Fast Omni-Directional Image Super-Resolution: Adapting the Implicit Image Function with Pixel and Semantic-Wise Spherical Geometric Priors

Xuelin Shen[2*], Yitong Wang[1,2*], Silin Zheng[1], Kang Xiao[2], Wenhan Yang[3], Xu Wang[1†]

[1]College of Computer Science and Software Engineering, Shenzhen University
[2]Guangdong Laboratory of Artificial Intelligence and Digital Economy(SZ)
[3]Peng Cheng Laboratory
shenxuelin@gml.ac.cn, ted.yitongwang@gmail.com, zhengpny@gmail.com, xiaokang2022@email.szu.edu.cn, yangwh@pcl.ac.cn, wangxu@szu.edu.cn

Abstract

In the context of Omni-Directional Image (ODI) Super-Resolution (SR), the unique challenge arises from the non-uniform oversampling characteristics caused by EquiRectangular Projection (ERP). Considerable efforts in designing complex spherical convolutions or polyhedron reprojection offer significant performance improvements but at the expense of cumbersome processing procedures and slower inference speeds. Under these circumstances, this paper proposes a new ODI-SR model characterized by its capacity to perform Fast and Arbitrary-scale ODI-SR processes, denoted as FAOR. The key innovation lies in adapting the implicit image function from the planar image domain to the ERP image domain by incorporating spherical geometric priors at both the latent representation and image reconstruction stages, in a low-overhead manner. Specifically, at the latent representation stage, we adopt a pair of pixel-wise and semantic-wise sphere-to-planar distortion maps to perform affine transformations on the latent representation, thereby incorporating it with spherical properties. Moreover, during the image reconstruction stage, we introduce a geodesic-based resampling strategy, aligning the implicit image function with spherical geometrics without introducing additional parameters. As a result, the proposed FAOR outperforms the state-of-the-art ODI-SR models with a much faster inference speed. Extensive experimental results and ablation studies have demonstrated the effectiveness of our design.

Code — https://github.com/GingaUL/FAOR

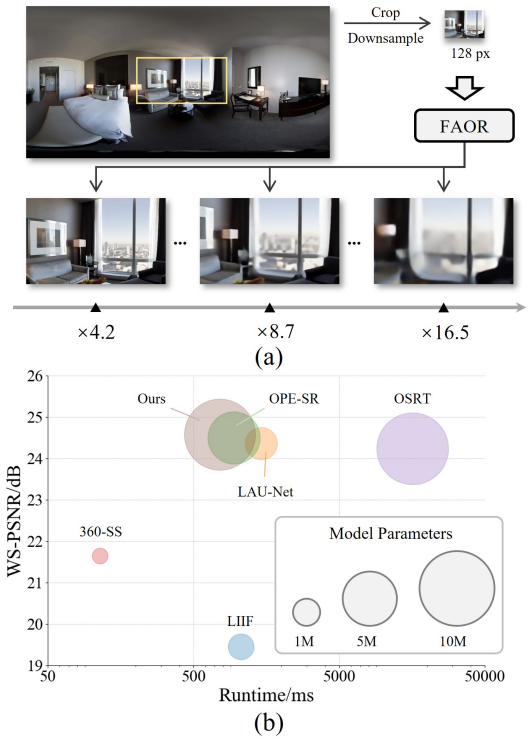

Figure 1: (a) The proposed FAOR provides a continuous ODI representation, enabling arbitrary-scale SR. (b) Comparisons with other state-of-the-art models regarding performance and inference speed on the $\times 8$ SR task

Introduction

Characterized by a wide $360° \times 180°$ field of view, Omni-Directional Images (ODIs) possess unique advantages in providing more comprehensive scene representations and immersive viewing experiences compared to conventional planar images. As a result, recent days have witnessed the blooming applications of ODIs in both machine and human vision-oriented domains, *e.g.*, augmented reality (AR), virtual reality (VR), autonomous driving (Yang et al. 2021; Xu et al. 2022), and robot navigation (Yang et al. 2020b; Kyoungkook and Sunghyun 2019). In practice, ODIs are usually stored and transmitted in Low Resolution (LR) to satisfy the requirements of real-time machine vision-oriented applications, and subsequently undergo Super-Resolution (SR) to achieve high perceptual quality for human vision.

As for ODI-SR, the main challenges arise from the distortion caused by sphere-to-planar projection. Specifically, raw ODIs are typically stored in EquiRectangular Projection (ERP) format, which straightforwardly maps longitude to equally spaced columns and latitude to equally spaced

*These authors contributed equally.
†Corresponding author.

rows. As a result, the pixels in high-latitude regions are oversampled, leading to significant content deformation in these areas. As such, directly adapting planar image SR models to ODI images is inappropriate, as they cannot comprehensively capture the scene's contextual information without considering ODI-specific characteristics. In addressing this challenge, numerous efforts have been devoted to various strategies, *e.g.,* developing specific convolution operators aligned with spherical characteristics (Lee et al. 2019), leveraging latitude-adaptive methodologies that allow different latitude regions to adopt distinct upscaling factors (Deng et al. 2021; Cai et al. 2024), or introducing new polyhedron-based reprojection methods to minimize sphere-to-planar distortion (Yoon et al. 2022). Although significant achievements have been made, these ODI-oriented modules involve cumbersome implementation procedures, lead to time-consuming inference, and present significant obstacles to practical implementation. This motivates us to explore a more efficient approach that incorporates spherical characteristics into the SR process, pursuing both performance and running speed.

To this end, this paper proposes a new ODI-SR method that adapts the implicit image function to the ERP domain by incorporating the sphere geometric priors to both the *latent representation* and *image reconstruction* stages. In general, the implicit image function aims at establishing map functions between the coordinates and corresponding latent representations to pixel values, favored by its capacity for continuous image representation and fast inference speed (Chen, Liu, and Wang 2021; Song et al. 2023). Herein, at the *latent representation* stage, we first explore leveraging affine transformation for sphere-to-planar distortion representation, enjoying the streamlined computation with standard image operators. Thus, we propose the Affine-Transformation-based Feature Modulating (ATFM) module and insert it into the feature encoder, within which the affine transformations are applied to the extracted ERP features, obtaining sphere-to-planar aware latent representations. Moreover, the ATFM module is guided by a set of external priors, including a pixel-wise stretching ratio map and an instance segmentation map, which are responsible for providing insights into the sphere-to-planar distortions from pixel-wise and semantic perspectives, respectively. During the *image reconstruction* stage, we introduce a spherical geodesic-based resampling function that leverages unbiased spherical locations for the latent representation resampling process. We highlight our main contributions as follows:

- To the best of our knowledge, we are the first to address ODI-SR by jointly considering performance and running speed. With architectures specially designed to capture spherical characteristics and ensure inference simplicity, the proposed FAOR is capable of outperforming existing methods while achieving significantly faster running speeds. as illustrated in Fig. 1.
- We explore the affine transformation for representing sphere-to-planar distortion, incorporating pixel-wise and semantic-wise spherical geometric priors into the ODI latent representation.
- A Spherical Geodesic-based Implicit Image Function (SGIF) is proposed to provide a continuous and spherically unbiased ODI representation, facilitating efficient and arbitrary-scale ODI-SR processes.

Related Works

Omni-Directional Image Machine Vision

ODI-oriented machine vision research emphasizes understanding the object deformation caused by non-uniform sampling due to sphere-to-planar projection. Preliminary works focused on modifying and adapting regular convolution kernels to produce deformation-aware features. In special, Su *et al.* (Su and Grauman 2017) made the first attempt that leverages the knowledge distillation to obtain adaptive kernel sizes for 2D convolution filters. The following studies have directed their attention toward the adaptation of sampling grid positions of convolution filters (Tateno, Navab, and Tombari 2018). SphereNet (Benjamin, Paul, and Andreas 2018), for instance, is a notable work that has demonstrated the effectiveness of this methodology in tasks such as classification and object detection. Zhao *et al.* (Zhao et al. 2018) further leveraged a non-regular grid for each pixel based on its distortion level and convolved the sampled grid using square kernels shared by all pixels, facilitating end-to-end training. Apart from adapting 2D CNN kernels, another approach involves establishing spherical convolution kernels to achieve rotational equivariance and invariance. For instance, Esteves *et al.* (Carlos et al. 2018) made the first attempt to implement spherical harmonic domain CNN filters via group convolutions. Meanwhile, Yang *et al.* (Yang et al. 2020a) and Perraudin *et al.* (Nathanaël et al. 2019) took a different approach by leveraging a graph to represent the spherical image and achieving resilient isometric equivariance transformation through via hierarchical equal area isolatitude pixelization.

Omni-Directional Image Super-Resolution

Preliminary ODR-SR models adopted a naive strategy that straightforwardly adopted the checkpoints of planar image SR models and fine-tuned them on the ODIs, resulting in limited performance (Ozcinar, Rana, and Smolic 2019). LAU-Net (Deng et al. 2021) proposed the first ODI-SR approach that considered the non-uniform sampling characteristics of ODR images. Specifically, a tile-based strategy was introduced, where the ERP images were first split into several regions according to latitude, and the upsampling factors were adaptively assigned. Yoon *et al.* (Yoon et al. 2022) explored the usage of icosahedron projection to minimize sphere-to-planar distortion and proposed a new kernel weight-sharing scheme aligned with the icosahedron projection. Although these methods successfully enhance the performance of ODI-SR by incorporating the spherical characteristics of ODI, the specially designed models involve tedious operations, posing significant obstacles to both running speed and practical implementation and hindering further research of ODI-SR. Moreover, other works have tended to incorporate sphere-to-planar knowledge as side information into the feature extraction process, such as

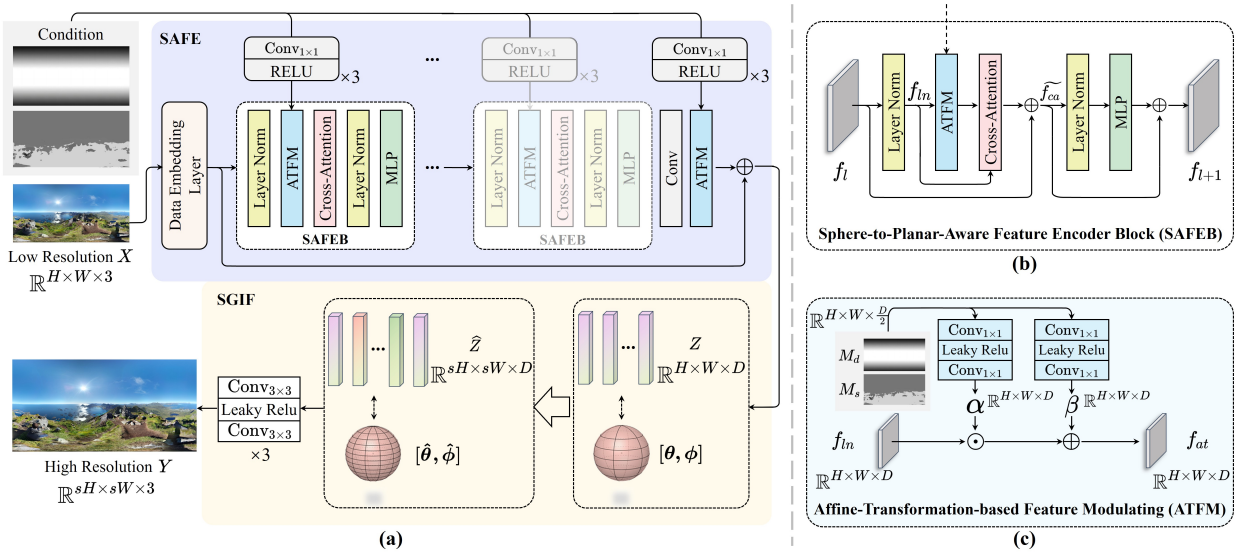

Figure 2: Overall architectures of the proposed FAOR and detailed design of key modules.

the stretching ratio or distortion map (Yu et al. 2023). However, the naive incorporation method and the sole focus on the feature extraction process prevent these approaches from achieving state-of-the-art performance.

Implicit Image/Neural Representation

The implicit neural representation (INR) originates from the modeling of 3D object-shape surfaces (Atzmon and Lipman 2020; Chen and Zhang 2019; Michalkiewicz et al. 2019) and involves leveraging a multi-layer perceptron function to map coordinates to the signal. Inspired by INR's favorable capacity for recovering fine details of shapes, it has started to be employed in representing planar images, such as in image restoration (Anokhin et al. 2021; Dupont, Teh, and Doucet 2022; Skorokhodov, Ignatyev, and Elhoseiny 2021) and image compression (Strümpler et al. 2022). Preliminary INR involves learning a specific mapping function for a single image, raising concerns about inference time in practical implementation scenarios. As such, recent works tend to explore an implicit representation function space shared by different images (Chen, Liu, and Wang 2021; Song et al. 2023; Nguyen and Beksi 2023), where the extracted deep features are also regarded as inputs alongside the coordinates for mapping to the signals, denoted as implicit image function. In the context of image SR, the implicit image function has obtained significant progress, especially in arbitrary-scale image SR, such as SIREN (Sitzmann et al. 2020) and LIIF (Chen, Liu, and Wang 2021), since the coordinate-to-signal function is well-aligned with continuous image representation. Furthermore, in the ODR-SR domain, SphereSR (Yoon et al. 2022) made the first effort to incorporate INR into spherical coordinates to achieve arbitrary-scale SR processing. However, the spherical implicit image function is performed on a reprojected icosahedron surface and involves a complex convolutional kernel modification process to align with the icosahedron ODI pixel representation. This inevitably limits the performance and inference speed.

Method

Motivation and Overview

As described in the background information, existing studies on ODI-SR are still insufficient. Shortages of exiting ODI-SR routes are summarized as follows, 1) Tile-based methods disrupt the spatial continuity of the input image, failing to comprehensively understand the scene context and provide satisfactory performance. 2) Reprojection-based methods involve specially designed convolution kernels to align with the corresponding pixel representation approach, resulting in a cumbersome and time-consuming inference process. 3) Sphere-to-planar prior-guided methods successfully address the above issues. However, the naive incorporation of prior information and the lack of consideration for the image reconstruction process prevent these methods from achieving state-of-the-art performance.

Motivated by this, this paper proposes a novel Fast and Arbitrary-Scale ODI Super-Resolution (FAOR) method that adapts the implicit image function to the ERP domain, incorporating spherical geometric priors into both the latent representation and image reconstruction processes. Bearing both fast inference speed and high SR performance in mind, the spherical geometric priors are integrated via specially designed lightweight modules. Moreover, although numerous projection methods exist for ODI representation, such as cube map, fisheye, and polyhedron, it has to be mentioned that almost all raw ODIs are recorded and stored in ERP format. With other projection types derived from ERP, patterns across them are reusable when distortions are correctly rectified. The architecture of the proposed FAOR framework is illustrated in Fig. 2, with its main modules highlighted as follows:

- **Sphere-to-planar-Aware Feature Encoder (SAFE).** The SAFE is responsible for extracting the latent representation $Z \in \mathbb{R}^{H\times W\times D}$ of the input LR ODI $X \in \mathbb{R}^{H\times W\times 3}$, where D denotes the feature map channels. In general, the SAFE is characterized by an Affine-Transformation-based Feature Modulating (ATFM) module and a Cross-Attention (CA)-based feature enhancement module. These are specially designed to incorporate spherical geometric priors and further enhance sphere-to-planar awareness, respectively.
- **Affine-Transformation-based Feature Modulating (ATFM).** The ATFM module jointly leverages a pixel-wise stretching ratio map M_d and an instance segmentation map M_s to generate a set of affine parameters, thereby obtaining sphere-to-planar aware latent representations.
- **Spherical Geodesic-based Implicit Neural Function (SGIF).** To align the image representation process with the spherical characteristics, the SGIF leverages a spherical geodesic-based resampling method, obtaining the latent representation $\hat{Z} \in \mathbb{R}^{sH\times sW\times D}$ of the HR-ODI, where s is the arbitrary upsampling scale. The $\hat{Z}$ and the corresponding spherical coordinates $[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}]$ are subsequently fed to the continuous implicit image function, obtaining the HR-ODI $Y \in \mathbb{R}^{sH\times sW\times 3}$.

Details of the modules will be presented in the following subsections.

Sphere-to-Planar-Aware Feature Encoder

The proposed SAFE is characterized by L duplicated feature encoder blocks, with details illustrated in Fig. 2. To gain intuition into the encoding process, denote the input feature of the $l-th$ encoder block as f_l, it would be first fed to a Layer Norm (LN) layer and subsequently to the ATFM module to generate the sphere-to-planar aware features f_{at},

$$f_{ln} = \text{LN}(f_l),\ \ f_{at} = \text{ATFM}(f_{ln}), \tag{1}$$

where the ATFM applies the affine transforms on the input features according to the spherical geometric and semantic priors, corresponding details will be presented in the following subsection. To further enhance awareness of the sphere-to-planar distortion, we incorporated a Cross-Attention (CA) module that conducts channel-wise attention between features with and without the affine transform. In special, consider the q_{ln} as the query extracted from f_{ln}, k_{at} and v_{at} are the key and value associated with f_{at}, the CA process can be formulates as,

$$f_{ca} = \text{MLP}(\frac{q_{ln}\times k_{at}}{\sqrt{d_k}} v_{at}), \tag{2}$$

where d_k is a scaling factor, MLP denotes the multi-layer perceptron layer. Moreover, a skip connection bypasses the ATFM and CA modules, and an additional residual block containing an LN layer and an MLP layer is incorporated to further boost training performance.

$$\begin{aligned} \widetilde{f_{ca}} &= f_l \oplus f_{ca}, \\ f_{l+1} &= \widetilde{f_{ca}} \oplus \text{MLP}(\text{LN}(\widetilde{f_{ca}})), \end{aligned} \tag{3}$$

where f_{l+1} denotes the output of the $l-th$ encoder block.

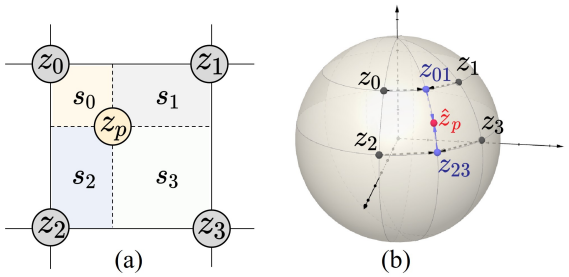

Figure 3: (a) The normalized interpolation method in LIIF, leveraging the areas of s_0, s_1, s_2, and s_3 to obtain normalized weights for z_p interpolation. (b) The proposed spherical geodesic-based interpolation.

Affine-Transformation-based Feature Modulating

The distortion pattern associated with sphere-to-planar projections in ERP is characterized by stronger stretching ratios at higher latitudes, whereas the stretching ratio along the longitude remains constant, and can be roughly formulated as,

$$D_{ERP}(h, w) = D_{ERP}(\theta, \phi) = \cos(\phi); \tag{4}$$

where (h, w) and (θ, ϕ) are equivalent coordinates in planar and spherical coordinate systems, respectively. To represent the nonuniform stretching artifact, this paper explores the use of affine transformations, which are capable of performing complex deformations, such as rotation and zooming, with streamlined computations.

In general, the ATFM module takes additional offline-generated priors as input to generate a set of element-wise affine parameters (α, β), performing the affine transformations on the input features. Details of the proposed module are illustrated in Fig. 2 (c), where f_{ln} and f_{at} denote the input and output features of the module, respectively.

As for the incorporated priors, a straightforward stretching ratio map M_d that reflects the pixel-wise sphere-to-planar distortion is generated by,

$$M_d(h, w) = 255 \times \cos(\frac{h + 0.5 - H/2}{H}\pi). \tag{5}$$

Moreover, we adopt an instance segmentation map M_s from (Zhang et al. 2022) to enhance awareness of spherical characteristics from the semantic deformation perspective. By constraining the perception perspective to the instance level, M_s is expected to provide a more thorough and accurate understanding of the sphere-to-planar distortion for the ATFM model. Thus, the entire ATFM process can be formulated as,

$$f_{at} = \text{ATFM}(f_{ln}, M_s, M_d) = \alpha \odot f_{ln} + \beta, \tag{6}$$

where f_{ln} and f_{at} denote the input and output features of the ATFM module, respectively.

Spherical Geodesic-based Implicit Image Function

In general, the implicit image function involves establishing a mapping function from latent vectors and corresponding coordinates to pixel values. With well-established mapping

Figure 4: Visual comparisons of $\times 8$ SR results on ODI-SR and SUN360 testing sets.

functions, the latent vectors closest to the targeted HR coordinate are fed into the mapping function and subsequently passed through a normalized interpolation, obtaining the targeted HR pixel values.

To align the image reconstruction process with spherical characteristics, the SGIF specifically incorporates a geodesic-based spherical resampling function. An intuitive comparison between the existing planar resampling function and the proposed spherical-oriented approach is illustrated in Fig. 3. Moreover, we propose a *resampling-then-representation* strategy that directly performs geodesic-based spherical resampling on the latent rather than on the pixel representation. This approach enables the implicit image function to be executed only once per HR pixel prediction, for the sake of fast inference speed.

In particular, with the obtained latent representation of the LR-ODI Z, we first obtain $H \times W$ latent vectors $[z_{11}, z_{12}, ..., z_{HW}]$, where $z_{hw} \in \mathbb{R}^{1\times1\times D}$ corresponding to the spherical coordinate $[\theta_h, \phi_w]$. Subsequently, the HR coordinates $\widehat{\boldsymbol{\theta}} = [\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_{sH}]$, and $\widehat{\boldsymbol{\phi}} = [\widehat{\phi}_1, \widehat{\phi}_2, ..., \widehat{\phi}_{sW}]$. are calculated and contributed to obtain the corresponding latent vectors $\widehat{Z} = [\widehat{z}_{11}, \widehat{z}_{12}, ..., \widehat{z}_{sHsW}]$ via spherical linear interpolation of Z. For simplicity, denote the target spherical coordinate as $[\widehat{\theta}_p, \widehat{\phi}_p]$, the four nearby reference latent vectors as z_0, z_1, z_2, z_3, as illustrated in Fig. 3(b). The spherical linear interpolation is conducted to obtain the HR latent vector $\widehat{z}_p$ corresponding to the target coordinate.

Specifically, we introduce a geodesic-based spherical linear interpolation since the ERP pixels within each row (colum) share the same latitude (longitude) (Fatelo and Martins-Ferreira 2022). We resample z_0 and z_1 to z_{01}, z_2 and z_3 to z_{23},

$$\begin{aligned} z_{01} &= \frac{\sin(1-t_{01})\delta_{01}}{\sin\delta_{01}} z_0 + \frac{\sin t_{01}\delta_{01}}{\sin\delta_{01}} z_1, \\ z_{23} &= \frac{\sin(1-t_{23})\delta_{23}}{\sin\delta_{23}} z_2 + \frac{\sin t_{23}\delta_{23}}{\sin\delta_{23}} z_3, \end{aligned} \tag{7}$$

where δ_{01} denote the longitude angle subtended by z_0 and z_1, as well as δ_{23}. Since the z_{01} and z_{23} are with the same longitude, the latent vector at the targeted coordinate can be obtained by,

$$\widehat{z}_p = \frac{\sin(1-t_{02})\delta_{02}}{\sin\delta_{02}} z_{01} + \frac{\sin t_{02}\delta_{02}}{\sin\delta_{02}} z_{23}, \tag{8}$$

where the δ_{02} and t_{02} are calculated based on the latitude of z_0 and z_2.

As illustrated in Fig. 2, the obtained latent vector $\widehat{z}_p$ and its spherical coordinate would be fed to the SGIF, predicting the corresponding pixel value y_p,

$$y_p = \text{SGIF}(\widehat{z}_p, [\widehat{\theta}_p, \widehat{\phi}_p]). \tag{9}$$

Along this vein, the predicted HR ODI image $Y = [y_{11}, y_{12}, ..., y_{sHsW}]$ can be obtained in a pixel-by-pixel manner.

Dataset	ODI-SR							
Method	$\times 2$		$\times 4$		$\times 8$		$\times 16$	
	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM
Cubic	27.61	0.8156	24.95	0.6923	19.64	0.5908	17.12	0.4332
LIIF	27.34	0.8214	22.29	0.6626	19.45	0.5692	17.59	0.5231
OPE-SR	29.20	0.8522	<u>26.48</u>	0.7435	<u>24.50</u>	0.6543	**22.82**	0.5992
360-SS	27.14	0.8095	23.20	0.6613	21.65	0.6417	19.65	0.5431
LAU-Net	29.33	0.8633	26.34	0.7352	24.36	**0.6602**	22.07	0.5901
OSRT	<u>29.61</u>	<u>0.8700</u>	26.26	<u>0.7443</u>	24.24	0.6532	22.49	**0.6030**
Ours	**30.12**	**0.8796**	**26.69**	**0.7560**	**24.58**	<u>0.6581</u>	<u>22.78</u>	<u>0.6008</u>
Dataset	SUN360 Panorama							
Cubic	28.01	0.8321	24.90	0.7083	19.72	0.5403	17.56	0.4638
LIIF	28.20	0.8424	22.17	0.6708	19.07	0.5706	16.59	0.5258
OPE-SR	30.67	0.8782	<u>27.42</u>	<u>0.7809</u>	<u>24.38</u>	<u>0.6848</u>	<u>22.50</u>	<u>0.6161</u>
360-SS	27.97	0.8294	23.19	0.6725	21.48	0.6352	19.62	0.5308
LAU-Net	29.11	0.8555	26.65	0.7479	24.02	0.6708	21.82	0.5824
OSRT	<u>31.20</u>	<u>0.8958</u>	26.93	0.7780	24.24	0.6533	22.31	**0.6211**
Ours	**32.06**	**0.9096**	**27.66**	**0.7950**	**24.77**	**0.6914**	**22.53**	**0.6211**

Table 1: Quantitative comparison with state-of-the-art SR methods on ODI-SR and SUN360 testing sets. The bold and underlined font indicate the best and second-best results, respectively.

Training and Supervision

A specifically designed self-supervised training strategy is employed, allowing the proposed SR model to be trained just once to gain the capacity for arbitrary-scale ODI-SR. In detail, the pristine HR ERP images are first cropped into a series of patches with resolution $128r \times 128r$ as ground truths, where r denotes random floating-point numbers. Subsequently, these ground-truth patches are downsampled to a uniform size of 128×128 using cubic interpolation, creating LR-HR pairs with randomized scales. Moreover, to achieve arbitrary-scale SR capacity within a single training approach, a fixed number of 128×128 pixels are randomly selected from different resolution ground-truth patches and contributed to the SGIF optimization. By constraining the selected pixel coordinates and r to uniform distributions, the SGIF is supplied with sufficient latent vector-pixel pairs at arbitrary coordinates. Moreover, the entire model optimization process is supervised by pixel-wise $\mathcal{L}_1$ loss.

Experiment

In this section, extensive experiments are conducted to demonstrate the superiority of the proposed method against the state-of-the-art regarding inference speed and perceptual quality. Furthermore, comprehensive ablation studies have demonstrated the effectiveness of our design.

Experimental Setting

Dataset: In the experimental stage, the ODI-SR (Deng et al. 2021) and SUN360 (Xiao et al. 2012) datasets are employed. In particular, the training sets of ODI-SR and SUN360 are jointly leveraged for training, comprising a total of 1151 ODIs with a resolution of 2048×1024. The performance on their testing sets, each containing 100 images, is separately reported. During the training process, corresponding LR versions are down-sampled via Cubic.

Benchmark: As for the benchmark, six SR models with various architectures and mechanisms are employed to provide a comprehensive comparison. This includes three state-of-the-art ODI-SR models: 360-SS (Ozcinar, Rana, and Smolic 2019), LAU-Net (Deng et al. 2021), and OSRT (Yu et al. 2023); two cutting-edge 2D SR models, LIIF (Chen, Liu, and Wang 2021) and OPE-SR (Song et al. 2023); and the conventional cubic interpolation method. Specifically, the 360-SS, LAU-Net, and OSRT models perform fixed-scale SR and are retrained for specific scales. Meanwhile, the OPE-SR and LIIF models achieve arbitrary-scale SR. All the learning-based models are retrained with the same dataset as ours to ensure fair comparisons. Regarding the evaluation criteria, WS-PSNR (Sun, Lu, and Yu 2017) and WS-SSIM (Zhou et al. 2018) have been adopted.

Implementation Details: We employ the Adam optimizer (Kingma and Ba 2014) with an initial learning rate of 10^{-4}, which will be halved at the 30k, 50k, 100k, and 400k-th iterations. The L indicating the number of SAFEB is set to 36. The networks are implemented in PyTorch 1.10.2 and Python 3.8.16. Training is performed on a machine equipped with 8 NVIDIA RTX 3090 GPUs.

Experimental Results

Quantitative Results: Comparisons between the proposed FAOR and the employed anchors are made at $\times 2$, $\times 4$, $\times 8$, and $\times 16$ scale SR tasks, corresponding results regarding reconstruction performances on testing set of ODI-SR and SUN360 are provided in Table 1. In particular, ODI-oriented SR methods have demonstrated overwhelming advantages over 2D SR models, underscoring the necessity of incorpo-

rating ODI characteristics into SR models. Moreover, compared to the state-of-the-art ODI-SR method, OSRT, our proposed method achieves an average gain of 0.39 dB in WS-PSNR and 0.006 in WS-SSIM on the ODI-SR dataset, and a 0.59 dB gain in WS-PSNR and 0.017 in WS-SSIM on the SUN 360 dataset.

Qualitative Comparison: A set of comparisons on $\times 8$ scale SR is provided in Fig. 4. As shown, the overwhelming advantages of the proposed FAOR over LIIF, 360-SS, and LAU-Net can be easily observed, as FAOR is capable of providing more natural object structuring and greatly improving the recovery of fine textures, *e.g.,* the house and tree regions in the above figure. Meanwhile, LAU-Net and 360-SS failed to recover detailed information and suffered from significant erasing artifacts. Moreover, in complex scenarios, the limitations of cutting-edge methods such as OPE-SR and OSRT become evident, with failures to reconstruct the texture details in the window regions, inevitably resulting in blurring artifacts. Meanwhile, the proposed FAOR continues to recover fine texture details effectively, owing to the incorporation of semantic and pixel-wise sphere-to-planar-aware priors.

Inference Speed Comparison: A comprehensive comparison regarding inference running speed, model parameters, and reconstruction performance on $\times 8$ scales is provided in Fig. 1 (b). Specifically, 360-SS achieves a faster running speed as it leverages a naive adaptation method that directly adapts 2D image SR models to the ODI domain using GAN-based supervision. However, the lack of consideration for spherical characteristics also results in poor visual representation. Compared to LIIF, our method achieves a faster running speed despite having more parameters, primarily due to the incorporation of the *resampling-then-representation* strategy. Unlike the vanilla local implicit image representation in LIIF, which requires executing the implicit function four times to predict one pixel, the *resampling-then-representation* strategy further boosts the running speed while maintaining the reconstruction performance. Moreover, compared to the cutting-edge OSRT, the proposed FAOR has a similar model size but operates much faster. The underlying reason lies in this paper's exploration of a new path for spherical characteristics incorporation, utilizing the streamlined affine transformation for sphere-to-planar awareness instead of complex spherical operators and reprojection processing, thereby striking a good balance between inference speed and SR performance.

Ablation Study

Extensive ablation studies are conducted to demonstrate the effectiveness of the incorporated spherical geometric priors, in both the SAFE and SGIF. The corresponding results and analysis are presented in this subsection.

Ablation of Incorporated Priors in SAFE In the latent representation stage, the pixel-wise stretching ratio map M_d and the instance segmentation map M_s are adopted to provide a comprehensive understanding of the sphere-to-planar distortion. Herein, we examine their contributions to SR performance on the ODI-SR dataset by gradually ablating them, while keeping all other implementation details strictly the same as in our full version. The corresponding results are listed in Table 2. The effectiveness of the M_d and M_s is evident, as they both contribute to an overall performance gain across multiple SR scales.

Method	Metrics	$\times 2$	$\times 4$	$\times 8$	$\times 16$
Ours	WS-PSNR	**30.12**	**26.69**	**24.58**	**22.78**
	WS-SSIM	0.8796	**0.7560**	**0.6581**	**0.6008**
$w/o\ M_d$	WS-PSNR	30.11	26.67	24.47	22.69
	WS-SSIM	**0.8850**	0.7558	0.6575	0.6004
$w/o\ M_d + M_s$	WS-PSNR	29.79	26.45	24.35	22.64
	WS-SSIM	0.8741	0.7482	0.6502	0.5949

Table 2: Ablation study of incorporated priors in SAFE. All models are tested on ODI-SR testing set.

Method	Metrics	$\times 2$	$\times 4$	$\times 8$	$\times 16$
Ours	WS-PSNR	**30.12**	**26.69**	**24.58**	**22.78**
	WS-SSIM	**0.8796**	**0.7560**	**0.6581**	**0.6008**
$w/o\ sphere$	WS-PSNR	29.95	26.56	24.42	22.78
	WS-SSIM	0.8761	0.7510	0.6533	0.5974

Table 3: Ablation study of spherical geodesic-based resampling function.

Ablation of Spherical Geodesic-based Resampling Function To demonstrate the effectiveness of the incorporated geodesic-based resampling function, we replace it with the normalized interpolation method in LIIF (denoted as '*w/o sphere*'), corresponding comparison results on the ODI-SR dataset are provided in Table 3. The benefit of the incorporated spherical geometric prior is easily observed, resulting in an improvement of 0.12 dB in WS-PSNR and 0.004 in WS-SSIM, respectively.

Conclusion

Aiming to perform fast and arbitrary-scale omnidirectional image super-resolution, this paper adapts the implicit image function from the planar image domain to the ERP image domain by incorporating multiple spherical geometric priors into both the latent representation and image reconstruction processes. Specifically, to obtain sphere-to-planar aware latent representations, we leverage a pixel-wise stretching ratio map and an instance segmentation map to enhance sphere-to-planar awareness from both pixel-wise and semantic perspectives. Additionally, we introduce a spherical geodesic-based resampling function to align the implicit image function with spherical characteristics. All incorporated modules are specially designed with respect to both complexity and efficiency. As a result, the proposed method achieves superior performance compared to state-of-the-art methods, with significantly faster inference speeds.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62371310, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515011236 and 2024A1515010454, in part by the Basic and Frontier Research Project of PCL, in part by the Major Key Project of PCL, and in part by by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under Grant No. GML-KF-24-27.

References

Anokhin, I.; Demochkin, K.; Khakhulin, T.; Sterkin, G.; Lempitsky, V.; and Korzhenkov, D. 2021. Image generators with conditionally-independent pixel synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14278–14287.

Atzmon, M.; and Lipman, Y. 2020. Sal: Sign agnostic learning of shapes from raw data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2565–2574.

Benjamin, C.; Paul, C. A.; and Andreas, G. 2018. SphereNet: Learning Spherical Representations for Detection and Classification in Omnidirectional Images. In *Proceedings of the European Conference on Computer Vision*, 62–78.

Cai, Q.; Li, M.; Ren, D.; Lyu, J.; Zheng, H.; Dong, J.; and Yang, Y.-H. 2024. Spherical pseudo-cylindrical representation for omnidirectional image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 873–881.

Carlos, E.; Christine, A.; Ameesh, M.; and Kostas, D. 2018. Learning SO (3) equivariant representations with spherical CNNs. In *Proceedings of the European Conference on Computer Vision*, 52–68.

Chen, Y.; Liu, S.; and Wang, X. 2021. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on computer vision and Pattern recognition*, 8628–8638.

Chen, Z.; and Zhang, H. 2019. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5939–5948.

Deng, X.; Wang, H.; Xu, M.; Guo, Y.; Song, Y.; and Yang, L. 2021. Lau-net: Latitude adaptive upscaling network for omnidirectional image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9189–9198.

Dupont, E.; Teh, Y. W.; and Doucet, A. 2022. Generative Models as Distributions of Functions. In *International Conference on Artificial Intelligence and Statistics*, 2989–3015.

Fatelo, J. P.; and Martins-Ferreira, N. 2022. Mobi spaces and geodesics for the N-sphere. *Cahiers de Topologie et Geometrie Differentielle Categoriques*, 63(1).

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kyoungkook, K.; and Sunghyun, C. 2019. Interactive and automatic navigation for 360 video playback. *ACM Transactions on Graphics*, 38(4): 1–11.

Lee, Y.; Jeong, J.; Yun, J.; Cho, W.; and Yoon, K.-J. 2019. Spherephd: Applying CNNs on a spherical polyhedron representation of 360deg images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9181–9189.

Michalkiewicz, M.; Pontes, J. K.; Jack, D.; Baktashmotlagh, M.; and Eriksson, A. 2019. Implicit surface representations as layers in neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4743–4752.

Nathanaël, P.; Michaël, D.; Tomasz, K.; and Raphael, S. 2019. Deepsphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications. *Astronomy and Computing*, 27: 130–146.

Nguyen, Q. H.; and Beksi, W. J. 2023. Single image super-resolution via a dual interactive implicit neural network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 4936–4945.

Ozcinar, C.; Rana, A.; and Smolic, A. 2019. Super-resolution of omnidirectional images using adversarial learning. In *International Workshop on Multimedia Signal Processing*, 1–6.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33: 7462–7473.

Skorokhodov, I.; Ignatyev, S.; and Elhoseiny, M. 2021. Adversarial generation of continuous images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10753–10764.

Song, G.; Sun, Q.; Zhang, L.; Su, R.; Shi, J.; and He, Y. 2023. OPE-SR: Orthogonal position encoding for designing a parameter-free upsampling module in arbitrary-scale image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10009–10020.

Strümpler, Y.; Postels, J.; Yang, R.; Gool, L. V.; and Tombari, F. 2022. Implicit neural representations for image compression. In *Proceedings of the European Conference on Computer Vision*, 74–91.

Su, Y.; and Grauman, K. 2017. Learning spherical convolution for fast features from 360 imagery. *Advances in Neural Information Processing Systems*, 30.

Sun, Y.; Lu, A.; and Yu, L. 2017. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE Signal Processing Letters*, 24(9): 1408–1412.

Tateno, K.; Navab, N.; and Tombari, F. 2018. Distortion-aware convolutional filters for dense prediction in panoramic images. In *Proceedings of the European Conference on Computer Vision*, 707–722.

Xiao, J.; Ehinger, K. A.; Oliva, A.; and Torralba, A. 2012. Recognizing scene viewpoint using panoramic place representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2695–2702.

Xu, H.; Zhao, Q.; Ma, Y.; Li, X.; Yuan, P.; Feng, B.; Yan, C.; and Dai, F. 2022. PANDORA: A Panoramic Detection Dataset for Object with Orientation. In *Proceedings of the European Conference on Computer Vision*, 237–252.

Yang, K.; Zhang, J.; Reiss, S.; Hu, X.; Xinxin; and Stiefelhagen, R. 2021. Capturing Omni-Range Context for Omnidirectional Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1376–1386.

Yang, Q.; Li, C.; Dai, W.; Zou, J.; Qi, G.; and Xiong, H. 2020a. Rotation equivariant graph convolutional network for spherical image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4303–4312.

Yang, S.; Zhu, W.; Xu, H.; and Zhang, X. 2020b. Graph learning based head movement prediction for interactive 360 video streaming. *IEEE Transactions on Multimedia*, 22(9): 2316–2327.

Yoon, Y.; Chung, I.; Wang, L.; and Yoon, K.-J. 2022. 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*, 5677–5686.

Yu, F.; Wang, X.; Cao, M.; Li, G.; Shan, Y.; and Dong, C. 2023. Osrt: Omnidirectional image super-resolution with distortion-aware transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13283–13292.

Zhang, J.; Yang, K.; Ma, C.; Reiß, S.; Peng, K.; and Stiefelhagen, R. 2022. Bending reality: Distortion-aware transformers for adapting to panoramic semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 16917–16927.

Zhao, Q.; Zhu, C.; Dai, F.; Ma, Y.; Jin, G.; and Zhang, Y. 2018. Distortion-aware CNNs for Spherical Images. In *International Joint Conference on Artificial Intelligence*, 1198–1204.

Zhou, Y.; Yu, M.; Ma, H.; Shao, H.; and Jiang, G. 2018. Weighted-to-spherically-uniform SSIM objective quality evaluation for panoramic video. In *Proceedings of the IEEE International Conference on Signal Processing*, 54–57.