OmniSSR: Zero-shot Omnidirectional Image Super-Resolution using Stable Diffusion Model

Runyi Li†, Xuhan Sheng†, Weiqi Li, and Jian Zhang ✉

School of Electronic and Computer Engineering, Peking University, China
{lirunyi,shengxuhan,liweiqi}@stu.pku.edu.cn
{zhangjian.sz}@pku.edu.cn

Abstract. Omnidirectional images (ODIs) are commonly used in real-world visual tasks, and high-resolution ODIs help improve the performance of related visual tasks. Most existing super-resolution methods for ODIs use end-to-end learning strategies, resulting in inferior realness of generated images and a lack of effective out-of-domain generalization capabilities in training methods. Image generation methods represented by diffusion model provide strong priors for visual tasks and have been proven to be effectively applied to image restoration tasks. Leveraging the image priors of the **S**table Diffusion (SD) model, we achieve **omni**directional image **S**uper **R**esolution with both fidelity and realness, dubbed as **OmniSSR**. Firstly, we transform the equirectangular projection (ERP) images into tangent projection (TP) images, whose distribution approximates the planar image domain. Then, we use SD to iteratively sample initial high-resolution results. At each denoising iteration, we further correct and update the initial results using the proposed Octadecaplex Tangent Information Interaction (OTII) and Gradient Decomposition (GD) technique to ensure better consistency. Finally, the TP images are transformed back to obtain the final high-resolution results. Our method is zero-shot, requiring no training or fine-tuning. Experiments of our method on two benchmark datasets demonstrate the effectiveness of our proposed method.

Keywords: Omnidirectional Imaging · Super-Resolution · Latent Diffusion Model

1 Introduction

Omnidirectional images (ODIs) capture the entire scene in all directions, exceeding the narrow field of view (FOV) offered by planar images. Super-Resolution

† Equal contributor ✉ Corresponding author

This work was supported in part by National Science Foundation of China (No. 62372016) and Shenzhen Research Project Grant (No. JSGGZD20220822095800001).

(SR) techniques enhance the visual quality of ODIs by increasing their resolution, thereby revealing finer details and enabling more accurate scene analysis and interpretation. This becomes particularly crucial in applications like virtual reality and surveillance, where high-resolution ODIs are essential for precise perception and decision-making.

Current research in omnidirectional image super-resolution (ODISR) explores various methodologies to enhance the resolution of ODIs [14, 36]. SphereSR [59] addresses non-uniformity in different projections by learning upsampling processes and ensuring information consistency using LIIF [5]. OSRT [60] designs a distortion-aware Transformer to modulate equirectangular projection (ERP) distortions continuously and self-adaptively. Without a cumbersome process, OSRT outperforms previous methods remarkably. However, existing ODISR methods face the following challenges: (1) The majority are end-to-end models that can only produce a deterministic output, always better data fidelity but worse visual perception quality [17]. It's promising to develop a generation-based model, but requiring high data demands, yet high-resolution ODIs are high cost to collect [54, 55]. (2) Most methods directly perform SR on ERP format ODIs, while users usually watch ODIs in a narrow FOV using tangent projection (TP). So another promising direction is to use off-the-shelf planar models on TP images. Recent times have witnessed the introduction and widespread application of diffusion models [23, 43], especially Stable Diffusion (SD) [38], which have provided a robust backbone for visual tasks [21, 24, 56, 61], including SR [30, 40, 47, 51, 52, 62]. However, if TP images are trivially one-by-one processed using diffusion-based SR models, they will exhibit discrepancies in the overlapping region when reprojected onto the ERP image. As a result, the global continuity is compromised.

Leveraging the strong image prior provided by SD, we propose the first diffusion-based zero-shot method for ODISR, named OmniSSR. Specifically, we propose Octadecaplex Tangent Information Interaction (OTII). OTII entails iterative conversion of intermediate SR results between ERP and TP, bridging the domain gap between ODIs and planar images. Building upon OTII, we further employ an approximate analytical solution of gradient descent, namely as Gradient Decomposition, to guide high-fidelity, high-quality omnidirectional image SR. By capitalizing on SD's effective image prior, our approach strikes a balance between *fidelity* and *realness*, ensuring that the restored ODIs exhibit both fidelity to the input data and realistic visual details. This method shows potential for advancing the current state of ODISR, providing improved resolution and visual quality across various applications. Fig. 1 showcases results fully demonstrating the superiority and performance of our proposed methods.

Our main contributions are summarized as follows:

- We propose OmniSSR, the first zero-shot ODISR method, using an off-the-shelf diffusion-based model, requiring no training or fine-tuning, leveraging existing image generation model priors to solve ODISR task.
- To bridge the domain gap between ODIs and planar images, we introduce Octadecaplex Tangent Information Interaction by repeatedly transforming

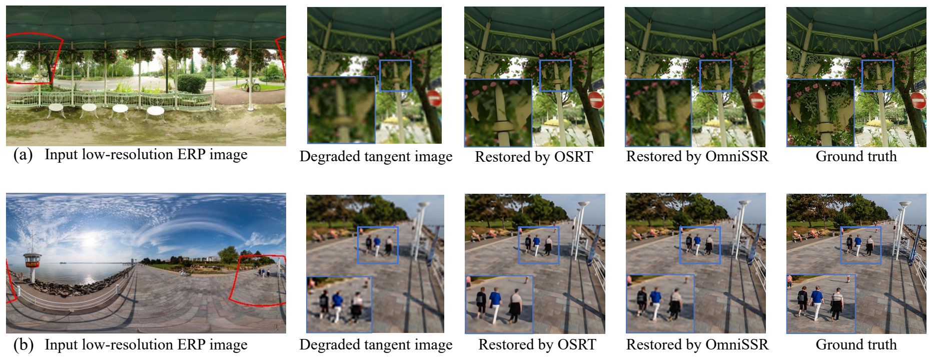

Fig. 1: We address omnidirectional image super-resolution in a *zero-shot* manner via OmniSSR. Presented above are select outcomes that sketch the essence of OmniSSR compared with current state-of-the-art approach OSRT [60]. Part (a) and (b) illustrate that OmniSSR upholds fidelity and visual realness at the same time, providing vivid and realistic details, while OSRT outputs over-smoothed and distorted results. Zoom in for more details.

ODIs between ERP format and TP format, enabling ODISR task with pre-trained diffusion models on planar images.

- By iteratively updating images using the developed Gradient Decomposition technique, we introduce consistency constraints into the sampling process of the latent diffusion model, ensuring a trade-off between fidelity and realness in the reconstructed results.
- Extensive experiments are conducted on the benchmark datasets, demonstrating the superior performance of our method over existing state-of-the-art approaches, which validate the effectiveness of OmniSSR.

2 Related Work

2.1 Single Image Super-Resolution (SISR)

Image super-resolution methods based on deep learning have undergone significant development over an extended period. Currently, they can be broadly classified into two categories of solutions. The first category involves end-to-end network training methods, which utilize image pairs consisting of low-resolution degraded images and high-resolution ground truth images for network training [6–8, 27, 31, 32, 63, 66]. The network architectures employed in this category include CNN [16], Transformers [46], etc. The second category employs image generation models as priors, such as GAN [20], diffusion models [23, 43], etc., where low-resolution images are used as conditions to generate high-resolution images. We will mainly introduce the methods using generative prior.

Single Image SR using GAN prior In SR works utilizing GAN priors [3, 12, 33, 37, 58], including real-world senarios [8, 49, 50, 65], pre-trained GAN networks are employed to transform image features into latent space, where the corresponding latent code for the high-resolution image is searched, ultimately yielding the reconstructed high-resolution result.

Single Image SR using diffusion prior The diffusion model provides a powerful image prior, and the diffusion sampling process can generate highly realistic images. This strong prior distribution can be applied to various image restoration tasks, including super-resolution [9, 10, 19, 40, 42, 51]. Image-domain diffusion models directly provide prior distributions of image-domain data. DDNM [51] based on the mathematical method of Range-Null space Decomposition, iteratively refines content on the zero space, combining image prior content in the value domain to achieve image restoration. DDRM [25] uses SVD decomposition to obtain restoration results, which is similar to DDNM. DPS [9] transforms the image super-resolution problem into an optimization problem with consistency constraints, using gradient descent algorithms to guide the generation of image-domain diffusion models. GDP [19] further uses such gradient to update the degradation operator to tackle blind image inverse problems. Other methods including MCG [10], DDS [9] and unified control of diffusion generation [19, 42] use same strategy for image restoration, especially image super-resolution.

2.2 Omnidirectional Image Super-Resolution

Omnidirectional image super-resolution (ODISR) aims to enhance the resolution of omnidirectional or 360-degree images, which are commonly captured by cameras with a wide field of view. This field has garnered increasing attention due to its applications in virtual reality, omnidirectional video streaming, and surveillance. Several approaches have been proposed to address the unique challenges of ODISR [1, 2, 35, 44]. For instance, Kämäräinen et al. [18] propose a deep learning-based approach for omnidirectional super-resolution, leveraging convolutional neural networks to effectively upscale low-resolution omnidirectional images while preserving spatial details. Similarly, Smolic et al. [36] introduce a novel omnidirectional super-resolution algorithm utilizing generative adversarial networks (GANs) to enhance the visual quality of omnidirectional images by hallucinating high-frequency details.

For evaluation purposes, researchers commonly utilize datasets such as the ODI-SR dataset from LAU-Net [13], and SUN 360 Panorama dataset [53]. These datasets enable the quantitative assessment of ODISR algorithms across various scenarios and facilitate fair comparisons between different methods.

3 Method

In this section, we first briefly introduce the preliminary background of our method (Sec. 3.1), and give an overall view of our proposed OmniSSR (Sec. 3.2). Then, we discuss the designs of Octadecaplex Tangent Information Interaction, which transform ODIs between ERP and TP formats with pre-upsampling strategy (Sec 3.3), and the Gradient Decomposition correction (Sec. 3.4).

3.1 Preliminaries

ERP↔TP Transformation The essence of projection transformations between ERP and TP lie in determining the positions of target image pixels within the source image and computing their corresponding pixel values using interpolation algorithms, as digital images are always stored discretely [28]. Therefore, the ERP→TP transformation involves locating the TP image pixels on the ERP imaging plane, and vice versa. Gnomonic projection [11] provides the correspondence between ERP image pixels and TP image pixels.

For a pixel $P_e(x_e, y_e)$ within the ERP image, we first find its corresponding pixel $P_s(\theta, \phi)$ on the unit sphere using Eq. 1:

$$\theta = 2\pi x_e/W,\ \phi = \pi y_e/H, \tag{1}$$

where H and W are the height and width of the ERP image. The Cartesian coordinates of the ERP image and the angular coordinates on the unit sphere exhibit a straightforward one-to-one linear relationship, suggesting a conceptual equivalence between these two projection formats.

Given the spherical coordinates of the tangent plane center (θ_c, ϕ_c), The transformation from $P_s(\theta, \phi)$ to $P_t(x_t, y_t)$, i.e. ERP→TP, is defined as:

$$\begin{aligned} x_t &= \big(\cos(\phi)\sin(\theta-\theta_c)\big)/\zeta, \\ y_t &= \big(\cos(\phi_c)\sin(\phi) - \sin(\phi_c)\cos(\phi)\cos(\theta-\theta_c)\big)/\zeta, \end{aligned} \tag{2}$$

where $\zeta = \sin(\phi_c)\sin(\phi) + \cos(\phi_c)\cos(\phi)\cos(\theta - \theta_c)$.

The corresponding inverse transformation, i.e. TP→ERP, is:

$$\begin{aligned} \theta &= \theta_c + \arctan\big((x_t \sin(c))/(\rho\cos(\phi_1)\cos(c) - y_t \sin(\phi_c)\sin(c))\big), \\ \phi &= \arcsin\big(\cos(c)\sin(\phi_c) + y_t \sin(c)\cos(\phi_c)/\rho\big), \end{aligned} \tag{3}$$

where $\rho = \sqrt{x_t^2 + y_t^2}$ and $c = \arctan(\rho)$.

With Eq. 2 and Eq. 3, we can build one-to-one forward and inverse mapping functions between pixels on the ERP image and pixels on the TP images. An illustration of the ERP→TP transformation is shown in Fig. 2(a).

Iterative Denoising for Super-Resolution Utilizing the rich image priors provided by SD, we can super-resolve planar images. During initialization, the images are passed through the encoder $\mathcal{E}$ of SD to obtain latent codes, which are then added to pure noise to generate initial noise $\mathbf{z}_T$. Following the approach proposed by StableSR [47], we pass the images through a time-aware adapter $\mathcal{T}$. This adapter network structure is similar to the down-sampling part in denoising UNet, taking the image and the time step t of diffusion sampling as inputs to obtain the latent code feature for step t. This feature, along with the latent code $\mathbf{z}_t$ for each step and the time step t, is then passed through denoising UNet to calculate the denoised result $\mathbf{z}_{0|t}$ and the latent code $\mathbf{z}_{t-1}$ for the next sampling step. By iterating this process T times, we can obtain the final super-resolution result via decoder $\mathcal{D}$ of SD, yielding high-resolution images.

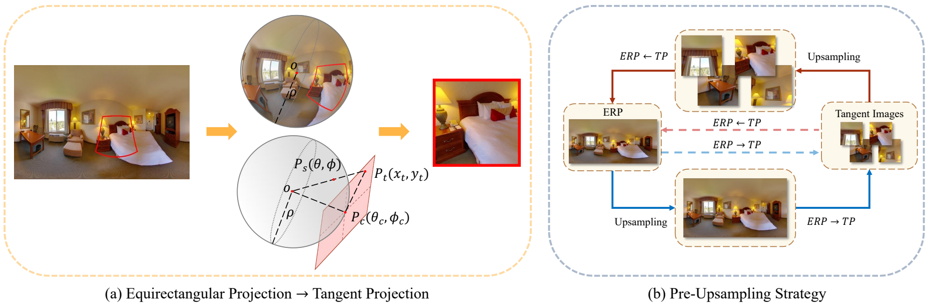

Fig. 2: Details about gnomonic transformations. (a) conversion from ERP to TP. (b) pre-upsampling proposed in Octadecaplex Tangent Information Interaction (Sec. 3.3) mitigating loss during transformation.

3.2 Overview

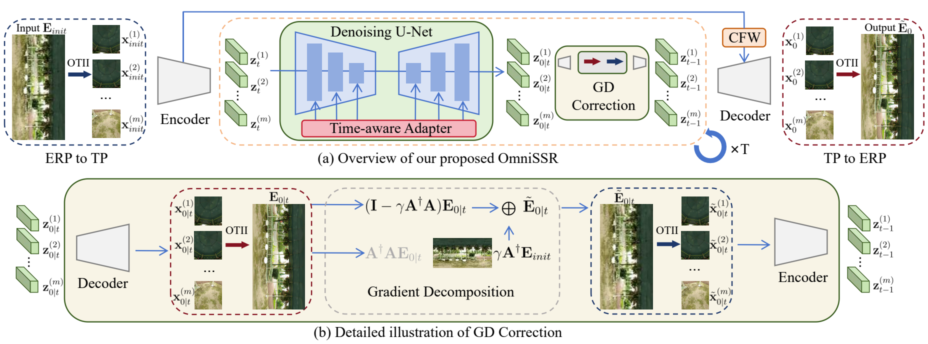

Fig. 3: Overview of our proposed OmniSSR. Input low-resolution omnidirectional image $\mathbf{E}_{init}$ in ERP format is first projected onto Tangent Projection (TP) images $\mathbf{x}_{init}^{(1)}, \mathbf{x}_{init}^{(2)}, ..., \mathbf{x}_{init}^{(m)}$, then iteratively refined via Stable Diffusion (SD) with a time-aware adapter and controllable feature wrapping (CFW) module. In each step of diffusion sampling, we adopt the Gradient Decomposition (GD) correction technique to introduce consistency constraints for the restored intermediate results. After T steps of sampling, we obtain the final result $\tilde{\mathbf{E}}_0$ with high resolution and better visual quality.

Our approach can be divided into three parts. The first part is pre-processing, where we initially up-sample the low-resolution ERP images $\mathbf{E}_{init}$ with bicubic interpolation to target high-resolution size, then project them onto tangent planes to obtain a series of TP images. These TP images are transformed to the latent space by the SD encoder, iteratively processed through denoising UNet and time-aware adapter network, and then decoded to obtain high-resolution TP images. During each denoising step, these TP images are transformed back via inverse transformation to ERP images, employing the Gradient Decomposition correction to ensure consistency constraints in diffusion sampling. After T

Algorithm 1: OmniSSR Pipeline

Input: $\mathbf{E}_{init}$, $\mathcal{F}$, $\mathcal{F}^{-1}$, $\mathbf{A}$, $\mathbf{A}^{\dagger}$, $\mathcal{E}$, $\mathcal{D}$, T
Output: SR result $\tilde{\mathbf{E}}_0$
$\{\mathbf{x}_{init}^{(1)}, \mathbf{x}_{init}^{(2)}, ..., \mathbf{x}_{init}^{(m)}\} = \mathcal{F}(\mathbf{E}_{init})$
for $i = 1$ **to** m **do**
  $\mathbf{z}_{init}^{(i)} = \mathcal{E}(\mathbf{x}_{init}^{(i)})$
  $\boldsymbol{\epsilon}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{z}_T^{(i)} = \sqrt{\overline{\alpha}_T}\mathbf{z}_{init}^{(i)} + \sqrt{1-\overline{\alpha}_T}\boldsymbol{\epsilon}^{(i)}$
end
Get$\{\mathbf{z}_0^{(1)}, \mathbf{z}_0^{(2)}, ..., \mathbf{z}_0^{(m)}\}$ from Algo. 2
for $i = 1$ **to** m **do**
  $\mathbf{x}_0^{(i)} = \mathcal{D}(\mathbf{z}_0^{(i)})$
end
$\mathbf{E}_0 = \mathcal{F}^{-1}(\{\mathbf{x}_0^{(1)}, \mathbf{x}_0^{(2)}, ..., \mathbf{x}_0^{(m)}\})$
$\tilde{\mathbf{E}}_0 = \mathbf{E}_0 + \gamma_p \mathbf{A}^{\dagger}(\mathbf{E}_{init} - \mathbf{A}\mathbf{E}_0)$
return $\tilde{\mathbf{E}}_0$

Algorithm 2: Iterative Denoising with GD Correction

Input: $\mathbf{E}_{init}$, $\mathcal{F}$, $\mathcal{F}^{-1}$, $\mathbf{A}$, $\mathbf{A}^{\dagger}$, $\mathcal{E}$, $\mathcal{D}$, $\mathcal{T}$, T
Output: Latent code $\{\mathbf{z}_0^{(1)}, \mathbf{z}_0^{(2)}, ..., \mathbf{z}_0^{(m)}\}$
for $t = T$ **to** 1 **do**
  for $i = 1$ **to** m **do**
    $\boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_\theta(\mathbf{z}_t^{(i)}, \mathcal{T}(\mathbf{z}_{init}^{(i)}, t), t)$
    $\mathbf{z}_{0|t}^{(i)} = \frac{1}{\sqrt{\overline{\alpha}_t}}(\mathbf{z}_t^{(i)} - \boldsymbol{\epsilon}_t\sqrt{1-\overline{\alpha}_t})$
    $\mathbf{x}_{0|t}^{(i)} = \mathcal{D}(\mathbf{z}_{0|t}^{(i)})$
  end
  $\mathbf{E}_{0|t} = \mathcal{F}^{-1}(\{\mathbf{x}_{0|t}^{(1)}, \mathbf{x}_{0|t}^{(2)}, ..., \mathbf{x}_{0|t}^{(m)}\})$
  $\tilde{\mathbf{E}}_{0|t} = \mathbf{E}_{0|t} + \gamma_e \mathbf{A}^{\dagger}(\mathbf{E}_{init} - \mathbf{A}\mathbf{E}_{0|t})$
  $\{\tilde{\mathbf{x}}_{0|t}^{(1)}, \tilde{\mathbf{x}}_{0|t}^{(2)}, ..., \tilde{\mathbf{x}}_{0|t}^{(m)}\} = \mathcal{F}(\tilde{\mathbf{E}}_{0|t})$
  for $i = 1$ **to** m **do**
    $\tilde{\mathbf{z}}_{0|t}^{(i)} = (1-\gamma_l)\mathbf{z}_{0|t}^{(i)} + \gamma_l \mathcal{E}(\tilde{\mathbf{x}}_{0|t}^{(i)})$
    $\mathbf{z}_{t-1}^{(i)} \sim p(\mathbf{z}_{t-1}^{(i)} | \mathbf{z}_t^{(i)}, \tilde{\mathbf{z}}_{0|t}^{(i)})$
  end
end
return $\{\mathbf{z}_0^{(1)}, \mathbf{z}_0^{(2)}, ..., \mathbf{z}_0^{(m)}\}$

iterations, the final super-resolution result is obtained. A formulaic description for OmniSSR is shown in Algo. 1. Fig. 3 shows the overview of our pipeline.

3.3 Octadecaplex Tangent Information Interaction (OTII)

Motivation To apply SD for ODISR, a straightforward way is to perform the ERP→TP transformation on the input ERP image. Then, each obtained TP image is fed into the SD-based model for SR. Finally, the TP→ERP transformation yields the ultimate super-resolved ERP image. OmniFusion [28] employs a similar approach for depth estimation. However, this simplistic strategy fractures the inherent global coherence of ODIs, leading to pixel-level discontinuities in the fused ERP images. Moreover, interpolation algorithms cause significant information loss in the original projection transformations, resulting in more blurred images. If applied multiple times, this exacerbates the information loss even further. To mitigate this, a trivial solution is to increase the pixel count (resolution) of the intermediate projection imaging plane. However, excessively high resolutions in TP images can introduce unnecessary computational overhead during the denoising stage and potentially compromise the denoising performance. (see Supplementary Materials for details)

Information Interaction and Pre-upsampling Based on the analysis of the **Motivation** in 3.3, we propose OTII by alternating the intermediate results between ERP and TP formats at each denoising step, where a single ERP image is represented by 18 TP images. From Sec. 3.1, we can achieve the ERP→TP transformation (denoted as $\mathcal{F}(\cdot)$) and the TP→ERP transformation (denoted as $\mathcal{F}^{-1}(\cdot)$). Through the ERP→TP transformation, we can convert distorted ERP images into TP images with content distributions that approximate those of planar images. This enables the use of the original SD super-resolution method for planar images. Conversely, through the TP→ERP transformation, we can fuse information between different TP images holistically, while providing ERP-format input for the subsequent GD Correction in Sec. 3.4. To handle information loss during projection transformation, we further propose to pre-upsample the source image before projection transformations, as shown in Fig. 2(b). Our experiments in Sec. 4.4 demonstrate that this pre-upsampling strategy can significantly mitigate the information loss caused by projection transformations.

3.4 Gradient Decomposition (GD) Correction for Fidelity

SD-based methods, as introduced in Sec. 3.1, can perform SR on sliced TP images. However, relying solely on the SR results from SD may lack consistency and fail to accurately preserve the original semantic information and details of the low-resolution image.[1] To enhance the consistency of the SR results from SD, we opt to use convex optimization methods to iteratively refine them. Modeling the SR task as an image inverse problem, the following equation is formulated:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{4}$$

where $\mathbf{x}$ represents the original image, $\mathbf{y}$ denotes the degraded result, $\mathbf{A}$ is the degradation operator (e.g., bicubic downsampling for super-resolution), and $\mathbf{n}$ is random noise. The objective we aim to solve can be expressed as the following convex optimization problem:

$$\underset{\mathbf{x}}{\operatorname{argmin}} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \lambda \mathcal{R}(\mathbf{x}), \tag{5}$$

where the first term is the data-fidelity term, ensuring the consistency of image reconstruction, and the second term is the regulation term, ensuring the sparsity of the reconstruction result, thus making the image more realistic. The regularization term can be the 1-norm, Total Variation, etc. The aforementioned convex optimization problem can be solved using a series of algorithms, such as gradient descent, ADMM [4], etc. Considering the trade-off between time and performance, we turn to find a solution based on gradient descent, and provide an approximate analytical solution composed of a *fidelity* term and a *realness* term, named "Gradient Decomposition (GD)":

$$\begin{aligned} \tilde{\mathbf{E}}_{0|t} &= \mathbf{E}_{0|t} + \alpha \nabla_{\mathbf{E}_{0|t}} ||\mathbf{E}_{init} - \mathbf{A}\mathbf{E}_{0|t}||_F = \mathbf{E}_{0|t} + \alpha \times 2(\mathbf{A}^\dagger \mathbf{E}_{init} - \mathbf{A}^\dagger \mathbf{A}\mathbf{E}_{0|t}) \\ &= \mathbf{E}_{0|t} + \gamma \mathbf{A}^\dagger (\mathbf{E}_{init} - \mathbf{A}\mathbf{E}_{0|t}) = \gamma \mathbf{A}^\dagger \mathbf{E}_{init} + (\mathbf{I} - \gamma \mathbf{A}^\dagger \mathbf{A})\mathbf{E}_{0|t} \end{aligned} \tag{6}$$

where $\mathbf{A}^\dagger$denotes pseudo-inverse of degradation operator $\mathbf{A}$, $\mathbf{E}_{init}$ denotes initial low-resolution ERP input, $\mathbf{E}_{0|t}$ denotes restored result by SD, $\tilde{\mathbf{E}}_{0|t}$ denotes corrected result by GD, α denotes the learning rate of gradient descent, and γ

[1] This claim will be further illustrated in subsequent experiments.

denotes the simplified hyper-parameter which is further tuned using grid search. The final setting of γ on different stages is shown in Sec. 4.1, and the ablation studies of parameter choice are in Sec. 4.4.

This technique could be seen as a step of gradient descent optimization, and the optimized result could be decomposed of (1) $\gamma\mathbf{A}^{\dagger}\mathbf{E}_{init}$, which ensures the consistency of the generated result, and (2) $(\mathbf{I}-\gamma\mathbf{A}^{\dagger}\mathbf{A})\mathbf{E}_{0|t}$, which serves as the iteratively updated generated result to improve its realness; γ is a hyper-parameter balancing the data fidelity and visual quality. For a better diversity and generality of the SR process, we expand this solution to latent space, and obtain the denoising result from both denoising UNet and corrected TP images (Algo. 2 line 11). A more detailed understanding of the iterative denoising process and application of GD correction could be referred to Algo. 2.

4 Experiments

4.1 Implementation Details

Datasets and Pretrained Models We choose the test set of ODI-SR dataset from LAU-Net [13] and SUN 360 Panorama dataset [53], comprising 97 and 100 omnidirectional images respectively, for experimental evaluation. The ground truth images are of size 1024×2048 pixels. In SR methods such as GDP [19] and PSLD [39] for planar images, we partitioned the images into several 256×256 patches and performed super-resolution on each patch individually.

For pre-trained models, we adopt from StableSR [47], which provided a SR network for planar images based on SD. This network architecture includes a time-aware adapter, a controllable feature wrapping (CFW) module, and the original SD structure from HuggingFace. All of them are kept untrained in our proposed OmniSSR.

Settings We set diffusion sampling steps to 200, which is the same as StableSR. The steps for other diffusion-based methods are set the same as their default settings (e.g. 1000 steps for PSLD). The degradation for low-resolution ERP images is bicubic down-sampling, and the implementation of its pseudo-inverse can be referred from code of DDRM [25][2]. For choices of hyper-parameter γ in GD correction, we set $\gamma_p = 1.0$, $\gamma_e = 1.0$, $\gamma_l = 0.5$. Our code is developed via PyTorch on NVIDIA 3090Ti GPU. [3]

4.2 Comparison of OmniSSR with diffusion-based methods

To evaluate the performance of proposed OmniSSR, we compare our method with recent state-of-the-art zero-shot methods for single image SR task: DPS [9], DDRM [25], GDP [19] which are based on the image-domain diffusion model,

[2] https://github.com/bahjat-kawar/ddrm
[3] Code is at https://github.com/LiRunyi2001/OmniSSR.

Table 1: SR results under bicubic downsampling on ODI-SR and SUN 360 Panorama datasets. For tasks not implemented in those papers, we mark N/A in the corresponding results. The best results are shown in **bold**, and the second best results are underlined.

Method	Scale	ODI-SR				SUN 360 Panorama			
		WS-PSNR↑	WS-SSIM↑	FID↓	LPIPS↓	WS-PSNR↑	WS-SSIM↑	FID↓	LPIPS↓
Bicubic		28.14	0.8343	24.00	0.2164	28.67	0.8537	29.25	0.1933
DDRM [25]		27.90	0.8317	**12.28**	0.1661	29.55	0.8670	13.10	**0.1426**
DPS [9]		20.99	0.6194	148.30	0.5249	21.44	0.6598	148.83	0.5175
GDP [19]	×2	27.89	0.8157	26.56	0.2724	28.60	0.8376	28.02	0.2445
PSLD [39]		N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
DiffIR [52]		23.77	0.6583	57.23	0.4687	23.54	0.6775	58.06	0.4658
StableSR [47]		22.70	0.6458	44.87	0.3039	23.30	0.6907	43.49	0.2858
OmniSSR		**28.57**	**0.8540**	13.01	**0.1575**	**29.69**	**0.8781**	**12.99**	0.1459
Bicubic		25.43	0.7059	50.84	0.3755	25.49	0.7229	55.99	0.3656
DDRM [25]		25.43	**0.7367**	32.69	0.3206	25.83	0.7443	**32.93**	0.3304
DPS [9]		24.75	0.6594	120.74	0.4911	21.09	0.6119	175.2143	0.5541
GDP [19]	×4	23.16	0.6692	77.43	0.4260	23.75	0.6569	90.23	0.4240
PSLD [39]		21.72	0.5498	107.99	0.5329	21.75	0.5828	141.49	0.5461
DiffIR [52]		24.01	0.6770	54.14	0.4367	23.90	0.7014	50.37	0.4235
StableSR [47]		23.33	0.6577	49.95	0.3135	23.99	0.6998	46.03	0.3023
OmniSSR		**25.77**	0.7279	**30.97**	**0.2977**	**26.01**	**0.7481**	34.58	**0.2963**

and PSLD [39], which is based on latent diffusion model. We also choose supervised diffusion-based super-resolution approaches including StableSR [47] and DiffIR [52] for comparison. We conduct experiments on ×2 and ×4 SR with ERP bicubic downsampling, on ODI-SR test set and SUN test set. We choose WS-PSNR [45], WS-SSIM [67], FID [22], and LPIPS [64] as the main metrics.

Quantitative results are presented in Tab. 1. With proposed OTII and GD correction, OmniSSR out-performs previous methods in terms of both *Fidelity* (from WS-PSNR and WS-SSIM) and *Realness* (from FID, LPIPS), which shows superior performance to existing diffusion-based methods for ODISR tasks on different scales.

Qualitative results are shown in Fig. 4 and Fig. 5, which illustrates the visualization of SR results on SUN test set and ODI-SR test set with ×2 and ×4 scales, by different methods.

The visual results indicate that our OmniSSR exhibits superior capability for detail recovery compared to other methods, particularly evident in textual elements (e.g., the text "flapping" in upper part of Fig. 4), complex objects (e.g., the black desk with a screen in lower part of Fig. 4, patterns above the white door in lower part of Fig. 5), and small-scale objects (e.g., the person and clock behind the desk in upper part of Fig. 5). OmniSSR demonstrates the ability to recover highly detailed and realistic visual effects from TP images.

4.3 Comparison with end-to-end supervised methods

The experiments of comparison in Sec. 4.2 are mainly focused on zero-shot image super-resolution methods, and supervised single image super-resolution methods, where the approaches are not trained or fine-tuned on omnidirectional images. In this part, we will compare OmniSSR to supervised end-to-end methods with end-to-end training on ODI datasets, including SwinIR and OSRT. Besides the

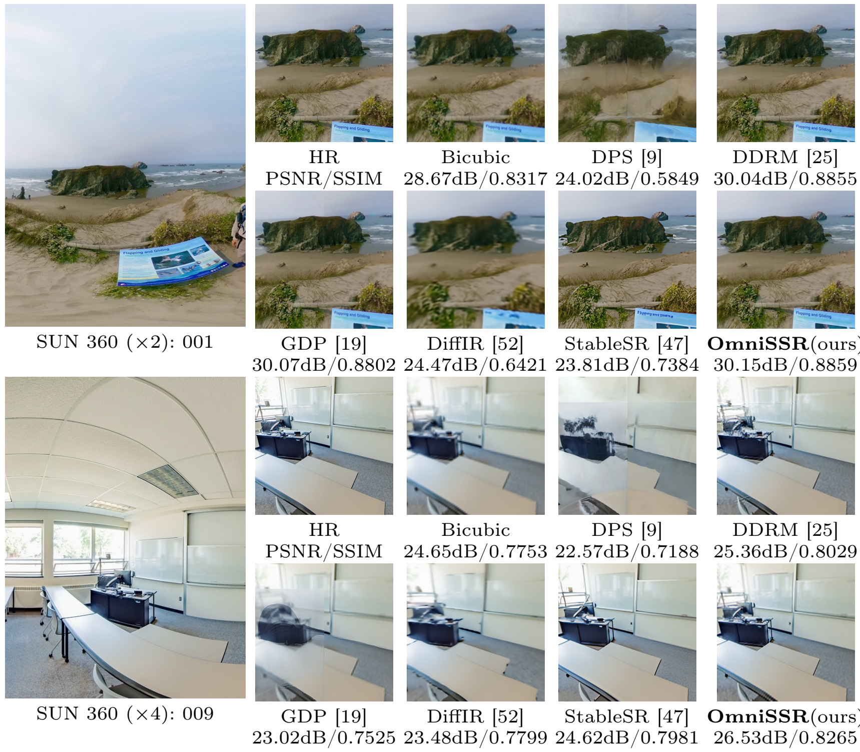

Fig. 4: Visualized comparison of ×2 and ×4 SR results on SUN 360 test set. 001 and 009 are the ID numbers in the test set filenames. We also calculate the PSNR and SSIM to HR ground truth of each SR result and downsampled image.

main metrics in Sec. 4.2, we also use NIQE [34] and DISTS [15] to evaluate the visual perception of SR outputs. Results are presented in Tab. 2, which shows that although our OmniSSR exhibits inferior fidelity metrics compared to end-to-end supervised methods trained directly on ODI datasets, it demonstrates notable improvements in the visual quality and authenticity of super-resolved images. Notably, end-to-end methods often produce smoothed reconstructions with distortions, whereas our approach preserves finer details and adheres more closely to the realistic distribution. Considering that our method has never been trained or tuned on ODI datasets, nor having omnidirectional images prior, this result is acceptable.

4.4 Ablation Studies

We first sequentially validate the performance improvement of the proposed strategy in OmniSSR including input image type, OTII and GD correction, on the ODI-SR test set with ×2 SR task, thereby demonstrating the significance of these strategies. The details are demonstrated as follows:

Fig. 5: Visualized comparison of ×2 and ×4 SR results on ODI-SR test set. 067 and 049 are the ID numbers in test set filenames. We also calculate the PSNR and SSIM between ground truth and each SR result as well as the downsampled image.

1) we do not use any proposed strategy in the SR task, which is equivalent to the vanilla StableSR baseline;

2) we transform the degraded ERP image to TP images and feed them separately into StableSR pipeline, instead of directly inputting ERP images;

3) based on 2), we add OTII strategy during the denoising process of SD (Algo. 2 line 7);

4) based on 2), we add GD correction at the *post-processing* stage (Algo. 1 line 12) of the overall pipeline;

5) based on 3) and 4), we add GD correction at *every step* and *post-processing* stage of sampling, to improve the consistency of the restored result.

Note that the execution of GD correction requires the execution of OTII in the denoising process simultaneously, there is no scenario where only GD correction is executed without the execution of OTII in the denoising process.

Quantitative results of ablation studies are shown in Tab. 3. From the result shown below, we could come to the claim that the OTII helps improve the performance on the domain level, and the transformation between ERP and TP images provides information fusion among adjacent TP images. Our proposal of

Table 2: Comparison on $\times 4$ SR task with supervised methods trained on ODI-SR dataset, including SwinIR and OSRT. The best results are shown in **bold**.

Method	Dataset	WS-PSNR↑	WS-SSIM↑	FID↓	LPIPS↓	NIQE↓	DISTS↓
SwinIR [29]	ODI-SR	26.76	0.7620	27.94	0.3321	5.3961	0.1710
OSRT [60]	ODI-SR	**26.89**	**0.7646**	**27.39**	0.3258	5.4364	0.1695
OmniSSR	ODI-SR	25.77	0.7279	30.97	**0.2977**	**5.2891**	**0.1541**
SwinIR [29]	SUN 360	26.02	0.7692	39.90	0.3419	5.2440	0.1325
OSRT [60]	SUN 360	**26.33**	**0.7766**	39.22	0.3364	5.2984	0.1312
OmniSSR	SUN 360	26.01	0.7481	**34.58**	**0.2963**	**5.1329**	**0.1299**

Table 3: Ablation studies of OmniSSR on input type, OTII, and GD correction, on the test set of the ODI-SR dataset. Best results are shown in **bold**.

Input type	OTII	GD Correction	WS-PSNR↑	WS-SSIM↑	FID↓	LPIPS↓
ERP	×	×	22.69	0.6458	44.87	0.3039
TP	×	×	23.53	0.6849	43.91	0.3113
TP	✓	×	23.74	0.6847	65.35	0.3748
TP	×	✓ (in post-process only)	26.77	0.8192	15.41	0.1691
TP	✓	✓	**28.58**	**0.8540**	**13.01**	**0.1575**

Gradient Decomposition corrects such restoration result, improving fidelity and realness significantly at the same time, and it would be better if it is applied at each step of the overall denoising pipeline. Tab. 4 shows the effect of mitigating information loss via proposed pre-upsampling strategy.

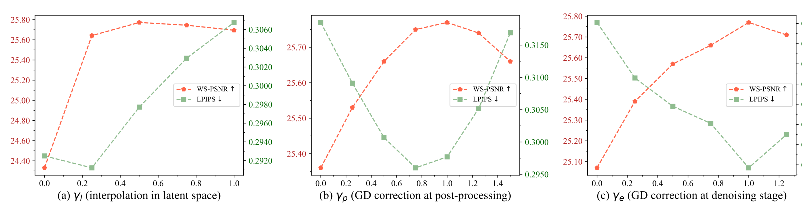

Fig. 6: Ablation of choices on γ_p, γ_e and γ_l. For better readability, WS-PSNR and LPIPS are chosen as evaluation metrics for fidelity and visual quality, respectively, to demonstrate the performance under different choices of the gamma parameter. We illustrate the results of (a) γ_p and γ_e fixed, while adjusting γ_l; (b) γ_e and γ_l fixed, while adjusting γ_p; (c) γ_p and γ_l fixed, while adjusting γ_e. It can be observed that when $\gamma_p = 1$, $\gamma_e = 1$, and $\gamma_l = 0.5$, OmniSSR achieves the relatively best performance.

For γ in the GD correction technique, we use grid search to obtain better results on ODI-SR dataset and $\times 4$ SR task. Fig. 6 shows performance on different choices of γ_p in Algo. 1 line 12, γ_e in Algo. 2 line 8, and γ_l in Algo. 2 line 11. The entire ablation of γ_p, γ_e and γ_l, with WS-PSNR, WS-SSIM, FID and LPIPS score all calculated and compared, will be provided in Supplementary Materials.

To evaluate the generalizability of our proposed modules, including Pre-Upsampling, OTII, and GD correction, we further conducted ablation studies

Table 4: Results of pre-upsampling strategy on different scales, where (x,y) denotes bicubic-based upsampling at $x\times$ scale to ERP before ERP→TP, and $y\times$ scale to TP before TP→ERP transformation. Best results are shown in **bold**.

ERP→TP→ERP	(1, 1)	(1, 4)	(4, 1)	(4, 2)	(2, 4)	(4, 4)
WS-PSNR↑	28.98	38.11	28.99	33.91	38.05	**38.18**
WS-SSIM↑	0.8859	0.9838	0.8862	0.9626	0.9837	**0.9841**

on two super-resolution backbones, StableSR and SwinIR. The results underscore substantial performance enhancements facilitated by our modules across both backbones, which is provided in Supplementary Materials.

5 Limitation and Discussion

Although OmniSSR bridges the gap between omnidirectional and planar images, achieving competitive performance and better visual results in ODISR, it still exhibits the following limitations: (1) The inference of the diffusion model requires a considerable amount of time, approximately 14 minutes per ERP-formatted omnidirectional image to be super-resolved into size 1024×2048, making real-time super-resolution challenging; (2) Multiple conversions between ERP and TP are required in the pipeline, leading to improved performance but consuming additional inference time; (3) Further exploration of the convex optimization properties of GD correction is warranted, such as designing gradient term coefficients adaptive to reconstruction results and degradation types.

This study explores the application of image generation models to ODISR tasks. In future work, the framework behind OmniSSR can be extended beyond the confines of image super-resolution in a single scenario and venture into more complex ODI-based real-world scenarios. These include ODI editing, ODI inpainting, enhancing the quality of 3D Gaussian Splatting scenes [26,41] obtained after super-resolving ERP images, as well as enhancing the quality of omnidirectional videos [48], and other possible diffusion-based applications [57, 68].

6 Conclusion

This paper leverages the image prior of Stable Diffusion (SD) and employs the Octadecaplex Tangent Information Interaction (OTII) to achieve *zero-shot* omnidirectional image super-resolution. Additionally, we propose the Gradient Decomposition (GD) correction based on convex optimization algorithms to refine the initial super-resolution results, enhancing the fidelity and realness of the restored images. The superior performance of our proposed method, OmniSSR, is demonstrated on benchmark datasets. By bridging the gap between omnidirectional and planar images, we establish a training-free approach, mitigating the data demand and over-fitting associated with end-to-end training. The application scope of our method can be further extended to various applications, presenting potential value across multiple visual tasks.

References

1. An, H., Zhang, X.: Perception-oriented omnidirectional image super-resolution based on transformer network. In: Proceedings of the IEEE International Conference on Image Processing (ICIP) (2023)
2. Arican, Z., Frossard, P.: Joint registration and super-resolution with omnidirectional images. IEEE Transactions on Image Processing (TIP) (2011)
3. Chan, K.C., Xu, X., Wang, X., Gu, J., Loy, C.C.: Glean: Generative latent bank for image super-resolution and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2022)
4. Chan, S.H., Wang, X., Elgendy, O.A.: Plug-and-play admm for image restoration: Fixed-point convergence and applications. IEEE Transactions on Computational Imaging (2016)
5. Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
6. Chen, Z., Zhang, Y., Gu, J., Kong, L., Yang, X., Yu, F.: Dual aggregation transformer for image super-resolution. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV) (2023)
7. Cheng, M., Ma, H., Ma, Q., Sun, X., Li, W., Zhang, Z., Sheng, X., Zhao, S., Li, J., Zhang, L.: Hybrid transformer and cnn attention network for stereo image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
8. Chong, M., Yanze, W., Xintao, W., Chao, D., Jian, Z., Ying, S.: Metric learning based interactive modulation for real-world super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV) (2022)
9. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. arXiv preprint arXiv:2209.14687 (2022)
10. Chung, H., Sim, B., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2022)
11. Coxeter, H.S.M.: Introduction to geometry. John Wiley & Sons, Inc. (1961)
12. Daras, G., Dean, J., Jalal, A., Dimakis, A.: Intermediate layer optimization for inverse problems using deep generative models. In: Proceedings of the International Conference on Machine Learning (ICML) (2021)
13. Deng, X., Wang, H., Xu, M., Guo, Y., Song, Y., Yang, L.: Lau-net: Latitude adaptive upscaling network for omnidirectional image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
14. Deng, X., Wang, H., Xu, M., Li, L., Wang, Z.: Omnidirectional image super-resolution via latitude adaptive network. IEEE Transactions on Multimedia (TMM) (2022)
15. Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Image quality assessment: Unifying structure and texture similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2020)
16. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2015)
17. Duan, H., Zhai, G., Min, X., Zhu, Y., Fang, Y., Yang, X.: Perceptual quality assessment of omnidirectional images. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS) (2018)

18. Fakour-Sevom, V., Guldogan, E., Kämäräinen, J.K.: 360 panorama super-resolution using deep convolutional networks. In: Proceedings of the Int. Conf. on Computer Vision Theory and Applications (VISAPP) (2018)
19. Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., Dai, B.: Generative diffusion prior for unified image restoration and enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
20. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2014)
21. Guo, L., Tao, T., Cai, X., Zhu, Z., Huang, J., Zhu, L., Gu, Z., Tang, H., Zhou, R., Han, S., et al.: Cas-diffcom: Cascaded diffusion model for infant longitudinal super-resolution 3d medical image completion. arXiv preprint arXiv:2402.13776 (2024)
22. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2017)
23. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2020)
24. Jiang, Y., Chan, K.C., Wang, X., Loy, C.C., Liu, Z.: Reference-based image and video super-resolution via c^2-matching. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2022)
25. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. In: Proceedings of the ICLR Workshop on Deep Generative Models for Highly Structured Data (ICLRW) (2022)
26. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (TOG) (2023)
27. Li, W., Chen, B., Zhang, J.: D3c2-net: Dual-domain deep convolutional coding network for compressive sensing. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) (2024)
28. Li, Y., Guo, Y., Yan, Z., Huang, X., Duan, Y., Ren, L.: Omnifusion: 360 monocular depth estimation via geometry-aware fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
29. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (2021)
30. Liu, J., Wang, Q., Fan, H., Wang, Y., Tang, Y., Qu, L.: Residual denoising diffusion models. arXiv preprint arXiv:2308.13712 (2023)
31. Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., Zeng, T.: Transformer for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2022)
32. Lugmayr, A., Danelljan, M., Timofte, R.: Ntire 2020 challenge on real-world image super-resolution: Methods and results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020)
33. Menon, S., Damian, A., Hu, S., Ravi, N., Rudin, C.: Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
34. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters (SPL) (2013)

35. Nishiyama, A., Ikehata, S., Aizawa, K.: 360° single image super resolution via distortion-aware network and distorted perspective images. In: Proceedings of the IEEE International Conference on Image Processing (ICIP) (2021)
36. Ozcinar, C., Rana, A., Smolic, A.: Super-resolution of omnidirectional images using adversarial learning. In: Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSPW) (2019)
37. Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C.C., Luo, P.: Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2021)
38. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
39. Rout, L., Raoof, N., Daras, G., Caramanis, C., Dimakis, A., Shakkottai, S.: Solving linear inverse problems provably via posterior sampling with latent diffusion models. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2023)
40. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2022)
41. Schönbein, M., Geiger, A.: Omnidirectional 3d reconstruction in augmented manhattan worlds. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2014)
42. Song, J., Zhang, Q., Yin, H., Mardani, M., Liu, M.Y., Kautz, J., Chen, Y., Vahdat, A.: Loss-guided diffusion models for plug-and-play controllable generation. In: Proceedings of the International Conference on Machine Learning (ICML) (2023)
43. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: Proceedings of the International Conference on Learning Representations (ICLR) (2020)
44. Sun, X., Li, W., Zhang, Z., Ma, Q., Sheng, X., Cheng, M., Ma, H., Zhao, S., Zhang, J., Li, J., et al.: Opdn: Omnidirectional position-aware deformable network for omnidirectional image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2023)
45. Sun, Y., Lu, A., Yu, L.: Weighted-to-spherically-uniform quality evaluation for omnidirectional video. IEEE Signal Processing Letters (SPL) (2017)
46. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2017)
47. Wang, J., Yue, Z., Zhou, S., Chan, K., Loy, C.: Exploiting diffusion prior for real-world image super-resolution. arXiv preprint arXiv:2305.07015 (2023)
48. Wang, Q., Li, W., Mou, C., Cheng, X., Zhang, J.: 360dvd: Controllable panorama video generation with 360-degree video diffusion model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
49. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
50. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision Workshops (ECCVW) (2018)
51. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. arXiv preprint arXiv:2212.00490 (2022)

52. Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Yang, W., Van Gool, L.: Diffir: Efficient diffusion model for image restoration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
53. Xiao, J., Ehinger, K.A., Oliva, A., Torralba, A.: Recognizing scene viewpoint using panoramic place representation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2012)
54. Yagi, Y.: Omnidirectional sensing and its applications. IEICE Transactions on Information and Systems (TOIS) (1999)
55. Yamazawa, K., Yagi, Y., Yachida, M.: Omnidirectional imaging with hyperboloidal projection. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (1993)
56. Yang, S., Zhou, Y., Liu, Z., Loy, C.C.: Rerender a video: Zero-shot text-guided video-to-video translation. In: Proceedings of the SIGGRAPH Asia 2023 Conference Papers (SIGGRAPH Asia) (2023)
57. Yang, S., Mou, C., Yu, J., Wang, Y., Meng, X., Zhang, J.: Neural video fields editing. arXiv preprint arXiv:2312.08882 (2023)
58. Yinhuai, W., Yujie, H., Jiwen, Y., Jian, Z.: Gan prior based null-space learning for consistent super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2023)
59. Yoon, Y., Chung, I., Wang, L., Yoon, K.J.: 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 (CVPR) (2022)
60. Yu, F., Wang, X., Cao, M., Li, G., Shan, Y., Dong, C.: Osrt: Omnidirectional image super-resolution with distortion-aware transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
61. Yu, J., Zhang, X., Xu, Y., Zhang, J.: Cross: Diffusion model makes controllable, robust and secure image steganography. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2023)
62. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. In: Advances in Neural Information Processing Systems (NeurIPS) (2023)
63. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
64. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
65. Zhang, W., Li, X., Shi, G., Chen, X., Qiao, Y., Zhang, X., Wu, X.M., Dong, C.: Real-world image super-resolution as multi-task learning. In: Advances in Neural Information Processing Systems (NeurIPS) (2023)
66. Zhang, X., Zhang, Y., Xiong, R., Sun, Q., Zhang, J.: Herosnet: Hyperspectral explicable reconstruction and optimal sampling deep network for snapshot compressive imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
67. Zhou, Y., Yu, M., Ma, H., Shao, H., Jiang, G.: Weighted-to-spherically-uniform ssim objective quality evaluation for panoramic video. In: Proceedings of the IEEE International Conference on Signal Processing (ICSP) (2018)

68. Zongying, L., Hao, L., Liuzhenghao, L., Bin, L., Junwu, Z., Yu-Chian, C.C., Li, Y., Yonghong, T.: Taxdiff: Taxonomic-guided diffusion model for protein sequence generation. arXiv preprint arXiv:2402.17156 (2024)