Fato: Frequency attention transformer for omnidirectional image super-resolution. In *Proceedings of the 6th ACM International Conference on Multimedia in Asia*, pages 1–7, 2024. 1, 6, 7

[2] Qing Cai, Mu Li, Dongwei Ren, Jun Lyu, Haiyong Zheng, Junyu Dong, and Yee-Hong Yang. Spherical pseudo-cylindrical representation for omnidirectional image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 873–881, 2024. 1, 7, 8

[3] Zhangjie Cao, Qixing Huang, and Ramani Karthik. 3d object classification via spherical projections. In *2017 international conference on 3D Vision (3DV)*, pages 566–574. IEEE, 2017. 1

[4] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12299–12310, 2021. 2

[5] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12312–12321, 2023. 1

[6] Jason Corso, Darius Burschka, and Gregory Hager. Direct plane tracking in stereo images for mobile navigation. In *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, pages 875–880. IEEE, 2003. 1

[7] Xin Deng, Hao Wang, Mai Xu, Yichen Guo, Yuhang Song, and Li Yang. Lau-net: Latitude adaptive upscaling network for omnidirectional image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9189–9198, 2021. 1, 2, 7, 8

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 2

[9] Debasish Dutta, Deepjyoti Chetia, Neeharika Sonowal, and Sanjib Kr Kalita. State-of-the-art transformer models for image super-resolution: Techniques, challenges, and applications. *arXiv preprint arXiv:2501.07855*, 2025. 1

[10] Ulas Gunes, Matias Turkulainen, Xuqian Ren, Arno Solin, Juho Kannala, and Esa Rahtu. Fiord: A fisheye indoor-outdoor dataset with lidar ground truth for 3d scene reconstruction and benchmarking. In *Scandinavian Conference on Image Analysis*, pages 3–17. Springer, 2025. 1

[11] Koji Koyamada and Takayuki Ito. Fast generation of spherical slicing surfaces for irregular volume rendering. *The Visual Computer*, 11(3):167–175, 1995. 1

[12] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2

[13] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1, 2, 8

[14] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 1, 2, 7

[15] Xuesi Ma, Ketao Zhang, and Jian S Dai. Novel spherical-planar and bennett-spherical 6r metamorphic linkages with reconfigurable motion branches. *Mechanism and Machine Theory*, 128:628–647, 2018. 1

[16] Cagri Ozcinar, Aakanksha Rana, and Aljosa Smolic. Super-resolution of omnidirectional images using adversarial learning. In *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2019. 1, 7, 8

[17] Hshmat Sahak, Daniel Watson, Chitwan Saharia, and David Fleet. Denoising diffusion probabilistic models for robust image super-resolution in the wild. *arXiv preprint arXiv:2302.07864*, 2023. 1

[18] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart. A flexible technique for accurate omnidirectional camera calibration and structure from motion. In *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, pages 45–45. IEEE, 2006. 1

[19] Fanjie Shang, Hongying Liu, Wanhao Ma, Yuanyuan Liu, Licheng Jiao, Fanhua Shang, Lijun Wang, and Zhenyu Zhou. Lightweight super-resolution with self-calibrated convolution for panoramic videos. *Sensors*, 23(1):392, 2022. 1

[20] Gyumin Shim, Jinsun Park, and In So Kweon. Robust reference-based super-resolution with similarity-aware deformable convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8425–8434, 2020. 1

[21] Junping Wang, An Liu, and Hyungsuck Cho. Direct path planning in image plane and tracking for visual servoing. In *Optomechatronic Systems Control III*, pages 44–51. SPIE, 2007. 1

[22] Jiangang Wang, Yuning Cui, Yawen Li, Wenqi Ren, and Xiaochun Cao. Omnidirectional image super-resolution via bi-projection fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5454–5462, 2024. 1, 7

[23] Li Wang, Ke Li, Jingjing Tang, and Yuying Liang. Image super-resolution via lightweight attention-directed feature aggregation network. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2):1–23, 2023. 1

[24] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 1

[25] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 1, 2

[26] Weilei Wen, Qianqian Zhao, and Xiuli Shao. Mambaosr: Leveraging spatial-frequency mamba for distortion-guided omnidirectional image super-resolution. *Entropy*, 27(4):446, 2025. 7

[27] LI Xiaohui, ZHOU Yinqing, and WANG Zulin. Spherical panorama creating algorithm based on curve surface mosaic. *JOURNAL-BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS*, 33(6):668, 2007. 1

[28] Hemant Yadav, Jalpesh Vasa, and Rudra Patel. Gan (generative adversarial network)-based image super-resolution: a technical perspective. In *International Conference on Information and Communication Technology for Intelligent Systems*, pages 283–293. Springer, 2023. 1

[29] Cuixin Yang, Rongkang Dong, Jun Xiao, Cong Zhang, Kin-Man Lam, Fei Zhou, and Guoping Qiu. Geometric distortion guided transformer for omnidirectional image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. 1, 7

[30] Ismail Enes Yigit and I Lazoglu. Spherical slicing method and its application on robotic additive manufacturing. *Progress in Additive Manufacturing*, 5(4):387–394, 2020. 1

[31] Youngho Yoon, Inchul Chung, Lin Wang, and Kuk-Jin Yoon. Spheresr: 360deg image super-resolution with arbitrary projection via continuous spherical image representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5677–5686, 2022. 1, 2, 7, 8

[32] Fanghua Yu, Xintao Wang, Mingdeng Cao, Gen Li, Ying Shan, and Chao Dong. Osrt: Omnidirectional image super-resolution with distortion-aware transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13283–13292, 2023. 1, 2, 7

[33] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36:13294–13307, 2023. 1, 2

[34] Dongxiao Zhang, Tangyao Qi, and Juhao Gao. Transformer-based image super-resolution and its lightweight. *Multimedia Tools and Applications*, 83(26):68625–68649, 2024. 1

[35] Saiping Zhang, Fuzheng Yang, Shuai Wan, and Peiyun Di. Spherical lanczos interpolation in planar projection or format conversions of panoramic videos. *IEEE Access*, 8:9655–9667, 2020. 1

[36] Xianming Zhang and Jiaojiao Feng. A novel blind restoration method for miner face images based on improved gfp-gan model. *IEEE Access*, 2024. 1

[37] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 1, 2, 7

[38] Xuan Zhu, Yue Cheng, Jinye Peng, Rongzhi Wang, Mingnan Le, and Xin Liu. Super-resolved image perceptual quality improvement via multifeature discriminators. *Journal of Electronic Imaging*, 29(1):013017–013017, 2020. 1