LAU-Net: Latitude Adaptive Upscaling Network for Omnidirectional Image Super-resolution

Xin Deng[1, *], Hao Wang[2, *], Mai Xu[2, †], Yichen Guo[2], Yuhang Song[3], Li Yang[2]

[1]School of Cyber Science and Technology, Beihang University, Beijing, China
[2]School of Electronic and Information Engineering, Beihang University, Beijing, China
[3]Department of Computer Science, University of Oxford, UK

{cindydeng, wang_hao, MaiXu, ycguo, LiYang2018}@buaa.edu.cn,
Yuhang.Song@some.ox.ac.uk

Abstract

The omnidirectional images (ODIs) are usually at low-resolution, due to the constraints of collection, storage and transmission. The traditional two-dimensional (2D) image super-resolution methods are not effective for spherical ODIs, because ODIs tend to have non-uniformly distributed pixel density and varying texture complexity across latitudes. In this work, we propose a novel latitude adaptive upscaling network (LAU-Net) for ODI super-resolution, which allows pixels at different latitudes to adopt distinct upscaling factors. Specifically, we introduce a Laplacian multi-level separation architecture to split an ODI into different latitude bands, and hierarchically upscale them with different factors. In addition, we propose a deep reinforcement learning scheme with a latitude adaptive reward, in order to automatically select optimal upscaling factors for different latitude bands. To the best of our knowledge, LAU-Net is the first attempt to consider the latitude difference for ODI super-resolution. Extensive results demonstrate that our LAU-Net significantly advances the super-resolution performance for ODIs. Codes are available at https://github.com/wangh-allen/LAU-Net.

1. Introduction

With the rapid development of virtual reality (VR), omnidirectional images (ODIs) are playing increasingly important roles in human's life. When viewing ODIs, people can obtain immersive and interactive experience via changing their viewports in the range of $360 \times 180°$. Typically, people watch ODIs through head-mounted displays (HMD), in which only the viewport with a limited range is visible. To

---

*Authors contributed equally.
†Corresponding author.

make this small viewport in high-resolution (HR), the whole ODI requires extremely high resolution [11]. However, due to the constraints of capture, storage and transmission , the resolution of ODIs cannot be sufficiently high.

Super-resolution (SR) is a common technique to address the aforementioned issue, which aims to restore an HR image from a single or a sequence of low-resolution (LR) images [12]. As a challenging ill-posed inverse problem, SR has received extensive study for decades [33, 45, 4, 36, 1]. However, the existing SR methods target at two-dimensional (2D) planar images, which are not appropriate for ODIs. For storage convenience, the spherical ODIs are usually projected into 2D planes. The widely used projection method is equirectangular projection (ERP), which leads to non-uniform pixel density across latitudes, in particular geometric distortion in high-latitude areas. As shown in Fig. 1, the density of pixels after ERP is in negative correlation to latitudes, i.e., the pixel distribution in higher latitudes tends to be more sparse than those in lower latitudes. In addition, the image patches at high-latitude areas usually have significant stretch distortion. Since the 2D SR methods do not consider these characteristics of ODIs, as verified in Finding 2, they often result in unsatisfactory SR results for ODIs.

For ODI SR, the existing methods primarily rely on assembling a sequence of LR ODIs to form an HR ODI. The representative works include Nagahara *et al.* [27], Arican *et al.*[2], and Bagnato *et al.* [3]. All these methods have the same disadvantage, i.e., their performance heavily depends on the number of LR images and the registration accuracy among them. Recently, Ozcinar *et al.* [28] proposed a generative adversarial network (GAN) to perceptually super-resolve the ODIs, and remove the artifacts in the spherical space. However, they merely treat the ERP projected ODI as a normal 2D image, without considering the varying pixel density across latitudes.

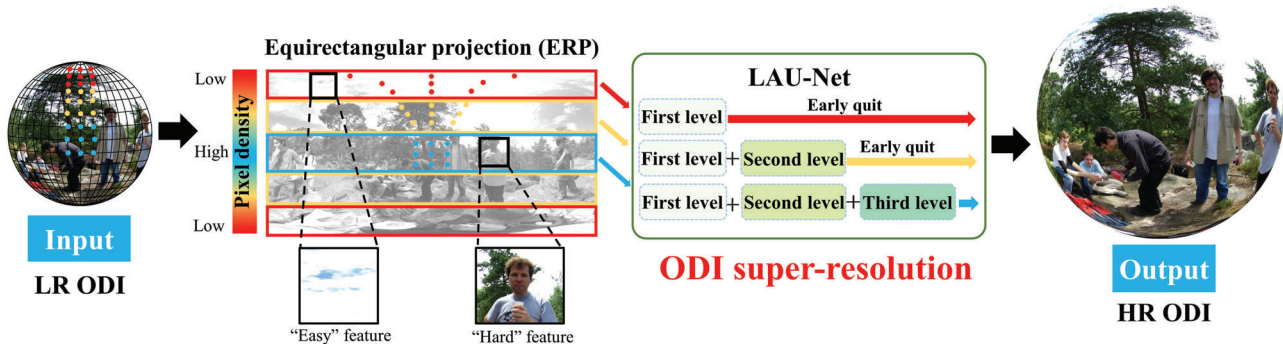In this paper, we propose a novel latitude adaptive upscal-

Figure 1. The basic framework of our method for omnidirectional image super-resolution.

ing network (LAU-Net), to dynamically upscale different latitude bands of ODIs with various upscaling factors. To determine the optimal upscaling factors for different latitude bands, we jointly train several evaluators for different bands with a multi-level CNN to find the optimal upscaling factor. The evaluators are trained by reinforcement learning (RL) with the reward encouraging both high SR performance and low computation complexity. As shown in Fig 1, "easy" patches with high latitude and low image complexity are stopped training at the first level, while "hard" patches with low latitude and high image complexity progressively go deeper until the last level. Using early quit strategy combined with RL network, our LAU-Net obtains better objective quality while saving computations effectively. The main contributions of our work are as follows:

- We establish a large database for ODI SR, which consists of 1,000 high-quality ODI images, with diverse image resolutions and content.

- We propose a new network named LAU-Net for ODI SR, in which different latitude bands are allowed to have distinct upscaling factors for resource efficiency.

- We develop an RL scheme to automatically select the optimal upscaling factors for different latitude bands, which significantly improves the SR performance using less computational resource.

2. Related work

Single image super resolution (SISR) is a long studied inverse problem. The traditional SISR methods include example-based [37, 43] and dictionary learning based [29, 44] approaches, while the recent SISR methods focus on deep neural networks [9, 18, 19, 10, 31, 19, 34, 5, 48, 7, 8, 6]. Dong *et al.* proposed the first SISR network called SRCNN [9], which achieves remarkable improvement over the traditional methods. Since then, many works are proposed to further enhance the SR performance. The representative works include VDSR [18], DRCN [19], SRResNet [38],

MemNet [35], EDSR [25], D-DPBN [15], and RCAN [48]. All these methods aim to improve the objective quality of the super-resolved images in terms of mean squared error (MSE). In order to improve the perceptual quality, Ledig *et al.* [22] proposed a generative adversarial network (GAN) for SISR, called SRGAN [22], which replaces the MSE loss with VGG loss. The SRGAN method inspires many follow-up works, like CX [26], ESRGAN [40], RankSRGAN [47], etc.

However, all the above SISR methods are proposed for standard 2D images. For ODI SR, Nagahara *et al.*[27] proposed to combine a series of LR ODIs using spatio-temporal nearest neighbor interpolation, to obtain a fused HR ODI. To handle the inaccurate alignment among LR ODIs, Arican *et al.*[2] cast the registration and SR problem as a joint least-square norm minimization problem, and solve it using a Levenberg-Marquardt method. Different from [27, 2] which require multiple LR images, Ozcinar *et al.*[28] proposed to use deep network for ODI SR, which only needs a single LR image. However, [28] did not consider the latitude difference in ODIs, i.e., all latitude bands are super-resolved by the same upscaling factor. Actually, since the high latitude area will be shrunk in the spherical domain, there is no need to upscale this area with a large factor.

Our work is the first attempt to achieve latitude adaptive ODI super-resolution. Rather than applying the same upscaling factor for all latitude bands, we allow each band to be super resolved by different factors, based on a multi-level separation and reinforcement learning scheme. To the best of our knowledge, this is the first time different upscaling factors are optimized and performed for different latitudes in ODI. This latitude adaptive mechanism can not only advance the SR performance, but also save the computing resource.

3. Database and analysis

3.1. ODI-SR database

We collected 1,000 high quality ODIs from Huang *et al.* [17] and the Internet for ODI SR. The resolution of
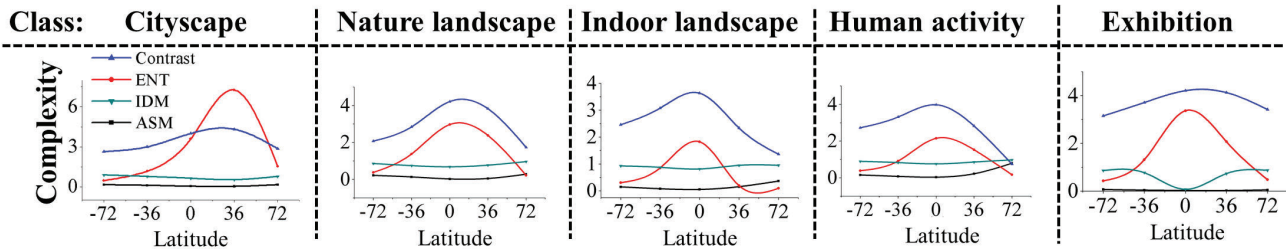
Figure 2. Texture complexity across different latitudes in ODI-SR database. Note that higher ENT and contrast values and lower IDM and ASM values indicate higher texture complexity.
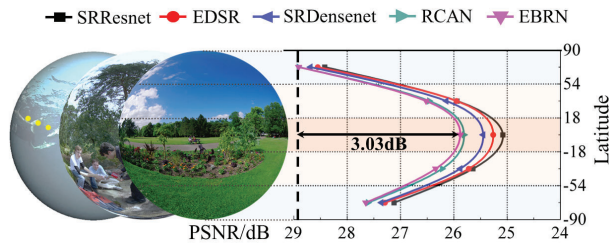

Figure 3. The PSNR results of different latitude bands reconstructed by different 2D SR methods.

these ODIs ranges from 2K, i.e., 1,920×1,080, to 24K, i.e., 24,048×12,024. To enrich diversity, these ODIs are selected to contain different kinds of content, including cityscape, natural scene, indoor scene, human activity and exhibition. From this database, we randomly select 800 images for training, 100 images for validation, and 100 images for testing.

3.2. ODI analysis

Through analyzing the ODI-SR database, we have the following inspiring findings about characteristics of ODIs, which play important roles in designing our LAU-Net.

Finding 1: *Compared to high latitude, low latitude areas tend to have higher texture complexity.*

Following [23, 13], we measure the texture complexity of different latitude areas in ODIs in terms of four common used textual features from [14]. These four features are entropy, contrast, angular second moment (ASM), and inverse differential moment (IDM). Note that the entropy and contrast values are in positive correlation with texture complexity, while ASM and IDM indices are in negative correlation with texture complexity.

Fig. 2 plots the change of the four features across latitudes in different categories in our ODI-SR database. We can see that the lower latitude areas tend to have higher entropy and contrast values, while the higher latitude areas tend to get higher ASM and IDM values. This demonstrates that the texture complexity is highly related with the latitude, and the low latitude areas show higher texture complexity. This completes the analysis of Finding 1.

Finding 2: *Compared to high latitude, it is more difficult to super-resolve low latitude areas using 2D SR methods.*

To verify this finding, we first equally split each ODI into five latitude bands, i.e., each band covers 36° (=180°/5) in latitude. For each band, we down-sample it by 4× and then perform SR using five state-of-the-art SISR methods, including SRResNet[22], EDSR[25], SRDenseNet[38], RCAN[48] and EBRN[30]. Fig 3 presents the PSNR results in different latitude areas with the five SISR methods. As can be seen, there is a big PSNR gap between the low and high latitude bands, i.e., the PSNR of lowest latitude is more than 2 dB lower than that of the highest latitude band. However, when people watch ODIs, the low latitude area usually attracts more attention, which should be reconstructed with higher accuracy. The existing SISR methods fail to achieve high PSNR in low latitude area. The possible reason is that they treat each band equally and assign equivalent computing resource to them. Actually, as analyzed in Finding 1, the low latitude area has higher texture complexity, and thus requires more computing resource to achieve similar reconstruction accuracy as the high latitude.

4. Latitude Adaptive Upscaling Network

In this section, we introduce our LAU-Net in detail. The multi-level architecture of LAU-Net is introduced in Section 4.1, and the structure of the spatial segmentation module and the evaluator is introduced in Section 4.2. Finally, Section 4.3 introduces the training algorithm and loss function.

4.1. Network architecture

Fig. 4 shows the network architecture of the proposed LAU-Net. As can be seen, the LAU-Net has a multi-level pyramid structure, with each level consisting of a channel attention dense subnet (CAD-net) and a spatial segmentation module (SSM). The role of CAD-net is to extract the high-level features from the input LR image, while SSM serves to dynamically drop the unnecessary latitude bands at the current level and send the remained bands to the next level. At the j-th level, the corresponding latitude bands can be upscaled by $2^j \times$. In other words, our network is able to achieve ODI SR at flexible upscaling factors by changing the number of levels. Here, for the sake of brevity, we only show in Fig. 4 the network architecture with 3 levels, i.e., 8× upscaling. Next, we introduce the details of each level.
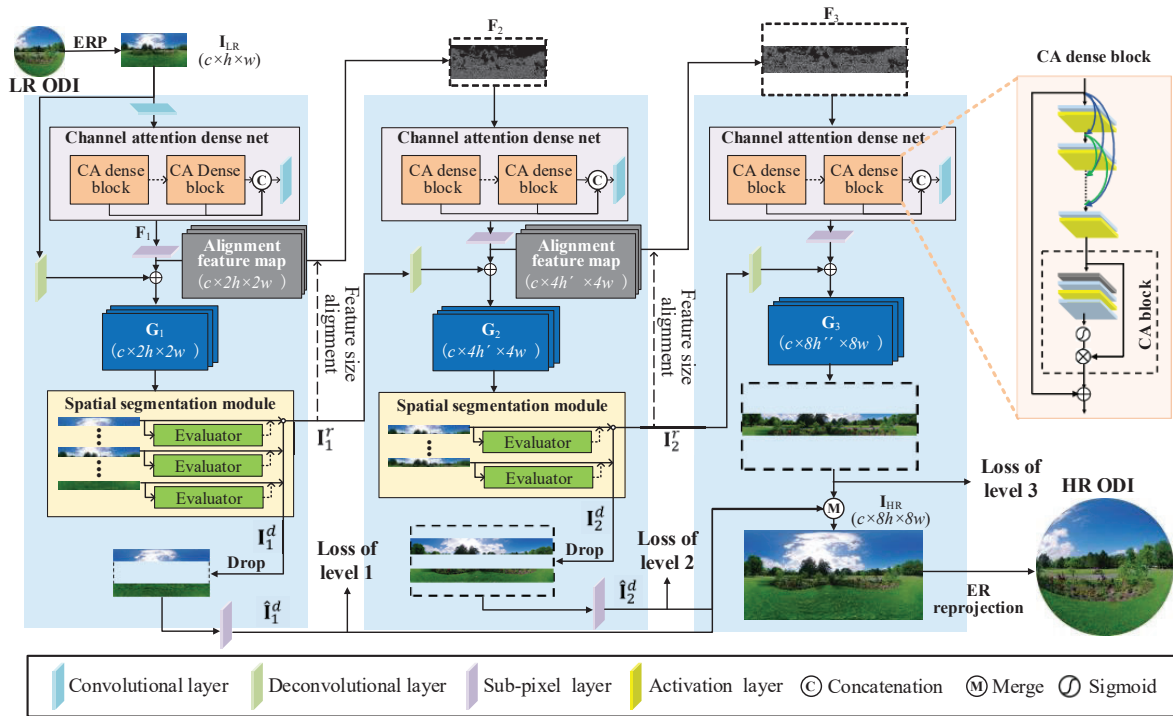
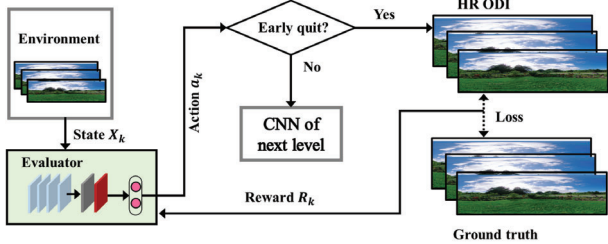Figure 4. The architecture of the proposed LAU-Net.

Figure 5. Framework of the proposed RL model for easy latitude band dropping.

First level: The first level aims to achieve $2\times$ upscaling on the whole LR ODI. The input to the first level is the LR ODI image $\mathbf{I}_{LR} \in \mathbb{R}^{c \times h \times w}$. After a convolutional layer and CAD-net, we can obtain the extracted high-level feature $\mathbf{F}_1$ as follows:

$$\mathbf{F}_1 = f_{\text{CAD}}(Conv(\mathbf{I}_{LR})), \tag{1}$$

where $Conv$ denotes single convolutional layer and f_{CAD} represents the operation of CAD-net. The CAD-net is composed of B channel attention dense blocks with global skip connections across each block, for extracting high level features. Each dense block contains 8 basic convolutional layers with local connections, and a channel attention (CA) block proposed in [48]. Then, $\mathbf{F}_1$ is upsampled using a sub-pixel convolutional layer to generate the super-resolved image $\mathbf{G}_1 \in \mathbb{R}^{c \times 2h \times 2w}$ as follows,

$$\mathbf{G}_1 = f_{\text{REC}}(\mathbf{F}_1) + f_{\text{UP}}(\mathbf{I}_{LR}). \tag{2}$$

Here, f_{REC} indicates the sub-pixel convolutional upsampling and f_{UP} is the deconvolutional upsampling layer. The super-resolved image $\mathbf{G}_1$ is then fed into SSM to drop the unnecessary latitude bands determined by the RL evaluator,

$$[\mathbf{I}_1^d, \mathbf{I}_1^r] = f_{\text{SSM}}(\mathbf{G}_1), \tag{3}$$

where $\mathbf{I}_1^d$ is the dropped latitude band, and $\mathbf{I}_1^r$ is the remained bands which are input to the next level. More details about SSM and RL evaluator are shown in Section 4.2.

Second and third levels: The second level aims to further upscale the remained latitude bands $\mathbf{I}_1^r$ from the first level. As shown in Fig 4, there are two inputs to the second level, i.e., $\mathbf{I}_1^r$ and alignment feature maps $\mathbf{F}_2$ from the first level. The definition of $\mathbf{F}_2$ is as follows,

$$\mathbf{F}_2 = f_{\text{FA}}(f_{\text{REC}}(\mathbf{F}_1)). \tag{4}$$

Here, f_{FA} indicates the feature size alignment operation, in which $\mathbf{F}_1$ is horizontally cropped to $\mathbf{F}_2$ to keep the same latitude range as $\mathbf{I}_1^r$. Then, we can obtain the super-resolved image $\mathbf{G}_2$ as follows,

$$\mathbf{G}_2 = f_{\text{REC}}(f_{\text{CAD}}(\mathbf{F}_2)) + f_{\text{UP}}(\mathbf{I}_1^r). \tag{5}$$

The SSM is also applied in the second level to drop the unnecessary latitude bands, and we can have $[\mathbf{I}_2^d, \mathbf{I}_2^r] = f_{\text{SSM}}(\mathbf{G}_2)$. Here, $\mathbf{I}_2^d$ indicates the dropped latitude at the second level and $\mathbf{I}_2^r$ is the remained latitude bands send to the next level.

For $8\times$ upscaling, the third level is also the final level. The inputs to the third level are $\mathbf{I}_2^r$ and $\mathbf{F}_3 = f_{\text{FA}}(\mathbf{F}_2)$. The super-resolved image $\mathbf{G}_3 = f_{\text{REC}}(f_{\text{CAD}}(\mathbf{F}_3)) + f_{\text{UP}}(\mathbf{I}_2^r)$. In summary, we obtain $\mathbf{I}_1^d$, $\mathbf{I}_2^d$ and $\mathbf{G}_3$ from the first, second and third levels, respectively. However, they are only partial latitude bands of the ODI. To obtain the complete HR ODI, we firstly upsample $\mathbf{I}_1^d$ and $\mathbf{I}_2^d$ to $\hat{\mathbf{I}}_1^d$ and $\hat{\mathbf{I}}_2^d$ by a sub-pixel convolution layer, to make them with the same width resolution as $\mathbf{G}_3$, and then merge them to produce the final reconstructed HR ODI image. To avoid boundary artifacts in merge process, following [49], we reserve overlapping areas and use weighted averaging to generate smooth boundaries.

4.2. Spatial segmentation module

In our method, we have several SSMs corresponding to different levels, and the input to the SSM at the j-th level is $\mathbf{G}_j$. For simplicity, we take the SSM at the first level for example. We firstly segment $\mathbf{G}_1$ into K stripes with the same size along latitude, i.e., $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K\}$. The height of each stripe is calculated by $h_d = \frac{2h}{K}$, where $2h$ is the height of $\mathbf{G}_1$. After segmentation, each stripe is fed into an independent evaluator f_{E_k} to determine whether it should be dropped or remained. The dropped "easy" stripes are forced to early exit, while the remained "hard" patches are fed to the next level. Except for the first SSM, number of evaluators in other SSM is determined by the number of remaining stripes. Noted that in two different SSMs, parameters are shared between evaluators which processing the same stripes.

Evaluator. The evaluator is the key component to achieve early quit strategy. As shown in Fig 5, a evaluator contains 4 convolutional layers, followed by a global pooling layer and a fully-connected layer. Since the process of determining early quit or not at each level is non-differentiable, we formulate it as a Markov Decision Process (MDP) and use reinforcement learning (RL) to train the evaluator. Next, we first describe the state and action, and then introduce the latitude-adaptive reward.

State and Action. For k-th evaluator at j-th level, the state is the input latitude stripe $\mathbf{X}_k^j$. Given the state $\mathbf{X}_k^j$, the evaluator f_{E_k} generates a dispersed distribution of dropping or not, which can be formulated as $f_{E_k}(\mathbf{X}_k^j) = \pi(a \mid \mathbf{X}_k^j), a \in \{0,1\}$. In the training phase, the action is sampled from this probabilistic distribution, denoted by $a_k^j \sim \pi(a \mid \mathbf{X}_k^j)$. In testing phase, the action is determined by the highest probability, i.e., $a_k^j = \arg\min_a \pi(a \mid \mathbf{X}_k^j)$.

Latitude-adaptive Reward. In an RL framework, the evaluator is trained to maximize a accumulated reward, and thus a proper design of reward function is critical. In this paper, to better serve the ODI SR task, we propose latitude-adaptive reward which not only considers the overall SR performance, but also the complexity of different latitudes of ODI. Inspired by [39, 46], the current reward and the accu-

mulated reward of k-th evaluator at j-th level is formulated as follows:

$$r_k^j = \alpha \cdot \mathbf{1}_{\{1\}}(a_k^j), \tag{6}$$

$$R_k^j = \sum_{i=j}^{J-1} \gamma^{i-1} r_k^i - \cos\theta_k \cdot \|\hat{\mathbf{I}}(k) - \mathbf{I}^{gt}(k)\|_2 \tag{7}$$

where α is the reward weight for quitting, which also serves as a trade-off between performance of network and computations. The $\mathbf{1}_{\{1\}}(\cdot)$ represents an indicator function. When the stripe is determined to be dropped, i.e., $a_k^j = 1$, the reward given to evaluators. We denote θ_k as the median latitude of k-th stripe. Similar to WS-PSNR defined in [32], we use θ_k to consider the non-uniform pixel distribution across latitudes. We use $\hat{\mathbf{I}}(k)$ and $\mathbf{I}^{gt}(k)$ to represent the MSE between the final output and groundtruth of k-th stripe. In addition, γ is the discount factor of future reward and J is number of total levels in LAU-Net.

4.3. Training policy

The First Stage. The training process is composed of two stages. In the first stage, we train the multi-level CNN without early exit or evaluator, i.e., RL network is not be involved in the first stage. All ODIs are trained through all the pyramid level. For CNN, given N pairs of training samples, we optimize the weighted ℓ_1 reconstruction loss between predicted HR latitude bands and the corresponding ground-truths across multiple levels. The loss function for the j-th level is defined as follows:

$$\mathcal{L}_j = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{W}_j(\hat{\mathbf{I}}_j^d(i) - \mathbf{I}_j^{gt}(i))\|_1, \tag{8}$$

where $\hat{\mathbf{I}}_j^d$ is the super-resolved output at the j-th level, and $\mathbf{I}_j^{gt}$ is the corresponding ground truth. $\mathbf{W}_j$ is the weight matrix which defines the importance of each pixel in terms of its latitude. In $\mathbf{W}_j$, the elements in the same row have the same value. Suppose that the latitude of the p-th row in $\mathbf{W}_j$ is q, following [32], we can have the values of the p-th row in $\mathbf{W}_j$ as $\cos(\frac{q+0.5-H/2}{H}\pi)$, where H is the height of $\mathbf{I}_j^{gt}$.

It is worth noting that without horizontal copping and merging operation in the first stage, complete ODI is output from each level. In this case, the total loss remains the same as the second stage. We train CNN network for more than 50 epochs for early convergence so that so that CNN and evaluator subnet are better associated and optimized.

The Second Stage. In the second stage, we jointly train the evaluator and the multi-level CNN as shown in Fig 5. In this stage, considering that the low latitude area is more important than the high latitude, we define the overall loss function across all the levels by giving more emphasis on

Table 1. The average and standard deviation of WS-PSNR (dB) and WS-SSIM results of different methods. The red values indicate the best and the blue values indicate the second best results.

Scale	8×				16×			
Method	ODI-SR		SUN 360 Panorama		ODI-SR		SUN 360 Panorama	
	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM
Bicubic	19.64±2.96	0.5908±0.0834	19.72±3.15	0.5403± 0.0862	17.12± 3.06	0.4332±0.0845	17.56± 3.06	0.4638± 0.0848
SRCNN	20.08± 1.65	0.6112±0.0712	19.46± 1.83	0.5701± 0.0819	18.08± 2.03	0.4501± 0.0806	17.95±2.12	0.4684±0.0813
VDSR	20.61± 1.74	0.6195± 0.0796	19.93± 1.91	0.5953± 0.0798	18.24± 2.35	0.4996± 0.0824	18.21± 2.47	0.4867±0.0829
LapSRN	20.72± 1.89	0.6214± 0.0823	20.05± 2.51	0.5998± 0.0816	18.45±2.54	0.5161±0.0861	18.46± 2.53	0.5068± 0.0841
MemNet	21.73±1.84	0.6284± 0.0802	21.08± 2.35	0.6015±0.0875	20.03±2.68	0.5411±0.0822	19.88±2.13	0.5401±0.0830
MSRN	22.29±1.86	0.6315±0.0815	21.34± 2.43	0.6002± 0.0918	20.05± 3.02	0.5416± 0.0968	19.87± 3.27	0.5316± 0.0976
EDSR	23.97± 1.74	0.6417± 0.0724	22.46± 2.32	0.6341± 0.0861	21.12± 2.58	0.5698± 0.0829	21.06± 2.49	0.5645± 0.0864
D-DBPN	24.15± 1.72	0.6573± 0.0758	23.70± 2.25	0.6421± 0.0858	21.25± 2.42	0.5714± 0.0831	21.08± 2.45	0.5646± 0.0918
RCAN	24.26± 1.68	0.6628± 0.0714	23.88± 2.02	0.6542± 0.0824	21.94± 1.75	0.5824± 0.0815	21.74± 2.28	0.5742± 0.0892
EBRN	24.29±1.72	0.6656±0.0698	23.89±2.04	0.6598±0.0832	21.86± 1.68	0.5809± 0.0792	21.78±2.12	0.5794±0.0842
360-SS	21.65± 1.91	0.6417± 0.0865	21.48± 2.56	0.6352±0.0872	19.65± 2.44	0.5431±0.0868	19.62±2.96	0.5308±0.0879
LAU-Net	24.36± 1.73	0.6801± 0.0736	24.02± 2.13	0.6708± 0.0801	22.07± 1.74	0.5901±0.0812	21.82± 2.36	0.5824± 0.0865

HR
WS-PSNR/WS-SSIM

Bicubic
19.65/0.5727

SRCNN
19.26/0.6093

VDSR
20.98/0.6472

LapSRN
21.47/0.6308

MemNet
21.89/0.6432

360-SS
21.86/0.6252

EDSR
22.91/0.6303
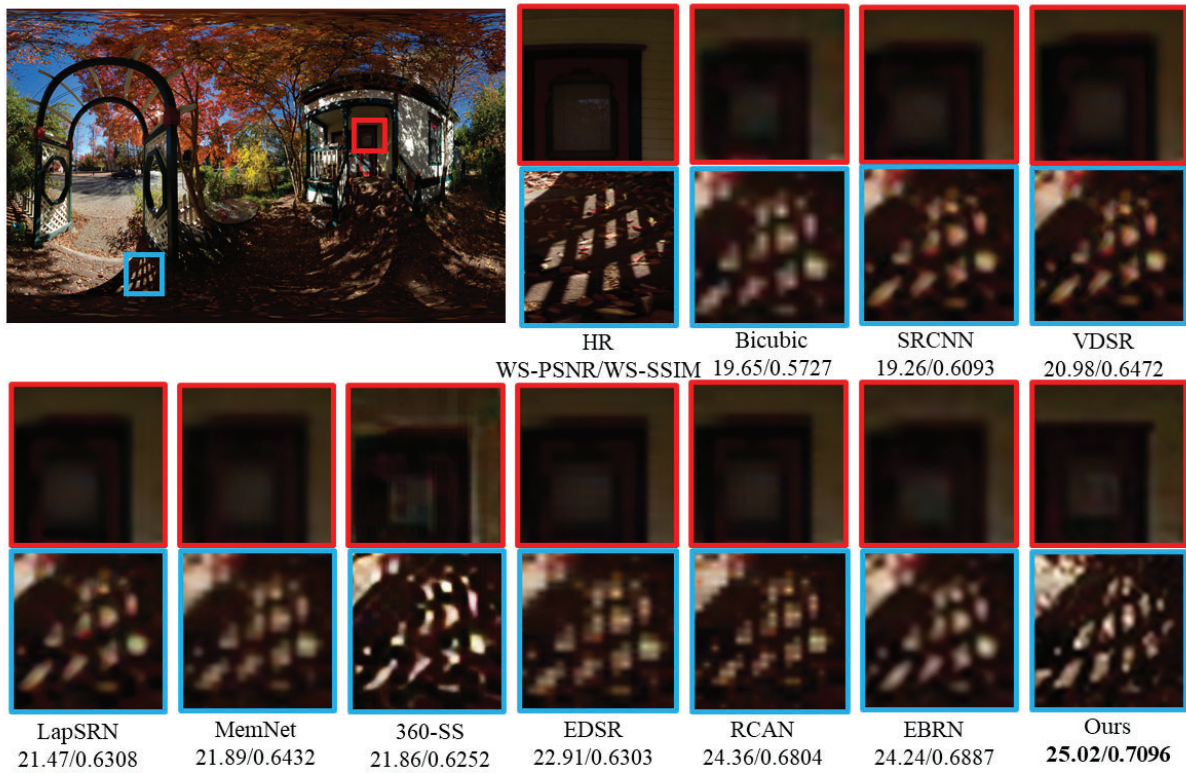
RCAN
24.36/0.6804

EBRN
24.24/0.6887

Ours
25.02/0.7096

Figure 6. Visual comparisons of 8× super-resolved images from the "nature landscape" category of ODI-SR dataset.

the low latitude regions:

$$\mathcal{L}_{\text{total}} = \sum_{j=1}^{J} \mathcal{L}_j \cdot 2^{j-1}. \qquad (9)$$

For evaluators, we update parameters following REIN-FORCE algorithm [41]:

$$\Delta \boldsymbol{\omega}_k = \nabla_{\boldsymbol{\omega}_k} \sum_{j=1}^{J-1} \log \pi(a_k^j \mid \mathbf{X}_k^j; \omega) R_k^j, \qquad (10)$$

$$\boldsymbol{\omega}_k := \boldsymbol{\omega}_k + \beta \Delta \boldsymbol{\omega}_k, \qquad (11)$$

where $\boldsymbol{\omega}_k$ denotes parameters of evaluators, and β denotes the learning rate.

5. Experiment

5.1. Dataset and implementation details

The network is trained using 800 images from our ODI-SR database. For testing, we used 100 images from ODI-SR database which are different from the training images, and 100 ODIs from the SUN 360 Panorama Database [42]. The LR ODIs are generated by bicubic downsampling on the HR ODIs. The number of level is set to 3 for 8× upscaling and

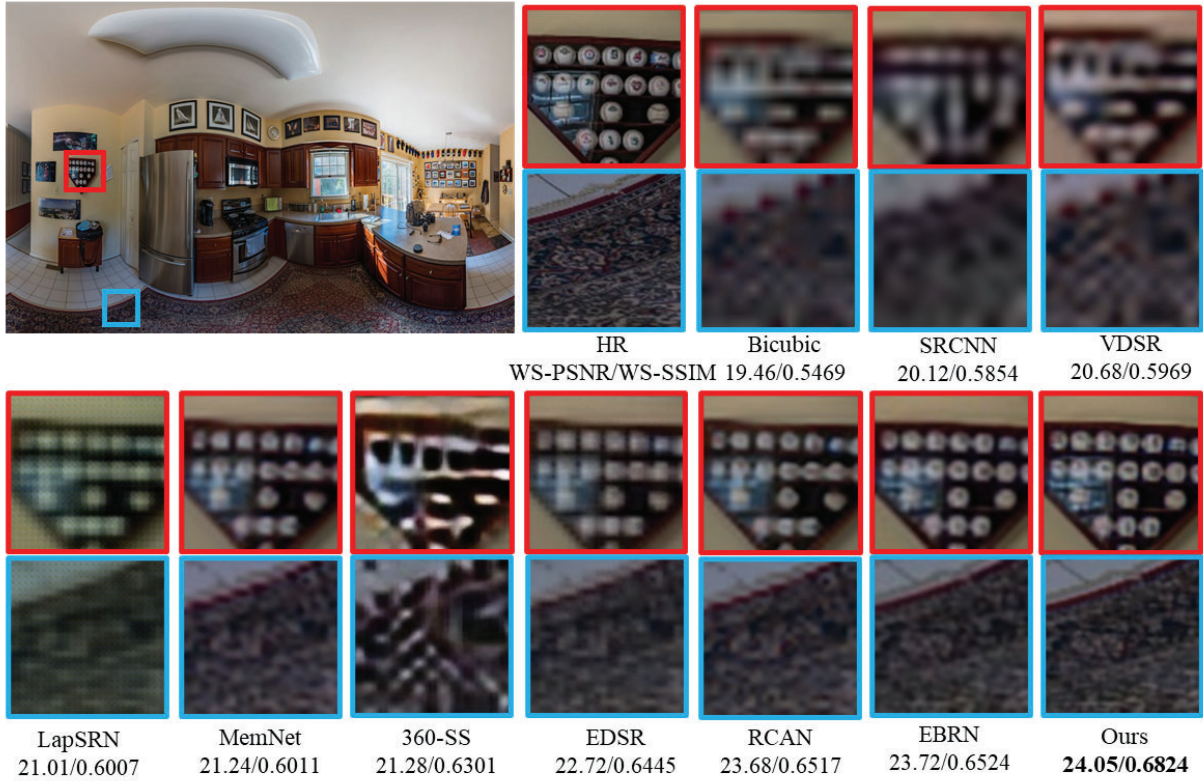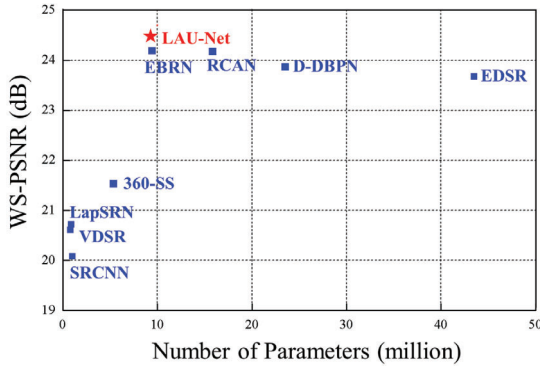Figure 7. Visual comparisons of $8\times$ super-resolved images using different methods on SUN 360 Panorama dataset.

HR
WS-PSNR/WS-SSIM 19.46/0.5469

Bicubic

SRCNN
20.12/0.5854

VDSR
20.68/0.5969

LapSRN
21.01/0.6007

MemNet
21.24/0.6011

360-SS
21.28/0.6301

EDSR
22.72/0.6445

RCAN
23.68/0.6517

EBRN
23.72/0.6524

Ours
24.05/0.6824

Figure 8. WS-PSNR vs. the number of parameters. The comparison is conducted on ODI-SR test set with the $8\times$ up-scaling factor.

Table 2. Computational complexity of different models.

Method	FLOPs	Network params	Running time
LapSRN	23G	1.3M	0.049s
EDSR	2473.4G	45.5M	2.231s
D-DBPN	766.4G	23.2M	0.682s
RCAN	617.9G	16M	0.416s
EBRN	595.5G	9.5M	0.403s
360-SS	15G	1.6M	0.010s
LAU-Net	342.8G	9.4M	0.352s

Table 3. Influence of CA dense block number on ODI-SR.

Number	1	2	3	4	5
WS-PSNR	23.85dB	24.08dB	24.19dB	24.36dB	24.37dB
WS-SSIM	0.6588	0.6656	0.6751	0.6801	0.6803

4 for $16\times$ upscaling. The number of dense blocks B is set to 4 in each CAD-net. The number of patches K is set to 12, which means latitude range is $15°$ for each stripe. To avoid boundary artifacts, an extra $\pm1.5°$ is added for each stripe. Since our network is latitude aware, the training patch should cover all latitudes. Thus, the training patch size is set to 128×32 for $4\times$ upscaling, 64×16 for $16\times$ upscaling and the batch size is 4. The model weights are initialized using the method in [16]. The Adam optimizer [20] is employed with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The learning rate is initially set to 10^{-4} and decreased by a factor of 10 every 100 epochs. Data augmentation techniques are utilized to enlarge the training data.

5.2. Comparison with SOTAs

To validate the effectiveness of the proposed LAU-Net, we compare it with 9 SISR methods for 2D images, including SRCNN [9], VDSR [18], LapSRN [21], MemNet [35], MSRN [24], EDSR [25], D-DBPN [15], RCAN [48], and EBRN [30], and one SISR method for ODI, i.e., 360-SS [28], which is the only method we can find for ODI SISR. For fair comparison, we retrain all the methods using the ODI-SR database. The weighted-to-spherically-uniform PSNR (WS-

Table 4. Influence of number of evaluators on ODI-SR.

K	4	8	12
WS-PSNR	22.64	24.15	24.36
WS-SSIM	0.6521	0.6710	0.6801

Table 5. Influence of loss function on ODI-SR.

Loss function	WS-PSNR	WS-SSIM
L_1 loss	24.31	0.6765
L_2 loss	24.28	0.6712
Ours	24.36	0.6801

PSNR) [32] and weighted-to-spherically-uniform SSIM (WS-SSIM), which are particularly designed for ODI quality measurement, are used as metrics to evaluate the performance of different methods in the experiments.

Quantitative results. Considering that the resolution required for ODIs in real-world conditions is much higher than that of 2D images, Table 1 presents the average and standard deviation of WS-PSNR and WS-SSIM results of different methods for $8\times$ and $16\times$ upscaling on ODI-SR and SUN 360 Panorama datasets. As we can see, our LAU-Net performs better than all other methods in terms of all metrics in both our ODI-SR database and SUN 360 Panorama database. In addition, our results also have a relatively low standard deviation, which means that the reconstruction performance of our network is more stable and generalized. Note that our results are obtained without any self-ensemble strategy.

Qualitative results. Fig 6 and 7 visualize the super-resolved images on ODI-SR and SUN 360 Panorama datasets using different methods for $8\times$ upscaling. As can be seen, our method is able to reconstruct clear textures and accurate structures at both high and low latitude areas. Other SR methods either leads to blurred edges or distorted structures.

Computational complexity. Computational complexity is important for ODI-SR in real applications. Fig 8 draws the number of parameters and the WS-PSNR results of different methods. As can be seen, our LAU-Net achieves higher WS-PSNR results than other methods, with fewer parameters than D-DBPN [15], RCAN [48], and EDSR [25]. This demonstrates that our LAU-Net can well balance the number of parameters and the reconstruction performance, owing to its well-designed architecture. We also present in Table 2 the FLOPs, number of network parameters and the average running time of different methods. As can be seen, the running time of our method is faster than the others. This is because the high latitude and low image complexity patches are early dropped in our network, which greatly reduces the running time without affecting too much objective results.

5.3. Ablation study

CA dense block. Firstly, we investigate the influence of the number of CA dense blocks on the SR performance. Table 3 shows the WS-PSNR results of our LAU-Net with different number of CA dense blocks. As can be seen, the WS-PSNR value improves with the increasing number of CA blocks. However, the increment becomes very small when the number is larger than 4. Thus, we choose to use 4 CA dense blocks in the CAD-net.

Evaluators. The number of evaluators K is an important

factor in our network. The more evaluators indicate more horizontal cropping operations and finer segmentation in SSM. Table 4 show the WS-PSNR and WS-SSIM results with the number of evaluators K ranging from 4 to 16. As we can see from this table, both WS-PSNR and WS-SSIM values increase with K is increased from 4 to 12. The smaller number of evaluators may not result in good results because a large band of the ODIs are forced to exit prematurely. When the number of evaluator is set to 12, our network has sufficient sampling density and is able to reasonably distinguish high latitude patches and low latitude patches.

Latitude-adaptive loss. In Eq. (8), we design a latitude-weighted loss function for each level in LAU-Net using a weight matrix W_j. In Eq. (9), we further give priority to the low-latitude area by multiplying 2^{j-1} to the weight of the j-th level. To investigate the effectiveness of latitude-weighted loss, we remove the W_j in Eq. (8), and directly use the conventional L_1 and L_2 loss to train the network. In Eq. (9), the total loss is simply the sum of the loss of each level without 2^{j-1}. Table 5 compare the WS-PSNR and WS-SSIM results with L_1, L_2 and our latitude-adaptive loss. As can be seen, our latitude-adaptive loss achieves the best performance, indicating its effectiveness.

6. Conclusion

In this paper, we propose a novel latitude adaptive upscaling network called LAU-Net for ODI SR. We first establish a large ODI database with diverse resolutions and image content. Based on our finding that the low latitude bands have higher texture complexity than the high latitude bands, we design a progressive pyramid network architecture in LAU-Net. The core component in LAU-Net is the spatial segmentation module, in which the ODI is split into different latitude bands, and several reinforcement learning based evaluators decide the optimal upscaling factor of the band. The consequence is that the high latitude bands quit the network from shallower levels, while the low latitude bands go deeper. Extensive quantitative and qualitative results on different ODI datasets demonstrate the superiority of the proposed method over the other state-of-the-art SR methods.

References

[1] Saeed Anwar, Salman Khan, and Nick Barnes. A deep journey into super-resolution: A survey. *ACM Comput. Surv.*, 53(3), 2020. 1

[2] Zafer Arican and Pascal Frossard. Joint registration and super-resolution with omnidirectional images. *IEEE TIP*, 20(11):3151–3162, 2011. 1, 2

[3] Luigi Bagnato, Yannick Boursier, Pascal Frossard, and Pierre Vandergheynst. Plenoptic based super-resolution for omnidirectional image sequences. In *ICIP*, pages 2829–2832. IEEE, 2010. 1

[4] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *European Conference on Computer Vision Workshops*, pages 334–355. Springer, 2019. 1

[5] J. Choi and M. Kim. A deep convolutional neural network with selection units for super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1150–1156, 2017. 2

[6] X. Deng. Enhancing image quality via style transfer for single image super-resolution. *IEEE Signal Processing Letters*, 25(4):571–575, 2018. 2

[7] X. Deng, R. Yang, M. Xu, and P. L. Dragotti. Wavelet domain style transfer for an effective perception-distortion tradeoff in single image super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3076–3085, 2019. 2

[8] X. Deng, Y. Zhang, M. Xu, S. Gu, and Y. Duan. Deep coupled feedback network for joint exposure fusion and image super-resolution. *IEEE Transactions on Image Processing*, 30:3098–3112, 2021. 2

[9] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016. 2, 7

[10] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*, pages 391–407. Springer, 2016. 2

[11] M. S. Elbamby, C. Perfecto, M. Bennis, and K. Doppler. Toward low-latency and ultra-reliable virtual reality. *IEEE Network*, 32(2):78–84, 2018. 1

[12] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *In IEEE International Conference on Computer Vision*, pages 349–356, 2009. 1

[13] W. Gomez, W. C. A. Pereira, and A. F. C. Infantosi. Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. *IEEE Transactions on Medical Imaging*, 31(10):1889–1899, 2012. 3

[14] Robert Haralick, K. Shanmugam, and Ih Dinstein. Textural features for image classification. *IEEE Trans Syst Man Cybern*, SMC-3:610–621, 1973. 3

[15] M. Haris, G. Shakhnarovich, and N. Ukita. Deep back-projection networks for super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1664–1673, 2018. 2, 7, 8

[16] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *In IEEE International Conference on Computer Vision*, pages 1026–1034, 2015. 7

[17] M. Huang, Q. Shen, Z. Ma, A. C. Bovik, P. Gupta, R. Zhou, and X. Cao. Modeling the perceptual quality of immersive images rendered on head mounted displays: Resolution and compression. *IEEE TIP*, 27(12):6039–6050, 2018. 2

[18] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 2, 7

[19] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1645, 2016. 2

[20] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *ICLR*, 2014. 7

[21] W. Lai, J. Huang, N. Ahuja, and M. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5835–5843, 2017. 7

[22] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 105–114, 2017. 2, 3

[23] S. Leigh, Z. Wang, and D. A. Clausi. Automated ice–water classification using dual polarization sar satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5529–5539, 2014. 3

[24] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *European Conference on Computer Vision*, pages 527–542. Springer, 2018. 7

[25] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1132–1140, 2017. 2, 3, 7, 8

[26] Roey Mechrez, Itamar Talmi, Firas Shama, and Lihi Zelnik-Manor. Maintaining natural image statistics with the contextual loss. In *ACCV*, pages 427–443, 2019. 2

[27] Hajime Nagahara, Yasushi Yagi, and Masahiko Yachida. Super-resolution from an omnidirectional image sequence. In *2000 26th Annual Conference of the IEEE Industrial Electronics Society.*, pages 2559–2564. IEEE, 2000. 1, 2

[28] Cagri Ozcinar, Aakanksha Rana, and Aljosa Smolic. Super-resolution of omnidirectional images using adversarial learning. In *MMSP*, pages 1–6. IEEE, 2019. 1, 2, 7

[29] E. Perez-Pellitero, J. Salvador, J. Ruiz-Hidalgo, and B. Rosenhahn. Psyco: Manifold span reduction for super resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1837–1845, 2016. 2

[30] Y. Qiu, R. Wang, D. Tao, and J. Cheng. Embedded block residual network: A recursive restoration model for single-image super-resolution. In *In IEEE International Conference on Computer Vision*, pages 4179–4188, 2019. 3, 7

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

[32] Y. Sun, A. Lu, and L. Yu. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE Signal Processing Letters*, 24(9):1408–1412, 2017. 5, 8

[33] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003. 1

[34] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2798, 2017. 2

[35] Y. Tai, J. Yang, X. Liu, and C. Xu. Memnet: A persistent memory network for image restoration. In *IEEE International Conference on Computer Vision*, pages 4549–4557, 2017. 2, 7

[36] Radu Timofte, Shuhang Gu, Luc Van Gool, Lei Zhang, and Ming-Hsuan Yang. Ntire 2018 challenge on single image super-resolution: Methods and results. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 965–96511, 2018. 1

[37] R. Timofte, R. Rothe, and L. V. Gool. Seven ways to improve example-based single image super resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1865–1873, 2016. 2

[38] T. Tong, G. Li, X. Liu, and Q. Gao. Image super-resolution using dense skip connections. In *In IEEE International Conference on Computer Vision*, pages 4809–4817, 2017. 2, 3

[39] Xin Wang, Fisher Yu, Zi-Yi Dou, and Joseph Gonzalez. Skipnet: Learning dynamic routing in convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 5

[40] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *European Conference on Computer Vision Workshops*, pages 63–79, 2019. 2

[41] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4):229–256, 1992. 6

[42] J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba. Recognizing scene viewpoint using panoramic place representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2695–2702, 2012. 6

[43] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1059–1066, 2013. 2

[44] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *IEEE TIP*, 21(8):3467–3478, 2012. 2

[45] W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue, and Q. Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12):3106–3121, 2019. 1

[46] K. Yu, C. Dong, L. Lin, and C. C. Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2443–2452, 2018. 5

[47] W. Zhang, Y. Liu, C. Dong, and Y. Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *In IEEE International Conference on Computer Vision*, pages 3096–3105, 2019. 2

[48] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*, pages 294–310, 2018. 2, 3, 4, 7, 8

[49] Z. Zhang and V. Sze. Fast: A framework to accelerate super-resolution processing on compressed videos. In *IEEE Conference on Computer Vision and Pattern RecognitionW*, pages 1015–1024, jul 2017. 5