Dual Enhancement in ODI Super-Resolution: Adapting Convolution and Upsampling to Projection Distortion

Xiang Ji$^1$, Changqiao Xu$^1$,∗, Lujie Zhong$^2$, Shujie Yang$^1$, Han Xiao$^1$ and Gabriel-Miro Muntean$^3$

$^1$State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications
$^2$Information Engineering College, Capital Normal University
$^3$School of Electronic Engineering, Dublin City University
{jxian_2019, cqxu, sjyang, xiaohan}@bupt.edu.cn, zhonglj@cnu.edu.cn, gabriel.muntean@dcu.ie

Abstract

Omnidirectional images (ODIs) demand considerably higher resolution to ensure high quality across all viewports. Traditional convolutional neural networks (CNN)-based single image super-resolution (SISR) networks, however, are not effective for spherical ODIs. This is due to the uneven pixel density distribution and varying texture complexity in different regions that arise when projecting from a sphere to a plane. Additionally, the computational and memory costs associated with large-sized ODIs present a challenge for real-world applications. To address these issues, we propose an efficient distortion-adaptive super-resolution network (ODA-SRN). Specifically, ODA-SRN employs a series of specially designed Distortion Attention Block Groups (DABG) as its backbone. Our Distortion Attention Blocks (DABs) utilize multi-segment parameterized convolution to generate dynamic filters, which compensate for distortion and texture fading during feature extraction. Moreover, we introduce an upsampling scheme that accounts for the dependence of pixel position and distortion degree to achieve pixel-level distortion offset. A comprehensive set of results demonstrates that our ODA-SRN significantly improves the super-resolution performance for ODIs, both quantitatively and qualitatively, when compared to other state-of-the-art methods.

1 Introduction

Omnidirectional images (ODIs), as a novel visual form of virtual reality (VR) [Zhang et al., 2023], are anticipated to become integral components of future social environments, enriching and enhancing the ways in which we share experiences. Consequently, ODIs are being introduced rapidly into people’s lives. As their name suggests, ODIs offer a $360° \times 180°$ field of view, and the expansive visual range naturally demands high resolution. The level of detail and sharpness in ODIs directly influences the realism of user experiences. However, due to storage and transmission constraints, omnidirectional images often lack the necessary resolution to provide users with a truly immersive experience. Super-resolution (SR) technology transforms low-resolution (LR) images into high-resolution (HR) ones. As a long-standing computer vision problem, single-image super-resolution (SISR) has attracted extensive attention from researchers [Anwar et al., 2021]. In recent years, several SISR studies based on convolutional neural networks (CNN) have achieved remarkable results [Senmao et al., 2023; Wang et al., 2021]. These studies utilize CNNs to extract features from images and subsequently employ the extracted features for upsampling. CNNs use convolution kernels to sense images through sliding windows, allowing them to capture local patterns. However, these works primarily target two-dimensional (2D) planar images, neglecting the global distortion information present in omnidirectional images (ODIs). For ODIs, equirectangular projection (ERP) from sphere to plane is widely employed for convenient storage and transportation. Regrettably, the distortion caused by ERP results in unevenly distributed features within ODIs. Fig. 1 illustrate the correlation between distortion degree and pixel position in ODI projection. Notably, stretch distortion is particularly severe in non-horizontal regions, often leading to unsatisfac-

$^*$Corresponding author.
tory feature extraction results for ODIs. This phenomenon raises the question: How can we achieve super-resolution in the presence of distortion and improve CNNs to effectively extract features from ODIs?

For ODI super-resolution (SR), traditional CNN-based feature extraction strategies have proven ill-suited for ODIs in practice. The issue arises from the fact that both existing CNN filters and image data are “flat”, making it challenging for these filters to achieve the right balance with ODIs that exhibit varying pixel densities and texture complexities. Recently, [Deng et al., 2021] proposed a multilevel CNN network that allocates additional computational resources to the central regions of ODIs. Their approach modified existing SR networks but did not fundamentally enhance the sensitivity of CNNs to ODI distortion. Conversely, most existing ODI SR research works [Ozcinar et al., 2019; Su and Grauman, 2017; Sun et al., 2021] tend to focus on the local resolution of ODIs. To mitigate the effects of distortion, these methods process tangent planes surrounding the sphere repeatedly. While this boosts accuracy, it also significantly heightens computational costs. These methods are heavily dependent on hardware performance, which poses a considerable burden for portable devices like head-mounted displays and cell phones.

In an effort to strike a balance between accuracy and cost, numerous CNN-based SR methods [Dai et al., 2019; Liu et al., 2020; Mei et al., 2020; Zhang et al., 2021] have introduced attention mechanisms. [Dai et al., 2019] proposed a deep second-order attention network to capture spatial context information over long distances. [Liu et al., 2020] incorporated spatial attention into the residual network to enhance model efficiency. Drawing on recent neuroscience research, [Zhang et al., 2021] proposed CRAN, which utilizes semantic knowledge as a factor influencing network attention. However, these attention mechanisms appear to be ineffective in addressing the distortion caused by projection. This is primarily because these methods focus on the comprehensive utilization of global information without considering distortion shifts. These challenges limit the representational power of CNNs, and as a result, the development of an optimal single ODI SR network remains an ongoing endeavor.

Motivated by the observations and analyses discussed above, we propose a novel method named Omnidirectional Distortion Attention Super-Resolution Networks (ODA-SRN). Contrary to existing strategies, the core idea of our ODA-SRN is the dual enhancement of two fundamental components in SR, convolution and upsampling, rather than designing intricately structured integrated network models. Our innovation resides in considering the global and local differences of ERP during the convolution process, and introducing pixel-level offsets in the upsampling phase to accommodate distortion. This focus on enhancing fundamental components, rather than simply stacking larger models, enables our ODA-SRN to efficiently conserve computational and memory resources.

Our main contributions can be summarized as follows:

• We have introduced novel convolution and upsampling techniques for ODIs. These components have been integrated into our ODA-SRN, achieving superior SR results.

• We propose extracting the potential representation of distortion information to obtain a set of routing weights. By leveraging the relationship between routing weights and segment positions, we enhance the distortion sensitivity of CNN.

• We introduce a Jacobian matrix of spatially transformed elements and mathematically adjust the sampling window of the upsampling layer. The reusable Jacobian matrix improves the efficiency of our ODA-SRN.

2 Related Work

Deep CNN for Super Resolution. Over the past decade, SR has been one of the most extensively studied inverse problems [Chen et al., 2022]. The combination of SR and CNN technology has led to significant progress in reconstruction error and practical applications [Chen et al., 2021; Ma et al., 2020; Senmao et al., 2023]. Since [Dong et al., 2014] first proposed SRCNN, numerous SR works based on deep CNN have been introduced. These works have made improvements in various aspects, such as increasing network depth to utilize more parameters, incorporating residual attention block mechanisms to enhance performance accuracy, and modifying the loss function [Li et al., 2022]. All of these works have achieved remarkable results in SR for 2D images. However, a common characteristic of CNN-based SISR methods is that their “flat” convolutional layers are inefficient for ODIs in feature extraction. Specifically, due to the uneven pixel density and varying texture complexity of ODIs, it is challenging to optimize a convolutional kernel that can account for the effects of all degrees of distortion.

Super Resolution for VR. With the rise of virtual reality (VR) technology, omnidirectional super-resolution has emerged as a challenging task for the research community. In recent years, viewport-based methods have been proposed as a promising approach for enhancing large-sized ODIs [Li et al., 2020]. [Su and Kristen, 2017] utilized different convolution kernels to extract features along various latitudes on ODIs; however, this approach required training a large number of convolution kernels. Inspired by the success of the Transformer architecture, [Shen et al., 2022] proposed PanoFormer, a deep neural network capable of estimating depth in ODIs. SphereSR [Yoon et al., 2022] incorporated RGB prediction into feature extraction and learned the relationships between different ODI formats to achieve more flexible SR. However, using Transformers for ERP-format ODIs poses challenges: their global attention may overlook local nuances critical for SR, and their computational demands can be burdensome given the large-size nature of ODIs.

Techniques for Projection Distortion. Projection distortion in ODIs tends to limit regions showing varied distortion degrees. In the realm of efficient adaptability, parameterized convolutions have emerged as a promising technique [Brandon Yang, 2020]. Unlike traditional convolutional layers that utilize static filters, parameterized convolutions change their filters based on the input. However, current
attention mechanisms are not guided by distortion information, which limits their ability to address the regional differences in ODIs caused by projection.

3 Proposed Method

Overview. We designed a concise network architecture, illustrated in Fig. 2. It demonstrates how a low-resolution ODI input ($I_{LR} \in \mathcal{R}^{c \times h \times w}$) generates a high-resolution ODI ($I_{SR} \in \mathcal{R}^{c \times sh \times sw}$), where $s$ is a multiple of the super-resolution, $c$ represents the number of channels, and $h$ and $w$ denote the ODI size.

3.1 Network Architecture

ODA-SRN mainly consists of two parts: the feature extraction part and the upsampling recovery part. First, shallow features $F_0$ are extracted from the input LR ODI:

$$F_0 = f_{conv}(I_{LR}),$$

where $f_{conv}$ represents a $3 \times 3$ convolution operation, and $F_0$ is used as input to DABGs (Distortion Attention Block Groups). Furthermore, each DABG contains several DAB modules. We adopted a residual-in-residual (RIR) structure as the backbone of our network. In this way, we aim to superimpose low-frequency and high-frequency feature information, producing an SR image with more texture details. The $i$-th DAB group is represented as follows:

$$F_i = F_{i-1} \oplus W_{SSC} f^m_{DAB}(\ldots f^1_{DAB}(F_{i-1}, M_d), \ldots, M_d),$$

where $\oplus$ represents dot multiplication, and $W_{SSC}$ is the weight associated with the DAB group output. $f^k_{DAB}$ represents the $k$-th distortion attention block operation. Distortion information $M_d$ contains all ODI distortion information $D_{x,y}$. The mathematical representation of the ERP distortion information is as follows:

$$D_{x,y} = \cos \left( \left( y - \frac{H}{2} + \frac{1}{2} \right) \frac{\pi}{H} \right),$$

$H$ represents the height of the ODI, and we denote $1 - D_{x,y}$ as the distortion degree of pixel $(x, y)$. Note that the higher the $D_{x,y}$ value, the less the distortion. Eq. (3) is derived from the JVET standard D0040 [Sun et al., 2016], which defines evaluation metrics for omnidirectional images. After the data stream passes through the DABGs, the deep information of the ODI is extracted and provided to the subsequent module for upsampling and reconstruction.

$$I_{SR} = f_{REC}(f_{UP}(F_0 + W_{LSC} * F_n)), \quad (4)$$

Eq. (4) shows the upsampling and reconstruction part. Similar to Eq. (2), $W_{LSC}$ represents the weight associated with the deep feature $F_n$. Eq. (4) indicates how, after fusion, the features are sent to the distortion offset upsampling layer $f_{UP}$. Finally, through the reconstruction process $f_{REC}$, the upsampled features are converted to SR ODI, which can be upsampled by $s \times (s$ is an integer, e.g., $2, 3, \ldots$).

3.2 Enhanced Distortion Attention Block (DAB) Methodology

The main challenge of feature extraction is the distortion caused by ERP. Previous CNN-based feature extraction methods treat the features of all LR regions equally and cannot sense the distortions in a targeted way. As shown in Fig. 3, we designed a Distortion Attention Block (DAB) mixed with general convolution layers and a distortion attention layer to learn features in a “total-local” manner. The DA layer is responsible for piecewise learning according to the distortion degree. Meanwhile, the general convolutional layer appears alternately to supplement the global association information.

Distortion Attention layer. We attempt to incorporate the global distortion information into the convolution process. As shown in Fig. 3, the input features and distortion degree & scale information are fed into the Distortion Mask (DM) module. DM cuts the feature maps into $k$ segments (Section 4 provides a more comprehensive analysis and guidance on the choice of this parameter). The calculation of the segmentation position $p$ follows the equal distortion principle, which means that the difference between the maximum and
the minimum distortion degree of each segment are equal. After segmenting transversely \( k - 1 \) times, we represent the set of segments as \( \{G_1, G_2, \ldots, G_k\} \). As depicted in Fig. 3, this set of segments is introduced to a network to generate a two-dimensional routing weights matrix. The learning process of the routing weights matrix \( W \in \mathbb{R}^{k \times d} \) from our network is defined as:

\[
W = \sigma(\text{BN}(\text{ReLU}(\text{FC2}(\text{BN}(\text{ReLU}(\text{FC1}(\text{Pool}(G))))))))
\]  
(5)

where \( \sigma \) can be a Softmax function, BN signifies batch normalization, ReLU represents the Rectified Linear Unit activation function, and Pool indicates pooling. For each segment \( G_i \), we extract a weight vector \( w_i \) from the routing matrix \( W \). Each element \( w_{ij} \) in \( w_i \) quantifies the influence of the \( j \)-th expert (a conditionally applied convolutional kernel) on the \( i \)-th segment. To normalize weights, we employ a softmax operation over each expert pertaining to segment \( G_i \):

\[
\alpha_{ij} = \frac{\exp(w_{ij})}{\sum_{j=1}^{d} \exp(w_{ij})}
\]  
(6)

Given this weight normalization, we proceed to combine these weights with the routing matrix \( W \) through a Hadamard product. This results in a tailored distortion-aware filter \( P_i \). Through the procedures delineated above, we’ve refined traditional convolution layers. A standout feature of the DA Layer is its adaptability: it can adjust its convolutional filters based on the nuances of the incoming feature maps. Importantly, it ensures efficiency, circumventing the computational overhead associated with training a larger number of filters thus marrying efficiency with model precision.

A comparative analysis was conducted against the feature extraction attention block incorporated in RCAN [Zhang et al., 2018]. As shown in Fig. 5, with the deepening of the network, the traditional feature extraction block has poor performance for ODIs affected by distortion, and the features in the severely distorted regions appear to fade (Feature vanishing in the polar region.). We found that traditional feature extraction blocks cannot find a balance between regions affected by different distortion degrees. In contrast, our DABs can still efficiently extract features from different regions affected by distortion at the same time in deep networks.

3.3 Distortion Offset Upsampling (DOU) Layer

The upsampling layer is a pivotal component, leveraging deep information to reconstruct larger-sized feature maps. For ODIs, the increased sampling density at the sphere’s poles engenders information redundancy in these polar regions. We conceptualize the SR process for ODIs as a three-tiered progression: from the sphere to a low-definition plane, and to a high-definition plane, as illustrated in Fig. 4(a). Our analysis seeks to determine the tile-to-pixel area ratio before and after projection, termed as the zoom factor \( \delta \). This metric is instrumental in gauging the influence of planar pixels on spherical visual perceptions. With ERP mapping a point \((X, Y, Z)\) on


Table 1: Comparisons Between the SR results. The 1st and the 2nd best performances are highlighted in red and blue.

<table>
<thead>
<tr>
<th>Method</th>
<th>×4 (ODI-SR &amp; SUN 360)</th>
<th>×8 (ODI-SR &amp; SUN 360)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ERP Format</td>
<td>Viewport Format</td>
</tr>
<tr>
<td></td>
<td>WS-PSNR</td>
<td>WS-SSIM</td>
</tr>
<tr>
<td>Bicubic</td>
<td>27.64</td>
<td>0.6065</td>
</tr>
<tr>
<td>SRCNN[ Dong et al., 2014]</td>
<td>28.08</td>
<td>0.6148</td>
</tr>
<tr>
<td>MemNet[ Tai et al., 2017]</td>
<td>28.31</td>
<td>0.6442</td>
</tr>
<tr>
<td>RCAN[Zhang et al., 2018]</td>
<td>28.78</td>
<td>0.6581</td>
</tr>
<tr>
<td>EDSR[Lim et al., 2017]</td>
<td>28.74</td>
<td>0.6571</td>
</tr>
<tr>
<td>360-SS[Ozcinar et al., 2019]</td>
<td>28.19</td>
<td>0.6486</td>
</tr>
<tr>
<td>LAU-Net+[ Deng et al., 2023]</td>
<td>29.13</td>
<td>0.6751</td>
</tr>
<tr>
<td>OSRT[ Yu et al., 2023]</td>
<td>29.13</td>
<td>0.6751</td>
</tr>
<tr>
<td>ODA-SRN(ours)</td>
<td>29.47</td>
<td>0.6786</td>
</tr>
</tbody>
</table>

Figure 6: Visual comparison (ERP form) of ×4 SR results with different methods in the central and polar regions of ODI.

the sphere to \((x_{LR}, y_{LR})\) on the plane, we delve into its reverse process as delineated below:

\[
\begin{align*}
X &= \cos(\pi(0.5 - v)) \cdot \cos(2\pi(u - 0.5)), \\
Y &= \sin(\pi(0.5 - v)), \\
Z &= -\cos(\pi(0.5 - v)) \cdot \sin(2\pi(u - 0.5)),
\end{align*}
\]

where \((u, v)\) represents the normalized \((x_{LR}, y_{LR})\), the derivation of Eq. (5) is omitted for brevity. Based on linear algebra, we calculate the Jacobian \(J(u, v)\) of the space projection using the following approach:

\[
\bar{x}_{\text{sphere}} = \begin{bmatrix} \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} & \frac{\partial Z}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} & \frac{\partial Z}{\partial v} \end{bmatrix}^{T} \bar{x}_{LR} = J^{T}(u, v)\bar{x}_{LR},
\]

The Jacobian \(J(u, v)\) can be used to obtain the mapping of any 2D vector in 3D space. We use \((u, v)\) and \((\Delta u, \Delta v)\) to represent any pixel in the plane. Through Eq. (6), we can obtain the four vectors corresponding to the tile in 3D space. We approximate the tile area as the sum of two triangle areas, as shown in Fig. 4(a). We use the cross product of two vectors to compute the areas of \(A_1\) and \(A_2\). It should be noted that the zoom factor \(\delta_{u,v}\) and the Jacobian are reusable, making this approach computationally efficient and cost-friendly. In the process of converting low-resolution pixels to high-resolution pixels, we calculate the relative distance between LR pixels and HR pixels. The value of the HR pixel \((x_{HR}, y_{HR})\) is unknown, but we can obtain the position of the LR pixel \((x_{LR}, y_{LR})\) corresponding to the HR pixel. Further, we can determine their correlation distances \(R(x_{HR})\) and \(R(y_{HR})\). As shown in Fig. 4(a), where \(\lfloor \cdot \rfloor\) indicates integer rounding down. \(R(y_{HR})\) are calculated in the same way. The correlation distance represents the bias relationship between HR pixels and LR pixels. Crucially, excessive reconstruction of
The training goal is to minimize \( L \). ODI SR is currently underexplored, and few features of 4.1 Datasets and Metrics

4.2 Implementation Details

For the model setup, the final version of ODA-SRN consists of 5 DABGs, with each DABG containing 2 DABs. The number of experts in the DA layer is set to 3, Coefficent \( \gamma \) is set to 0.1. Our model was trained using the ADAM optimizer [Kingma and Ba, 2015] with \( \beta_1 = 0.9, \beta_2 = 0.999 \), and \( \varepsilon = 10^{-8} \). The learning rate was set to \( 10^{-4} \), decaying by half every 50 epochs. The weight initialization of the model follows [He et al., 2015]. We used PyTorch [Paszke et al., 2017] to implement our model. During training, random crops were not used to preprocess ODIs in order to retain complete distortion information.

4.3 Comparison with State-of-the-Art Methods

In our experiments, we compared our ODA-SRN with other SOTA methods. To ensure consistent conditions across various models, each was deployed within its dedicated Docker container. This methodology was adopted to mitigate the inconsistencies stemming from divergent code implementations.
Quantitative Analysis. Table 1 shows the ODA-SRN consistently outperforms its contemporaries in a wide spectrum of evaluations. This overarching superiority of our proposed method becomes particularly pronounced when diving deeper into the results of the viewport format, a pivotal metric that simulates what a user would witness through their Head-Mounted Display (HMD). Such an unequivocal lead testifies to the meticulous engineering of the ODA-SRN, tailored to replicate and enhance the user’s visual experience. An underpinning idea of our model is the segmental processing of the ODI. By partitioning the ODI regionally, each segment is processed through individualized Distortion-aware filter. This strategy addresses a perennial challenge: mediating the inherent tug-of-war between varying distortion regions during the feature extraction phase. By localizing the processing, the ODA-SRN can give specialized attention to each region’s idiosyncratic distortions. The results, as illuminated in the Table 1, attest to the efficacy of this approach.

Qualitative Analysis. Fig. 6 visually presents the results of the ×4 super-resolution performances by various algorithms. Distinct regions from the image, central and polar, are highlighted using red and green boxes for clearer examination. On the other hand, methods such as EDSR, 360-SS, and OSRT, despite their sharp outputs, occasionally exhibit discrepancies in their outline representations. On the left of Fig. 7, the transformation of an image before and after ERP projection is showcased. Through VR equipment, users experience a spherical field of view. However, when this image undergoes ERP projection onto a two-dimensional plane, perceptual distortions arise. For instance, the skyscrapers that span across both the central and polar regions of the image exhibit window outlines that, while straight in the original three-dimensional context, become curved upon projection.

The right portion of Fig. 7 offers a comparative analysis of the results generated by various algorithms. Traditional 2D-oriented methods, like EDSR and RCAN, tend to restore features by treating the ERP-projected image as a regular 2D grid. This method, however, fails to accommodate the curvatures introduced by the ERP projection. As a result, when their outputs are translated back to a spherical perspective, the previously straight window contours appear misshapen and curved. While methods explicitly designed for ODIs, such as 360SS and LAU-Net+, make commendable efforts to tackle the anomalies caused by ERP projection, they still face challenges, particularly in extreme polar regions. These areas, depicted by the side of the blue building and the apex of the white structure in Fig. 7, are marked by results that are overly smoothed and lacking in contour fidelity. This limitation stems from the inherent sparsity of feature information in these polar regions, making the precise restoration of details a formidable task. ODA-SRN is designed to address the scarcity of features in the polar regions, the distortion-offset upsampling layer of ODA-SRN adopts a unique approach. By leveraging a receptive field offset mechanism, it pulls richer feature details from more central areas. This allows ODA-SRN to supplement the sparse features in the extremities and bring a marked improvement in restoring details. The result is a more accurate representation of even subtle elements, such as the high-rise window outlines, when they are projected back to the spherical viewpoint.

Anti-distortion Analysis. ERP projection significantly stretches polar regions in images, where inaccuracies can manifest during spherical conversion. We illustrate this with a comparison against various SR algorithms. Fig.8 shows that vents in ERP images are notably stretched. Our method effectively reduces this distortion, with ODA-SRN outperforming other tested methods in anti-distortion capability.

Model Efficiency Analysis. We have included Table 2. This table provides a comprehensive performance comparison of our method with other algorithms, covering various aspects such as FLOPs, WS-PSNR, network parameters, and runtime. By presenting this multidimensional comparison, we aim to offer a thorough evaluation of the effectiveness and efficiency of our approach compared to existing methods. ODA-SRN strikes a balance between performance and...
efficiency, delivering the highest WS-PSNR using fewer parameters, reduced FLOPs, and enhanced speed.

4.4 Ablation Study

Ablation Study on Weighted Loss Function. This study examines the zoom factor $\delta$ in loss calculation, which reflects a pixel’s projection scale from planar to spherical views. Table 3 demonstrates the superiority of the $\delta$-weighted loss function over conventional $L_1$ loss in SR.

Ablation Study on Distortion Attention Layer (DA). In the absence of the DA layer. Considering the unique patterns in ODIs in ERP format, Table 3 shows that integrating the DA layer significantly enhances WS-PSNR and WS-SSIM. Our DA layer, focusing on specific ODI features, improves structural similarity in SR outcomes. This is due to ODA-SRN’s adaptive feature extraction for different ODI regions.

Ablation Study on Distortion Offset Upsampling (DOU). Data in Table 3 emphasizes the effectiveness of DOU in ODI super-resolution. Standard upsampling doesn’t adequately address ODI specifics, leading to the creation of the DOU layer. This layer computes $\theta_{\text{offset}}$ to amend inherent distortions, allowing polar-proximate pixels to utilize central ODI information and enhance output quality. Compared to the pixel shuffle baseline, DOU consistently achieves superior WS-PSNR and WS-SSIM results.

Effect of Segments Number. We propose DAB for extracting features of ODI. Unlike typical CNNs, DAB adopts a segmented strategy to handle regions with varying degrees of distortion. We attempted to change the number of segments in DAB to observe the impact of different segment numbers on network performance. Table 4 shows the experimental results. With the increase in the number of segments, the performance of the network improves. However, the improvement in FLOPs is linear, which means that more segments do not always yield better results. We need to choose an appropriate number of segments to achieve a balance between performance and cost. When the number of segments increases to more than 5, the model becomes difficult to converge and performance decreases. Therefore, considering multiple factors, we finally chose 5 as the number of segments in ODA-SRN.

5 Conclusion

We addressed the SR challenges of ODIs. Our solution was not rooted in proposing a completely new network. Rather, our main contribution centers on the innovative redesign of two plug-and-play modules: the Distortion Attention Block and the Distortion Offset Upsampling Layer. These modules, crafted to counter the unique challenges posed by ODIs, not only enhance super-resolution performance but also offer the flexibility of being integrated into various existing network architectures. Comprehensive experiments highlight that our modular approach excels in metrics such as PSNR, SSIM, WS-PSNR, and WS-SSIM, and also stands out in terms of model efficiency.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) under grant No. 62225105, 62394323 and 62301070. Han Xiaow wishes to acknowledge support from the Postdoctoral Science Foundation of China under grant No. 2022M720518. G.-M. Muntean wishes to acknowledge the Science Foundation Ireland’s support via grant nos. 21/FFP-P/10244 (FRADIS) and 12/RC/2289 P2 (Insight).

References


