Super-resolution of Omnidirectional Images Using Adversarial Learning

Cagri Ozcinar*, Aakanksha Rana*, and Aljosa Smolic

V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Ireland.

{ozcinarc, ranaa, smolica}@scss.tcd.ie

Abstract—An omnidirectional image (ODI) enables viewers to look in every direction from a fixed point through a head-mounted display providing an immersive experience compared to that of a standard image. Designing immersive virtual reality systems with ODIs is challenging as they require high resolution content. In this paper, we study super-resolution for ODIs and propose an improved generative adversarial network based model which is optimized to handle the artifacts obtained in the spherical observational space. Specifically, we propose to use a fast PatchGAN discriminator, as it needs fewer parameters and improves the super-resolution at a fine scale. We also explore the generative models with adversarial learning by introducing a spherical-content specific loss function, called 360-SS. To train and test the performance of our proposed model we prepare a dataset of 4500 ODIs. Our results demonstrate the efficacy of the proposed method and identify new challenges in ODI super-resolution for future investigations.

Index Terms—omnidirectional image, virtual reality, super-resolution, generative adversarial network, spherical-content loss

I. INTRODUCTION

With recent advances in virtual reality (VR), the omnidirectional image (ODI) represents an increasingly important imaging format for immersive technologies. This format provides realistic VR experiences in which viewers have a sense of being present where the content was captured. Existing VR displays, such as a head-mounted displays (HMD), navigate a given ODI input with three degrees of freedom. ODIs are captured with 360° camera systems and stored in 2D planar representations (*e.g.*, equirectangular projection (ERP) or cubemap projection) [1] to be compatible with existing image processing pipelines. For display, the ODI is projected onto a sphere and then rendered through a VR display. When using this emerging technology, viewers receive a more immersive experience compared to watching a standard 2D image. ODIs are used in a variety of applications such as entertainment [2], advertising, communication [3], health-care, and education.

However, this image representation introduces significant technical challenges because it requires very high resolution to cover the entire 360° viewing space. Existing HMDs can use only part of a given ODI, called the viewport [4], so even higher resolutions are required to provide an acceptable level of quality of experience (QoE) [5] in VR. For instance, according to recent studies [6], the resolution of the captured

ODI should be 21600×10800 (*i.e.*, 60 pixels per degree) to provide a sufficiently high-quality ODI for existing HMD devices. At this level, neither current consumer capturing systems nor network bandwidths are capable of processing and transmitting such a large amount of data.

High-resolution content using a single ODI at a lower resolution can be reconstructed with super-resolution (SR) techniques. Early studies such as [7, 8] have shown that high-resolution ODIs with high-quality could be generated using handcrafted image features. A high-resolution ODI was constructed by interpolating the missing information between low-resolution ODI pixels.

In today's era of deep learning, the core SR algorithms for traditional images have shifted from interpolation to learning[9] for constructing the high-resolution image from a given low-resolution image using predictive or generative models to find better optimal solutions [10, 11]. Convolution neural networks (CNNs) have shown great successes in the SR field [12]. In particular, for standard images, high-quality reconstruction performance can be achieved by using recent high-capacity deep learning algorithms such as the generative adversarial networks (GANs) [13–16].

In this paper, we address the problem of SR with deep learning for ODIs by introducing an efficient CNN-based solution for reconstructing ODIs with high resolution from given their low-resolution ODIs. In particular, we design a GAN model that inputs a low resolution ODI and outputs a corresponding high-resolution image ($2\times$, $4\times$, $8\times$) of higher quality compared to what existing algorithms can reconstruct. Following a generative adversarial network paradigm, our model consists of a generator (G_{sr}) and a discriminator (D_{sr}), which compete with each other. The generator trying to fool the discriminator by producing high-resolution output ODIs for the given low resolution ODIs. The discriminator tries to distinguish between real and synthetically generated ODI pairs. To enhance the performance of this adversarial learning , we also introduce an ODI-structure-preserving loss function, called 360-SS loss. Our approach handles typical artifacts appearing at the polar regions of the conventional ERP format ODIs, and thus preserving more details of the spherical content in the polar and equatorial regions.

Our contributions in this paper are two fold. Firstly, we explore the super-resolution of ODIs problem using adversarial learning paradigm for the first time, and propose the use of a PatchGAN [17] discriminator to constrain the generator

for predicting high-quality output effectively. Secondly, we introduce the 360-SS loss function to precisely measure the objective quality of the ODI in spherical observation space. In this loss function, the error of each pixel on the projection plane is multiplied by a weight to account for the influence of the spherical mapping on the loss estimation. The code for the proposed method is provided with this paper[1]. We hope that our work can inspire further research within the context of deep learning-based SR to reconstruct high-resolution ODIs.

The rest of this paper is organized as follows. Section II presents the related work on super-resolution, focusing on early work for ODIs and recent deep learning-based approaches. The proposed model and experimental results are presented in Sections III and IV followed by our conclusions in Section V.

II. RELATED WORK

In previous two decades, many super-resolution methods have been proposed, and we briefly describe them in this section. For more details and examples from this field, comprehensive literature reviews are available in [18, 19].

Early work, such as [7, 8], proposed innovative super-resolution techniques that utilized handcrafted methods or simplistic learning based techniques [20] for ODIs. In [7], for instance, the authors first performed a registration between successive frames of omnidirectional video using the plenoptic geometry of the 360-degree scene. All the visual information was then used to generate a high-resolution ODI. Later, Arican *et al.* [8] showed how multiple ODIs with arbitrary rotations could contribute to the super-resolution problem by leveraging the Spherical Fourier Transform (SFT). In their work, the joint registration and super-resolution problem was solved based on the total least squares norm minimization in the SFT domain.

Recent image super-resolution efforts benefited from advances in deep learning [21] and end-to-end architectures. For example, Zhang *et al.* [13] introduced a dimensional stretching strategy for a single deep-learning network to handle multiple quality degradations (*i.e.*, blur kernel and noise level) for super-resolution. Furthermore, Ledig *et al.* [14] highlighted the limitations of PSNR-based super-resolution solutions and introduced a generative network for producing super-resolutions, called SRGAN. In their work, a perceptual loss was introduced to optimize the super-resolution model in a future space instead of pixel space. Later, to recover more realistic textural details, Wang *et al.* [15] improved the discriminator of the SRGAN using a realistic average GAN [16]. Additionally, they enhanced the perceptual loss by incorporating features before activation.

However, super-resolution for ODI has not been investigated sufficiently. To the best of our knowledge, no research work on the super-resolution of ODIs using recent advances in deep learning exist.

---

[1] https://github.com/V-Sense/360SR

III. PROPOSED MODEL

Our objective is to generate a high-resolution ODI, I_{360-sr}, from a low-resolution input ODI, I_{360-lr}. To this end, we first prepared the ODI pairs, the output I_{360-sr} and input I_{360-lr}, to be used for training. We obtained the low resolution version following the standard methodology mentioned in [14], where a Gaussian filter followed by a down-sampling operator is applied to obtain the low-resolution ODIs. With these training image-pairs, we then propose to learn the generator **G** to generate artifact-free and a better quality high-resolution output images by optimizing the whole network with an adequate loss functions. In this paper, we employ the most commonly used ERP ODI representation.

In the following, we explain the proposed network architecture including the generator (**G**) and discriminator (**D**) and the proposed loss function (360-SS) to preserve the spherical content specific details.

A. Generator-Discriminator Architecture

Motivation: The fundamental idea behind a GAN framework is that both the generator and discriminator try to compete with each other. On the one hand, the generator tries to fool discriminator by producing a high resolution, real looking ODI by taking a low-resolution ODI input. While on the other hand, the discriminator tries to discriminate between real and synthetically generated (by the generator) high-resolution ODI image pairs. In this process by spanning the sub-space of the natural images, the generator learns to generate images that are much closer to their corresponding ground truth. This, on the contrary, is quite difficult to achieve with a simple CNN network using a Euclidean loss [17]. In this paper, inspired from several state-of-art approaches [14, 15], we explore the adversarial learning framework for our task at hand.

GAN: In order to design a model for super-resolving ODIs, we adopt the generator architecture from [14]. The layout of the generator architecture is shown in Figure 1. The network consists of a convolutional layer followed by five residual blocks which are capable of capturing the fine details with high-frequency, bypassing the need of a deeper convolutional layer network [14]. The network consists explicitly of two convolution layers and a final sub-pixel convolution layer which is designed to super-resolute the input image by the given factor.

Next, we propose to use the PatchGAN discriminator to discriminate between the real and generated ODI. The layout of the discriminator architecture is shown in Figure 2. The PatchGAN discriminator is applied on a 70×70 (overlapping) patch-size of its input, and the averaged scores are then used to finalize the decision whether the image is real or fake. In this process, the network learns to focus at finer patch level details. An added advantage of a PatchGAN is the smaller number of the parameters as compared to the discriminator adopted in [14], which makes the computation faster.

Objective Function: In this work, our objective is to construct high-resolution ODI with high-quality from a low reso-
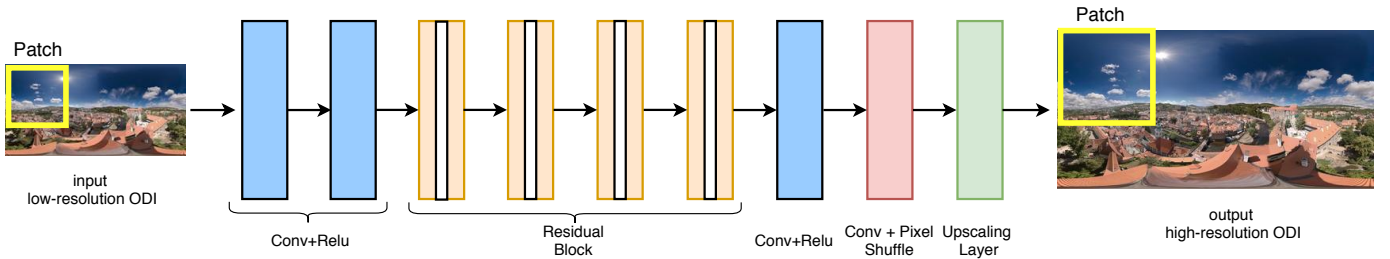
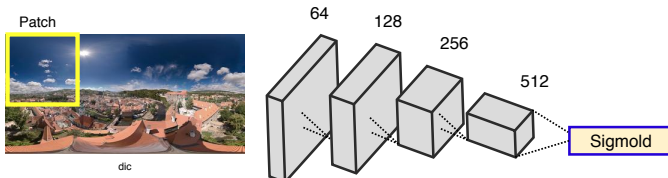Fig. 1: Architecture of Generator Network with corresponding block labels.

Fig. 2: Architecture of the used Discriminator Network.

lution ODI using super-resolution approach. For this aim, we formulate our objective as follows:

$$
\mathbf{G}^* = \arg\min_G \max_D [\mathcal{L}_{adv}(\mathbf{G}, \mathbf{D})] + \\ \beta \mathcal{L}_{feat}(\mathbf{G}) + \gamma \mathcal{L}_{360-SS}(\mathbf{G}) \quad (1)
$$

where we aim to optimize the objective over three loss functions which are used in the adversarial learning of the generator $\mathbf{G}$. In the next section we describe the 3 loss terms: adversarial loss $\mathcal{L}_{adv}$, feature (content) loss $\mathcal{L}_{feat}$ and spherical loss $\mathcal{L}_{360-SS}$. Initially, we detail the proposed loss function (360-SS) to preserve the spherical content specific details. Then, we discuss the feature and adversarial loss terms.

B. Loss Functions

a) 360-SS Loss: To account for the distortion of the spherical surface, we apply the weighted-to-spherically-uniform structural similarity (WS-SSIM) [22] quality score. This measurement considers mapping of spherical content onto the planar surface of a given ODI by adding the appropriate weight to the SSIM value [23]. As a traditional l_2 -based loss function is inferior to SSIM in many imaging problems, and the traditional SSIM [23] is not suitable for evaluating ODIs, we use a novel WS-SSIM-based loss function, called l_{360-SS}. In this loss function, the spherical surface is considered using a non-linear weighting in the SSIM calculation. SSIM relies on computation of luminance, contrast and structural similarities between distorted and original images and is more consistent with subjective quality evaluation than the traditional l_2 based loss function [14]. Similar as [24], the non-linear weights are calculated using the stretching ratio of the area that is projected from the planar surface to the spherical surface. Thus, the loss of 360-SS can be formulated as follows:

$$
\mathcal{L}_{360-SS} = \frac{1}{K} \sum_{i=1}^{K} d_{360-SS}^i, \quad (2)
$$

where K represents the number of samples, and

$$
d_{360-SS} = \frac{\sum_{x=1}^{W/r} \sum_{y=1}^{H/r} \left(SSIM\left(I_{360-sr}^{x,y}, \widehat{I}_{360-sr}^{x,y}\right) q_r^{x,y}\right)}{\displaystyle\sum_{x=1}^{W/r} \sum_{y=1}^{H/r} q_r^{x,y}}, \quad (3)
$$

where $W \times H$ is the resolution of the reconstructed version of the ERP ODI. Note that x and y denote the pixel coordinates of the ERP image, $\widehat{I}_{360-sr}$ and I_{360-sr} stand for the generated high-resolution and original high-resolution versions of the ODI, and $q_r^{x,y}$ represents the weighting intensity in (x, y) of the weight distribution of the ERP which can be calculated according to [24] with:

$$
q_r^{x,y} = cos\frac{(y + 0.5 - (H/2r)\pi}{(H/r)}. \quad (4)
$$

b) Feature Loss: To penalize the $\mathbf{G}$ for the distortions in the content of ODIs, we incorporate the VGG feature-based loss [25] function. This loss is computed as the distance between the feature maps $\mathcal{F}$ obtained by passing the ground truth ODI (I_{360-sr}) and the ODI generated by super-resolution ($\widehat{I}_{360-sr}$) through a pre-trained VGG-19 [25] network. The feature loss $\mathcal{L}_{feat}$ is given as

$$
\mathcal{L}_{feat} = \frac{1}{K} \sum_{i=1}^{K} (\mathcal{F}^i(I_{360-sr}) - \mathcal{F}^i(\widehat{I}_{360-sr})), \quad (5)
$$

c) Adversarial Loss: We apply adversarial loss [26] to aide the generator to effectively generate high-resolution ODIs from the manifold of natural image sets while fooling the discriminator $\mathbf{D}$ network. The adversarial loss term is computed over all the training samples and is given as:

$$
\mathcal{L}_{adv} = \sum_{i=1}^{K} (-log\mathbf{D}(\mathbf{G}(I_{360-lr}^i))). \quad (6)
$$

The adversarial loss is taken as the negative logarithm of the probability of the super-resoluted image, similar to [14].

IV. RESULTS & DISCUSSIONS

Next, we describe the used dataset, metrics, implementation details, and results obtained by our proposed model.
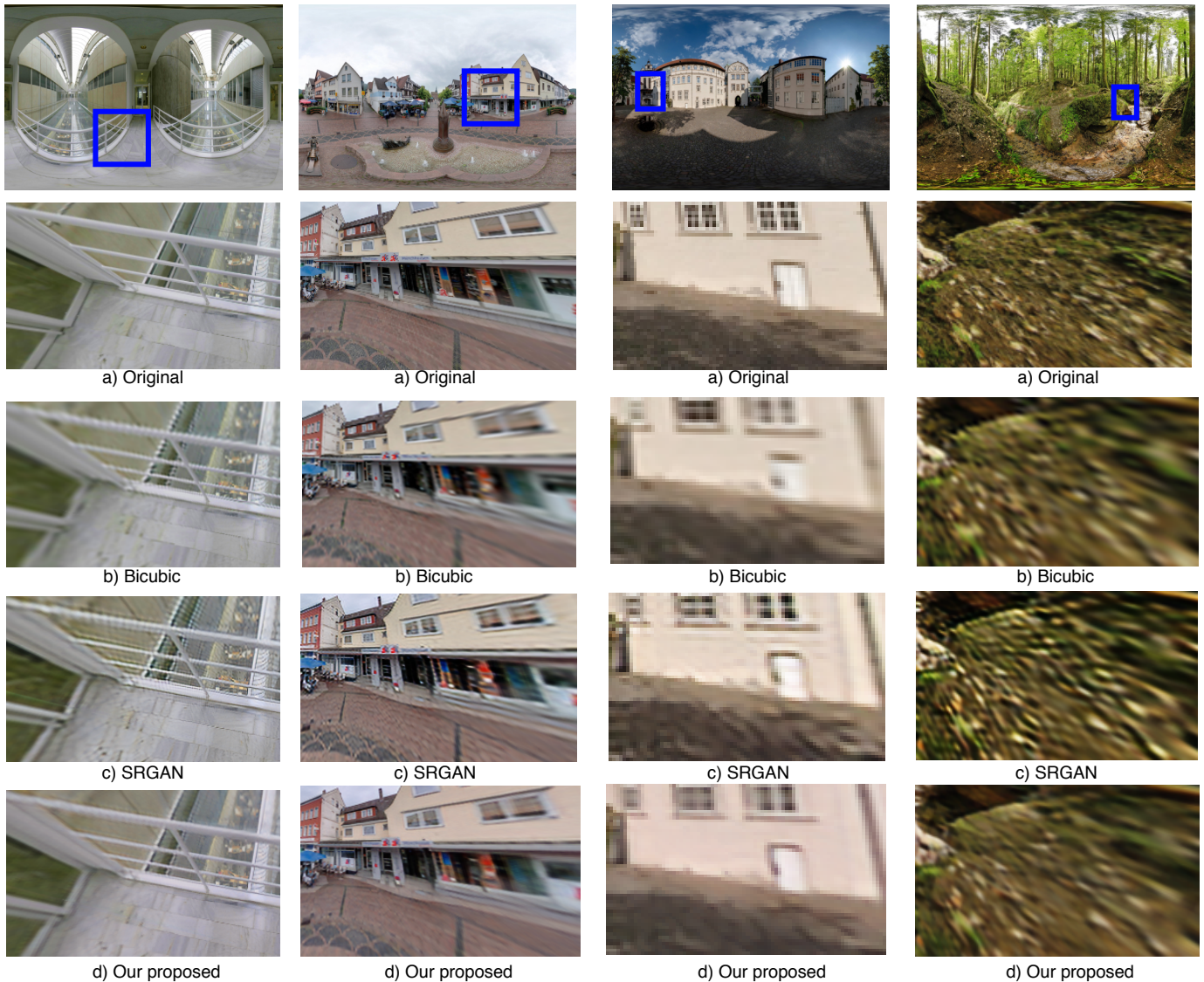
Fig. 3: Qualitative reconstruction results obtained with different benchmark methods and corresponding high-resolution original ODI. We applied rectilinear projection to visualize the most significant differences.

A. Dataset and Metrics

Dataset: We randomly selected a total of 4500 ODIs from the SUN 360 Panorama Database [27]. This database is the only publicly available ODI dataset that has many high-resolution panorama images with a wide variety of content, including indoor, outdoor, structures, human faces. Each selected ODI is in ERP format, and the resolution size is 1440×960. More specifically, we use the ERP format for training and evaluations. 3500 ODIs were randomly selected from the total set for training purposes, and 500 ODIs were used for validation and test sets.

Metrics: We evaluated the performance of our proposed model with four metrics: SSIM [23], PSNR, WS-SSIM [22], and WS-PSNR [28]. SSIM and PNSR are the well-known full-reference planar metrics that have been used in the literature to measure the reconstructed image quality using the original

image. To account for the visual distortion with stretching factors on planar representation, we used well-known ODI metrics, S-SSIM and WS-PSNR, to evaluate the performance of our proposed method. Both ODI quality metrics are the extended versions of their traditional metrics by considering the nonlinear relationship between the projection plane and the sphere surface. Subjective experiments in [5, 29–31] show that ODI metrics have high correlation with human votes compared to the standard planar metrics applied to the ERP formats.

TABLE I: Quantitative Results for Super-Resolution by a factor of 2 on 500 ODIs.

Method	$r = 2$			
	SSIM	PSNR	WS-SSIM	WS-PSNR
NN	0.92 ± 0.06	29.38 ± 0.04	$0.86 \pm .03$	$34.34 \pm .05$
Bicubic	0.93 ± 0.05	30.64 ± 0.06	$0.88 \pm .04$	$35.54 \pm .07$
SRGAN [14]	0.94 ± 0.05	32.56 ± 0.06	$0.90 \pm .06$	$36.35 \pm .06$
Ours	0.95 ± 0.04	33.20 ± 0.04	$0.92 \pm .04$	$37.68 \pm .05$
Ours+ 360-SS loss	0.95 ± 0.03	33.56 ± 0.04	$0.93 \pm .06$	$37.96 \pm .03$

TABLE II: Quantitative Results for Super-Resolution by a factor of 4 on 500 ODIs.

Method	$r = 4$			
	SSIM	PSNR	WS-SSIM	WS-PSNR
NN	0.83 ± 0.07	25.77 ± 0.05	0.71± .03	32.44 ± .05
Bicubic	0.85 ± 0.07	26.71 ± 0.03	0.74 ± .04	32.76 ± .04
SRGAN [14]	0.86 ± 0.02	27.11 ± 0.09	0.75 ± .06	34.76 ± .03
Ours	0.87 ± 0.05	27.19 ± 0.08	0.76 ± .05	35.89 ± .05
Ours+ 360-SS loss	0.87 ± 0.04	27.70 ± 0.03	0.77 ± .08	36.98 ± .06

TABLE III: Quantitative Results for Super-Resolution by a factor of 8 on 500 ODIs.

Method	$r = 8$			
	SSIM	PSNR	WS-SSIM	WS-PSNR
NN	0.83 ± 0.07	23.47 ± 0.05	0.64 ± .06	31.12 ± .04
Bicubic	0.85 ± 0.07	24.26 ± 0.03	0.66 ± .07	31.83 ± .06
SRGAN [14]	0.86 ± 0.02	25.10 ± 0.09	0.70 ± .06	33.00 ± .07
Ours	0.87 ± 0.05	26.24 ± 0.08	0.73 ± .07	34.68 ± .04
Ours + 360-SS loss	0.87 ± 0.04	26.56 ± 0.03	0.75 ± .06	35.54 ± .06

B. Training and Implementation Details

Training was performed on a set of 3500 images of resolution size 1440×960 each, where ODIs were down-scaled by a required factor $r \in \{2, 4, 8\}$. While training, random crops of size 512×512 were applied to the images. Additional data augmentation techniques such as rotation and flipping were applied to scale up the training dataset.

Our proposed model was implemented using the Pytorch deep learning library [32]. To train each model, we set the batch size to 16. The weights of the all the layers were initialized randomly and the network was trained from the scratch. Further, to optimize the network we used the ADAM solver [33] with learning rates of 10^{-4}. Both control parameters β and γ are set equal to 10. All our models were trained in an end-to-end fashion for 100 epochs. Training was done using a 12 GB NVIDIA Titan-X GPU on an Intel Xeon E7 core i7 machine which took approximately 2 hours. Inference time is 0.030 milliseconds for each ODI. Please not that the inference time is same as that of SRGAN due to the similar generator architecture.

C. Quantitative and Qualitative Analysis

Here, we compare the performance between our proposed method and other state-of-the-art methods using quantitative and qualitative analysis. For comparison, we used well-known interpolation techniques: nearest-neighbor (NN) and bicubic, and the state-of-the-art SR algorithm SR-GAN, which is also the base of our proposed method. Quantitative results are summarized in Tables I-III and visual examples are provided in Figure 3 for qualitative analysis.

In Tables I-III, we show mean and variation values for each method over 500 test ODIs. Looking at the tables, with respect to PSNR and SSIM, we see that the scores obtained using our method have a marginal gain with respect to the other methods. However, for ODI metrics, we obtain higher gains. This accounts to the idea of preserving finer details around the equatorial regions as compared to regions near the poles. Additionally, we observe that our model performs better with higher down-sampling factors (*e.g.*, $8\times$). This further validates

our idea of using a Patch-GAN based discriminator instead of the more-layered (deeper) discriminator of SRGAN [14]. This analysis can also benefit other high-resolution imagery studies [34], such as HDR imaging [35, 36]. Furthermore, our proposed loss function facilitates learning of spherically distorted content thereby, resulting in higher gains.

To provide qualitative visual comparison between methods, Figure 3 shows some examples of constructed high-resolution ODIs for *Bicubic*, *SRGAN*, and our *proposed method*. We magnify details to display the most significant differences using rectilinear projection. As can be seen in the results, the proposed method can construct high-resolution ODI with higher-quality, compared to the benchmark methods.

As shown in the experimental results, the proposed approach achieves the highest performance on ERP format. However, other formats (*e.g.*, cubemap, equi-angular cube map, pyramid formats, etc. [1]) could also be employed by modifying the Eq. 4.

V. CONCLUSION

In this paper, we studied the super-resolution problem for omnidirectional images (ODIs) and proposed a new generative adversarial network model optimized to handle the artifacts typically obtained in the spherical observation space of ODIs. The proposed method utilized a PatchGAN discriminator to constrain the generator for predicting high-quality ODI output effectively. A novel loss function called 360-SS was also introduced to precisely measure the objective quality of the ODIs in spherical observation space. The proposed method was compared with state-of-the-art super-resolution techniques developed for standard images, and its performance was verified with quantitative and qualitative results. As future work, we plan to improve and apply the proposed method in ODI streaming application scenarios.

REFERENCES

[1] Y. Ye, E. Alshina, and J. Boyce, "Algorithm descriptions of projection format conversion and video quality metrics in 360lib," ISO/IEC JTC1/SC29/WG11/N16888, Hobart, AU, Tech. Rep. JVET-F1003, March 2017.

[2] A. Rana, C. Ozcinar, and A. Smolic, "Towards generating ambisonics using audio-visual cue for virtual reality," in *44th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019.

[3] C. Ozcinar, J. Cabrera, and A. Smolic, "Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS)*, vol. 9, no. 1, pp. 217–230, March 2019.

[4] C. Ozcinar, A. De Abreu, and A. Smolic, "Viewport-aware adaptive 360° video streaming using tiles for virtual reality," in *International Conference on Image Processing (ICIP)*, Beijing, China, Sep 2017.

[5] S. Croci, C. Ozcinar, E. Zerman, J. Cabrera, and A. Smolic, "Voronoi-based objective quality metrics for omnidirectional video," in *11th International Conference on Quality of Multimedia Experience (QoMEX)*, 2019.

[6] M. S. Elbamby, C. Perfecto, M. Bennis, and K. Doppler, "Toward low-latency and ultra-reliable virtual reality," *IEEE Network*, vol. 32, no. 2, pp. 78–84, March 2018.

[7] L. Bagnato, Y. Boursier, P. Frossard, and P. Vandergheynst, "Plenoptic based super-resolution for omnidirectional image sequences," in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2010, pp. 2829–2832.

[8] Z. Arican and P. Frossard, "Joint registration and super-resolution with omnidirectional images," *IEEE Transactions on Image Processing (TIP)*, vol. 20, no. 11, Nov 2011.

[9] A. Rana, G. Valenzise, and F. Dufaux, "Learning-based tone mapping operator for efficient image matching," *IEEE Transactions on Multimedia (TMM)*, 2018.

[10] A. Rana, G. Valenzise, and F. Dufaux, "Optimizing tone mapping operators for keypoint detection under illumination changes," in *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*, Sep. 2016, pp. 1–6.

[11] A. Rana, G. Valenzise, and F. Dufaux, "Learning-Based Tone Mapping Operator for Image Matching," in *IEEE International Conference on Image Processing (ICIP)*. Beijing, China: IEEE, 2017.

[12] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," *arXiv cs.CV 1809.07517*, 20 Sep. 2018.

[13] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3262–3271.

[14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and Others, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 4681–4690.

[15] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced Super-Resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[16] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard (GAN)," *arXiv cs.LG 1807.00734*, 2 Jul. 2018.

[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 1125–1134.

[18] K. Nasrollahi and T. B. Moeslund, "Super-resolution: a comprehensive survey," *Machine vision and applications (MVA)*, vol. 25, no. 6, pp. 1423–1468, 2014.

[19] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *arXiv cs.CV 1902.06068*, 16 Feb. 2019.

[20] A. Rana, J. Zepeda, and P. Perez, "Feature learning for the image retrieval task," in *Asian Conference on Computer Vision (ACCV)*. Springer, 2014, pp. 152–165.

[21] E. Zerman, A. Rana, and A. Smolic., "Colornet - estimating colorfulness in natural images," in *The International Conference on Image Processing (ICIP)*, 2019.

[22] Y. Zhou, M. Yu, H. Ma, H. Shao, and G. Jiang, "Weighted-to-spherically-uniform ssim objective quality evaluation for panoramic video," in *14th IEEE International Conference on Signal Processing (ICSP)*, Aug 2018, pp. 54–57.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing (TIP)*, vol. 13, no. 4, pp. 600–612, April 2004.

[24] JVET, "AHG8: WS-PSNR for 360 video objective quality evaluation," JTC1/SC29/WG11, ISO/IEC, Chengdu, CN, Tech. Rep. JVET-D0040, Oct. 2016.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems (NeurIPS)*, 2014, pp. 2672–2680.

[27] 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 (CVPR)*. IEEE, 2012, pp. 2695–2702.

[28] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Processing Letters (SPL)*, vol. 24, no. 9, pp. 1408–1412, Sep. 2017.

[29] G. Luz, J. Ascenso, C. Brites, and F. Pereira, "Saliency-driven omnidirectional imaging adaptive coding: Modeling and assessment," in *IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2017, pp. 1–6.

[30] Y. Zhang, Y. Wang, F. Liu, Z. Liu, Y. Li, D. Yang, and Z. Chen, "Subjective panoramic video quality assessment database for coding applications," *IEEE Transactions on Broadcasting (TBC)*, vol. 64, no. 2, pp. 461–473, Jun. 2018.

[31] W. Sun, K. Gu, S. Ma, W. Zhu, N. Liu, and G. Zhai, "A Large-Scale compressed 360-degree spherical image database: From subjective quality evaluation to objective model comparison," in *IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, Aug. 2018, pp. 1–6.

[32] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. De-Vito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Workshop Autodiff Submission in Conference on Neural Information Processing Systems (NIPS-W)*, 2017.

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

[34] Y. Tarabalka and A. Rana, "Graph-cut-based model for spectral-spatial classification of hyperspectral images," in *IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, July 2014, pp. 3418–3421.

[35] A. Rana, G. Valenzise, and F. Dufaux, "An evaluation of hdr image matching under extreme illumination changes," in *The International Conference on Visual Communications and Image Processing (VCIP)*, Chengdu, China, Nov. 2016.

[36] A. Rana, G. Valenzise, and F. Dufaux, "Learning-based Adaptive Tone Mapping for Keypoint Detection," in *IEEE International Conference on Multimedia & Expo (ICME)*, Hong Kong, China, Jul. 2017.