SphereDRUNet: A Spherical Denoiser for Omnidirectional Images

Rita Fermanian, Thomas Maugey, Christine Guillemot

SphereDRUNet: A Spherical Denoiser for Omnidirectional Images

Rita Fermanian * Thomas Maugey † Christine Guillemot ‡

Inria Rennes – Bretagne-Atlantique

Figure 1: Omnidirectional image denoising performed on (a) Equirectangular projection (ERP) with the state-of-the-art denoiser DRUNet re-trained on ERP images and on (b) the Sphere using our proposed SphereDRUNet. The denoised equirectangular image has lost high frequency elements such as the details on the leaves or the stains on the train, whereas denoising the sphere successfully removes noise while also recovering details.

Abstract

Image denoising is a primary pre-processing task in image processing. Although it has garnered significant research attention in the context of traditional 2D images, omnidirectional image denoising has received relatively limited attention in the literature. Furthermore, extending processing models and tools designed for 2D images to the sphere presents many challenges due to the inherent distortions and non-uniform pixel distributions associated with spherical representations and their underlying projections. In this paper, we address the problem of omnidirectional image denoising and we aim to study the advantage of denoising the spherical image directly rather than its mapping. We introduce a novel network called SphereDRUNet to denoise spherical images using deep learning tools on a spherical sampling. We show that denoising directly the sphere using our network gives better performance, compared to denoising the projected equirectangular images with a similarly learned model.

Index Terms: Omnidirectional images—Inverse problems—Denoising—On-the-sphere learning

1 Introduction

From wavelet thresholding to advanced deep-learning architectures, image denoising is a classical yet still one of the most widely explored topics in image processing. Denoising an image is the process of recovering a clean image x, having a noisy measurement y, such that $y = x + \epsilon$; ϵ being the noise degradation. While different types of noise may arise in a given image due to several physical factors, the Central Limit Theorem [30] establishes that their accumulation often leads to a Gaussian distribution. Hence, the most popular assumption is that ϵ is an additive white Gaussian noise (AWGN) with standard deviation σ.

Throughout the past decades, multiple methods have been developed to solve image denoising. Traditional methods include non-local self similarity methods [3], such as BM3D [5] and LSSC [17], sparse representations such as NCSR [8], gradient models [20, 26] and Markov random field (MRF) models [13, 14]. Nevertheless, with the advancement of deep learning, many convolutional neural networks have been designed for denoising. These networks not only successfully outperform the traditional methods, but also are often computationally more effective during the test. Some of the most popular architectures are DnCNN [36], NLRN [16] and DRUNet [35]. The latter outperforms the state-of-the-art denoising algorithms. However, despite the massive work that has been done to denoise traditional 2D images, very little effort has been dedicated to omnidirectional image denoising.

In this paper, we address the problem of denoising omnidirectional images. Omnidirectional images, also known as spherical images, are high-resolution visual contents that cover a 360-degree field of view capturing the light field converging to a single point. They are captured by catadioptric or fish-eye based systems that are able to cover a whole hemisphere. Thanks to their human-friendly nature, they have gained a particular attention in the domains of Augmented Reality (AR) and Mixed Reality (MR). These cameras often suffer from inherent distortion and increased noise due to their wide field of view and unique optical characteristics. Consequently, the quality of captured omnidirectional images can be compromised by noise artifacts, limiting the accuracy of subsequent computer vision

*e-mail: rita.fermanian@inria.fr
†e-mail: thomas.maugey@inria.fr
‡e-mail: christine.guillemot@inria.fr

tasks in AR/MR. Therefore, omni-directional image denoising is a crucial pre-processing step to enhance the quality of the visual data used in AR/MR applications.

However, only a few works have addressed the problem of denoising omnidirectional images in the literature. Bigot et al. [2] adapt the Wiener filter and Tikhonov regularization to spherical images for the denoising application. Demonceaux et al. [6] denoise catadioptric inputs corrupted by White Gaussian noise by proposing a new neighborhood for Markov random fields (MRF) adapted for omnidirectional images. Furthermore, Alibouch et al. [11] adapt the Stein block thresholding method to omnidirectional images to remove additive white Gaussian noise. Finally, Phan et al. [23] solve poisson denoising by using a space-variant total-variation regularization on catadioptric images.

These methods use conventional techniques and are primarily tailored to address the challenges presented by catadioptric inputs, thus they depend on the nature of the acquisition. However, omnidirectional images rely on a specific geometry which is the sphere, and considering the sphere makes the approach more agnostic to the image acquisition.

On the other hand, in order to take advantage of computing tools and methods that have been developed for 2D images, omnidirectional images are often mapped to two-dimensional planar representations using equirectangular projection (ERP) [27] or cubemap projection (CMP) [18]. These non-linear mappings induce significant deformations leading to non-uniform spatial resolution, therefore the use of algorithms designed to process perspective 2D images without exploiting the true geometry behind the spherical framework yields limited results.

A commonly adopted approach for processing omnidirectional images involves applying two-dimensional processing techniques to their corresponding mappings, typically using the equirectangular projection [7, 19, 21, 31]. These algorithms are often taking into account the distortions caused by the underlying mapping.

An alternative approach is to work directly on the sphere to circumvent issues related to sampling distortion. However, spherical processing necessitates the specific design of processing tools to effectively address challenges arising from this geometry. Recently, Bidgoli et al. [1] have introduced a novel framework to perform an on-the-sphere learning for omnidirectional images. They solve spherical convolution in the pixel domain and reach the same filter expressiveness, consistency and complexity of 2D CNNs. Henceforth, we use the term "spherical" to refer to operations conducted directly on the sphere.

In this paper, we address the problem of omnidirectional image denoising on the sphere, which is an essential pre-processing step in AR/MR tasks. Our main goal is to show that image denoising gives better performance when it is performed directly on the sphere rather than via a mapping, even though it raises many challenges. Furthermore, we introduce a novel deep spherical denoiser (SphereDRUNet) by transforming the architecture of the DRUNet [35] using the elementary convolutional modules proposed by OSLO. Our network takes directly spherical inputs rather than equirectangular projections. We evaluate the performance of SphereDRUNet in comparison with the initial DRUNet retrained and applied on Equirectangular data. Our results show that working on the sphere leads to a significant performance gain for the task of image denoising. This also confirms that the OSLO solution can be successfully extended to inverse problems such as denoising. We further compare our SphereDRUNet to a graph-based convolution network and confirm that the convolution approach that we use is more efficient. The main contributions of this paper are two-folds: (i) We introduce the first spherical denoiser CNN designed and trained specifically for omnidirectional images, which gives promising results for different noise levels and (ii) we show that working directly on the sphere gives better denoising results rather than processing the projections.

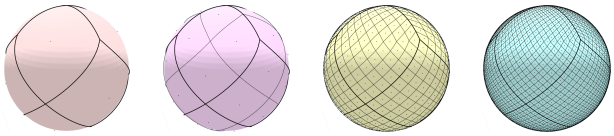
The remaining of the paper is organized as follows. In Section II, we elaborate spherical convolutions and explain the reasons behind choosing OSLO. We then detail the architecture of our proposed SphereDRUNet in Section III and present the comparison framework. In Section IV we show experimental results and validate the benefit of our approach. Finally, we conclude in Section V.

2 Related work

2.1 Spherical convolution

Extending conventional processing tools, such as convolutions designed for perspective imagery, to work effectively on spherical surfaces is not a straightforward task. In fact, the implementation of spherical convolutions comes with many challenges. First, considering registered images, a convolution is desired to be rotation equivariant for any rotation around the polar axis. Second, the filters should be able to increase their convolutional range and have anisotropic responses to ensure expressiveness. Finally, computational efficiency in spherical convolutions is very important for real-time processing and memory limitation.

Although different approaches for spherical convolutions have been proposed such as spectral learning [4, 24], 2D processing [28] and graph or multi-graph based representations [12, 22, 33], it hasn't been possible to assure these 3 key properties simultaneously to guarantee a proper spherical convolution. To overcome the limitations of existing methods, Bidgoli et al. [1] have proposed a novel framework for spherical processing (OSLO) where all aforementioned properties are successfully satisfied. For this purpose, they sample the sphere using HEALPix [9] and define the convolution in the pixel domain.



(a) N_{side}=1 (b) N_{side}=2 (c) N_{side}=8 (d) N_{side}=16

Figure 2: Visualization of HEALPix sampling over the Sphere. (a) Base resolution. (b) First resolution. (c) Third resolution. (d) Fourth resolution

2.2 HEALPix sampling

Hierarchical Equal Area isoLatitude Pixelation (HEALPix) [9] is a sampling method for spherical data that produces a configura-

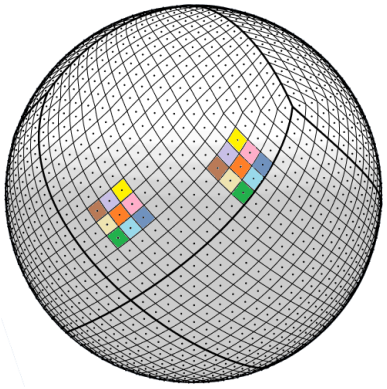


Figure 3: Spherical convolution in the pixel domain proposed by [1]. Each color represents a weight corresponding to a certain orientation. Filter weights remain the same, regardless of the position of the kernel on the sphere.

tion where the pixels are arranged in a diamond-shaped pattern. HEALPix starts by forming the base resolution and divides the surface of the sphere into 12 equal-area regions, each representing a specific area on the sphere. Then, each of these regions is iteratively partitioned into 2x2 equal-area sub-pixels, until reaching the target resolution. The HEALPix k^{th} resolution is defined by a parameter N_{side}=2^k and has a total of N_{pix}=$12N_{side}^2$ = 12 x 2^k x 2^k pixels (Fig. 2). Each pixel in the final arrangement has eight adjacent neighbors (only 24 pixels have seven neighbors, which is far from having a significant effect taking into account the resolution for an omnidirectional image). This ensures the regularity of the neighborhood, which is a fundamental condition for an expressive and effective convolution. Furthermore, the relative distance and orientation between neighboring pixels remains consistent throughout the entire sphere. Thus, HEALPix offers a rigid structure that was further proved in [1]. Both the regularity and the rigidity of the sampling method are essential properties to ensure an expressive and effective convolution.

2.3 On-the-sphere Learning for Omnidirectional Images (OSLO)

On-the-sphere Learning for Omnidirectional Images (OSLO) proposes a new framework that defines all necessary operators to build a CNN on the sphere. The method samples the sphere using HEALPix and performs convolutions in the pixel domain. Pooling, stride, skip-connections and patching are defined as well. The advantage of OSLO is that the convolutions are rotation equivariant, expressive and efficient at the same time.

In fact, let $\mathcal{N}_i(k)$ be the index assigned to the k^{th} neighbor for node i ($k = 1, ..., 8$ for a total of 8 adjacent neighbors of a vertex). Let L_{in} and L_{out} denote respectively the number of input and output features in the convolution. The convolution operation for an output feature l ($1 \leq l \leq L_{out}$) is defined as:

$$x_i^l = \langle \theta_0, x_i \rangle + \sum_{k=1}^{8} \langle \theta_k, x_{\mathcal{N}_i(k)} \rangle \cdot w_{\mathcal{N}_i(k),i} \qquad (1)$$

where θ_k represents the learnable filter weights and x_i is the input data/features at point i. $w_{\mathcal{N}_i(k),i}$ is set to 0 when the neighbor $\mathcal{N}_i(k)$ is missing and to 1 otherwise, in order to deal with the 24 exceptions of the pixels that have a missing adjacent neighbor.

The described convolution is anisotropic and consistent all over the sphere. The same weights are applied to compute the convolution output regardless of the position of the kernel on the sphere (Fig. 3). Yet, it only supports 1-hop neighborhood. In order to increase the potential size of kernels and extend the local support to n-hop neighborhood, the authors propose an iterative computation of the 1-hop convolution with a proper aggregation method such as concatenation, addition or max aggregation. This makes the convolution highly expressive. Furthermore, the convolution being in the pixel domain, it simply consists in translating the filter over the sphere, hence the complexity linearly increases with the number of pixels.

For these reasons, we choose OSLO to work on the sphere for omnidirectional image denosing. Further information about the other operations can be found in [1].

3 Spherical image denoising

The main objective of this paper is to explore whether omnidirectional image denoising is more efficient when performed directly on the sphere, compared to using a projection. For this purpose, we choose the DRUNet [35], which is a state-of-the-art Gaussian denoiser for perspective images, and use OSLO to transfer its architecture to the sphere. Hence, a novel spherical denoiser SphereDRUNet is obtained, which can be used to denoise spherical images corrupted with any noise level. By considering the sphere, our approach can be applied to all types omnidirectional images, regardless of their acquisition characteristics.

DRUNet is a bias-free model, which has been observed to be effective for generalization of denoisers over unseen noise levels. The model takes as input a noise level map and integrates Residual blocks (ResNet [10]) into U-Net [25]. It consists of four scales having 64, 128, 256 and 512 number of channels in each of the layers. Each scale in the downscaling consists of 4 residual blocks followed by a 2x2 strided convolution (SConv) whereas a 2x2 transposed convolution (TConv) is followed by 4 residual blocks during the upscaling. Four skip connections are set between 2x2 strided convolutions (SConv) downscaling and 2x2 transposed convolutions (TConv) blocks at each of the four scales.

We transfer DRUNet to the sphere using the OSLO framework. The architecture of the resulting SphereDRUNet is shown in Fig. 4. We use 1-hop neighborhood convolution for all the layers because all convolutions in DRUNet are of size 3x3. 2D stride operations of size 2x2 amount to stride 4 on the sphere. For implementation purposes, we replace the 2x2 transpose convolution by the sub-pixel convolution of size 4 (pixel-shuffle) since both unpooling methods have equivalent performance. Skip connections are set between the strided convolutions and the sub-pixel convolutions.

The network takes a noise level map as input along with the sphercial image. We note that various different approaches exist to estimate the noise level from the noisy image, including region-based methods, variational methods, as well as wavelet or transform domain methods [15].

3.1 Training details

For the purpose of a fair comparison between denoising a spherical image and its projection, we adapt our training framework to that of the DRUNet. We use the SUN360 equirectangular image dataset [32] and consider 2105 images for training, 22 images for validation and 43 images for testing. All images in the initial dataset have a spatial size of 9104x4552. For each epoch of the training, we degrade the training ERP images by AWGN with a random noise level σ chosen from [0,50], and concatenate a uniform map filled with σ having the same size as the ERP image. We perform HEALPix sampling of resolution 10 (i.e. 12,582,912 pixels) of the original image, the noisy image and the concatenation. SphereDRUNet takes as input the concatenated data in the spherical form, and learns to denoise the image by minimizing the l_1 loss between the estimated clean image and the ground-truth image using the ADAM optimization. The learning rate is set to 5e-5 and decreased by half every 50,000 iterations. The gradient update is performed once every 8 patches, patches being of resolution 8 (2^8 x 2^8 pixels). Finally, we train our network for a total of 25 epochs. Note that most of the modifications to the training framework compared to that of the DRUNet are essentially driven by the higher resolution of omnidirectional images in contrast to perspective images.

Furthermore, since our goal is to study denoising directly on the sphere compared to denoising the equirectangular image, applying the DRUNet trained on perspective images to ERP projections might not seem fair. Therefore, we re-train the DRUNet on ERP images in the same way as we trained the SphereDRUNet. In order to reduce the unfair bias caused by the presence of ground-truth images in ERP format, we resize the ERP images to 5056x2528 (12,781,568 pixels) so that both ERP and HEALPix have almost the same number of pixels.

3.2 Comparison framework

As discussed earlier, we proposed SphereDRUNet, a novel spherical denoising CNN and we re-trained the DRUNet with ERP images in order to apply it on spherical projections. We want to compare the performance of these two approaches in order to study whether omnidirectional image denoising is more efficient on the sphere or

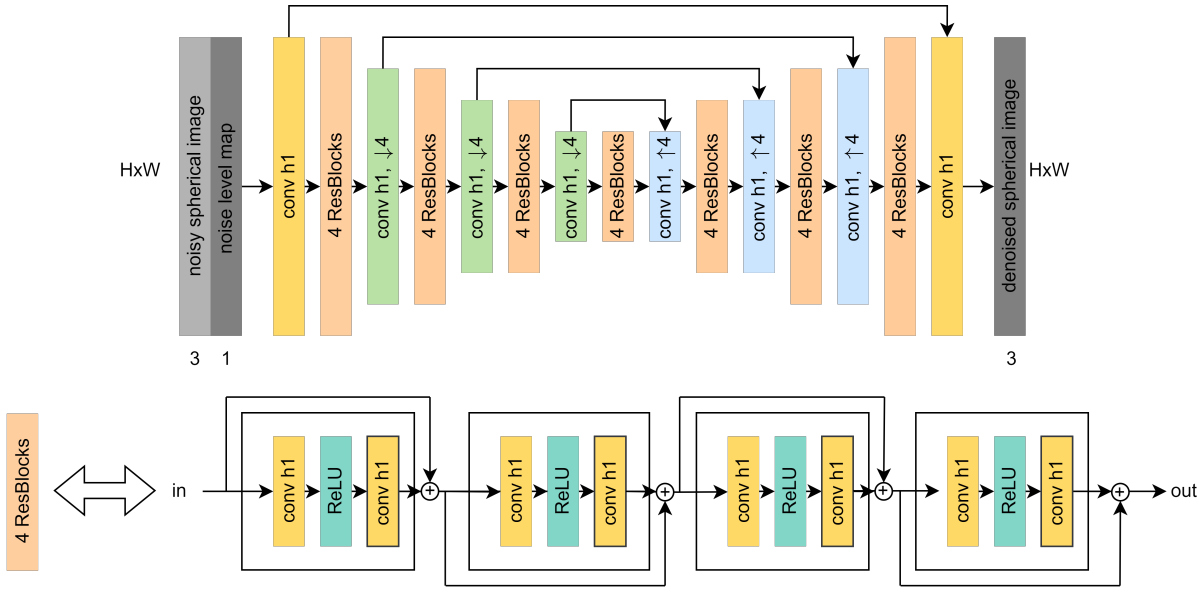


Figure 4: Architecture of the SphereDRUNet network with OSLO-based convolution.

on a corresponding projection. We also include a comparison with a cubemap projection: we map the equirectangular image to the six faces of a cube and each face is denoised with the DRUNet. We then reconstruct the ERP image by combining the six planes.

To avoid a biased comparison, for testing both spherical and mapping-based approaches, we first corrupt the high-resolution equirectangular input with Gaussian noise. Then, we downsample the corrupted image to get a HEALPix sampling of resolution 10 (12,582,912 pixels with $N_{side} = 2^{10}$) and resize the ERP image to 5056x2528 (12,781,568 pixels) in order to match the resolution of the HEALPix sampling. Thus, both test images are generated from the same corrupted input.

For quantitative evaluation, we calculate the Spherical PSNR (S-PSNR) [34] and the Weighted to Spherically uniform PSNR (WS-PSNR) [29]. The former maps the output and the ground-truth images (ERP or sphere) to a sphere by uniformly sampling 655,362 points, and computes the mean error between them to simulate the PSNR on the sphere. The sampling process for the S-PSNR calculation is different than the HEALPix sampling, hence the metric is not biased. S-PSNR is seen to be an estimation of the overall quality experienced by viewers across all potential views. On the other hand, in the calculation of the WS-PSNR, the mean squared error is weighted by the size of each pixel, as follows:

$$\text{WS-PSNR} = 10\log\left(\frac{\text{MAX}_I^2}{\text{W-MSE}}\right), \tag{2}$$

where MAX_I is the maximum pixel intensity in the image, and W-MSE is given as:

$$\text{W-MSE} = \frac{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}((\hat{I}(x,y) - I^{gt}(x,y))^2.w(y)}{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} w(y)}. \tag{3}$$

W and H denote the width and height of the image, $\hat{I}$ and I^{gt} represent respectively the reconstructed and the ground-truth images and $w(y)$ represents the weight of the corresponding pixel in the weight matrix.

We also show portions of mollweide projections of the denoised images for some visual comparison.

Table 1: Quantitative results of denoising Equirectangular, Cubemap and Spherical data in terms of WS-PSNR [dB].

	Noise level σ*255				
	10	20	30	40	50
Sphere + SphereDRUNet	**40.78**	**37.33**	**35.43**	**34.14**	**33.16**
ERP + DRUNet (trained on ERP)	39.85	35.38	32.46	30.15	28.16
CMP + DRUNet (trained on ERP)	30.30	29.65	28.73	27.61	26.38

4 Experimental results

We perform denoising of omnidirectional images corrupted with an Additive White Gaussian noise (AWGN) of noise level $\sigma * 255 \in [10, 20, 30, 40, 50]$. Table 1 and Table 2 show respectively the WS-PSNR and the S-PSNR of denoised equirectangular, cubemap and spherical images of the testing dataset. We can see that the on-the-sphere solution with our SphereDRUNet significantly outperforms denoising the equirectangular and the cubemap projections. This gain is even more important when the noise level is high. Fig. 1, Fig. 5 and Fig. 6 show a visual comparison of the denoised ERP and spherical images for a degradation of noise level σ= 30/255 and σ= 50/255. We observe that the DRUNet smoothes the ERP images and can not recover the high frequency details while the images denoised using SphereDRUNet on the sphere have less distortion and do not lose information (i.e. recover high frequency details such as texture). This is probably because in order to recover an equirectangular image with spatial distortions, a network needs a larger amount of weights to reconstruct high frequency components compared to recovering the same high frequency element on the sphere. For instance, a line on the sphere remains a line regardless of its position, whereas it can take different forms on the ERP depending on where it lies, so a network needs more weights to recover the same high frequency components on the ERP compared to the sphere.

4.1 Spherical convolution: OSLO-based vs. graph-based (DeepSphere)

Our hypothesis that omnidirectional image denoising is more efficient when performed on the sphere is verified. One popular approach to operate directly on irregular topology relies on graph signal processing. For the sake of completeness, we also want to prove that the OSLO-based spherical convolution that we have

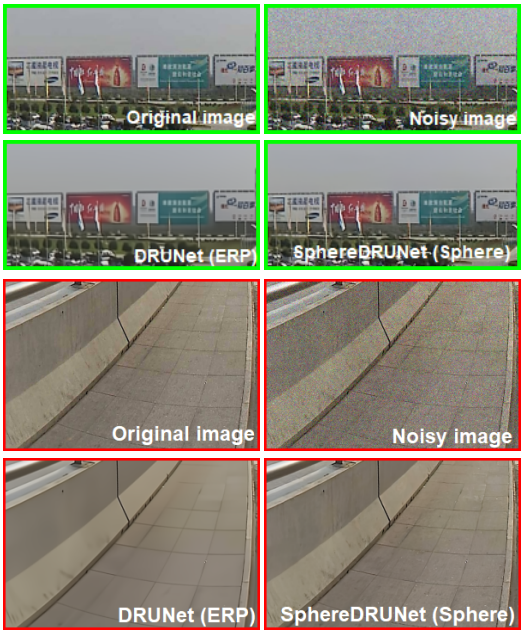


Figure 5: Zoomed portions of the mollweide projection of the Equirectangular image denoised by the DRUNet that was re-trained on ERP images and the Spherical image denoised by our SphereDRUNet. AWGN of $\sigma = 30/255$ was added on the original image.

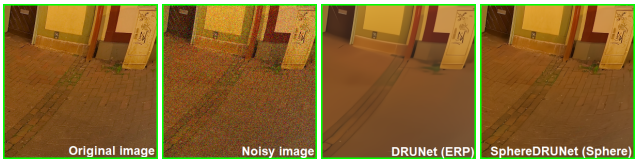


Figure 6: Zoomed portions of the mollweide projection of the Equirectangular image denoised by the DRUNet that was re-trained on ERP images and the Spherical image denoised by our SphereDRUNet. AWGN of $\sigma = 50/255$ was added on the original image.

chosen is the best approach. For this purpose, we consider a graph-based baseline: the DeepSphere architecture [22]. DeepSphere is also defined on the HEALPix sampling, but uses max-pooling and graph-based convolutions approximated by Chebyshev polynomial formulation. We re-train the same SphereDRUNet architecture with graph-based convolutions in the same training framework as our OSLO-based SphereDRUNet. For computational reasons, both networks are trained and evaluated on HEALPix images of resolution 8 (786,432 pixels) in this case.

Fig. 7 shows denoising results of spherical images restored with the OSLO-based SphereDRUNet and the graph-based SphereDRUNet. We can see that the graph-based convolution approach fails to recover the clean image from the noisy measurement. This limitation arises from the inherent nature of the graph, where the lack of directional information restricts the learning process to a single weight per neighborhood. By contrast, our OSLO-based approach demonstrates superior performance, successfully recovering the clean image from the noisy measurement. This highlights the importance of anisotropic filters and the effectiveness of the directional information embedded in the OSLO-based convolution, leading to enhanced denoising capabilities.

Table 2: Quantitative results of denoising Equirectangular, Cubemap and Spherical data in terms of S-PSNR [dB].

	Noise level σ*255				
	10	20	30	40	50
Sphere + SphereDRUNet	**22.28**	**20.47**	**19.49**	**18.83**	**18.32**
ERP + DRUNet (trained on ERP)	21.65	19.78	18.73	17.98	17.40
CMP + DRUNet (trained on ERP)	17.43	17.08	16.78	16.50	16.25

Table 3: Quantitative results of denoising Equirectangular images with the initial DRUNet network trained to denoise perspective 2D images and the DRUNet that we re-trained on ERP inputs. Results are in terms of WS-PSNR [dB].

	Noise level σ*255				
	10	20	30	40	50
ERP + DRUNet (trained on ERP)	**39.85**	**35.38**	**32.46**	**30.15**	**28.16**
ERP + DRUNet (trained on 2D images)	39.53	35.19	32.33	30.03	28.07
Gain	0.32	0.19	0.13	0.12	0.09

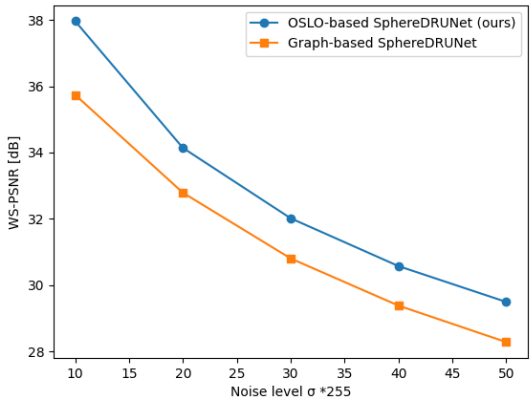


Figure 7: Denoising results of spherical images restored with the OSLO-based SphereDRUNet (proposed network) and graph-based SphereDRUNet. Results are in terms of WS-PSNR [dB].

4.2 Ablation study: training a denoiser on the equirectangular projection

Mapping-based solutions are a widely adopted approach when it comes to omnidirectional image processing. The reason is simple: 2D processing tools can be easily applied to these projections. However, traditional 2D operations (such as CNN tools) can not effectively process omnidirectional images, because these mappings display a significant distortion. In this section, we further demonstrate the limits of mapping-based learning. Table 3 shows the results (in terms of WS-PSNR [dB]) of denoising ERP images with the initial DRUNet network trained to denoise 2D images and the DRUNet that we re-trained on equirectangular projections. We observe that the performance of the DRUNet network does not greatly enhance when we re-train it over spherical projections compared to the initial DRUNet that was trained to denoise 2D images. For instance, the gain is only 0.09 dB for σ = 50/255, vs. a gain of 5.09 dB when we consider a processing on the sphere. This comparison further confirms that two-dimensional convolutions are not able to successfully learn spherical features. Due to their inherent flat nature, they cannot capture the curvatures and non-planar characteristics of spherical surfaces. Thus, 2D learning tools do not generalize well on 360 mappings in the task of denoising.

5 Conclusion

In this paper, we target omnidirectional image denoising, a critical pre-processing step in AR/MR applications. Our study investigates the effectiveness of performing denoising directly on the sphere, as opposed to employing a projection. To address this, we introduce SphereDRUNet, a novel approach that transfers a state-of-the-art

Gaussian denoiser to the sphere using the OSLO-based convolution. Notably, our network is the first spherical CNN that is trained specially for omnidirectional image denoising. Additionally, it has shown promising denoising results on the sphere for different noise levels. Our research demonstrates that denoising the spherical image gives superior results compared to denoising a corresponding mapping such as the equirectangular or the cubemap projections. This can be explained by the fact that a network needs less weights to recover the same high frequency components on the sphere compared to a projection. Future work includes leveraging our denoiser in spherical Plug-and-play algorithms to solve different inverse problems.

ACKNOWLEDGMENTS

The authors would like to thank Navid Mahmoudian Bidgoli for his assistance during this project. This work was supported by the french ANR research agency in the context of the artificial intelligence project DeepCIM.

REFERENCES

[1] N. M. Bidgoli, R. G. d. A. Azevedo, T. Maugey, A. Roumy, and P. Frossard. Oslo: On-the-sphere learning for omnidirectional images and its application to 360-degree image compression. *IEEE Transactions on Image Processing*, 31:5813–5827, 2022.

[2] S. Bigot, D. Kachi, S. Durand, and E. M. Mouaddib. Spherical image denoising and its application to omnidirectional imaging. In *VISAPP (1)*, pp. 101–108, 2007.

[3] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 2, pp. 60–65. IEEE, 2005.

[4] T. S. Cohen, M. Geiger, J. Köhler, and M. Welling. Spherical cnns. *arXiv preprint arXiv:1801.10130*, 2018.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[6] C. Demonceaux and P. Vasseur. Markov random fields for catadioptric image processing. *Pattern Recognition Letters*, 27(16):1957–1967, 2006.

[7] X. Deng, H. Wang, M. Xu, L. Li, and Z. Wang. Omnidirectional image super-resolution via latitude adaptive network. *IEEE Transactions on Multimedia*, 2022.

[8] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing*, 22(4):1620–1630, 2012.

[9] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759, 2005.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. doi: 10.1109/cvpr.2016.90

[11] A. Iazzi, A. Radgui, M. Rziza, et al. An adapted block thresholding method for omnidirectional image denoising. *Research Journal of Applied Sciences, Engineering and Technology*, 8(18):1966–1972, 2014.

[12] R. Khasanova and P. Frossard. Geometry aware convolutional filters for omnidirectional images representation. In *International Conference on Machine Learning*, pp. 3351–3359. PMLR, 2019.

[13] X. Lan, S. Roth, D. Huttenlocher, and M. J. Black. Efficient belief propagation with learned higher-order markov random fields. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II 9*, pp. 269–282. Springer, 2006.

[14] S. Z. Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.

[15] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang. Noise estimation from a single image. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 901–908. IEEE, 2006.

[16] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang. Non-local recurrent network for image restoration. *Advances in neural information processing systems*, 31, 2018.

[17] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *2009 IEEE 12th international conference on computer vision*, pp. 2272–2279. IEEE, 2009.

[18] K.-T. Ng, S.-C. Chan, and H.-Y. Shum. Data compression and transmission aspects of panoramic videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(1):82–95, 2005.

[19] A. Nishiyama, S. Ikehata, and K. Aizawa. 360° single image super resolution via distortion-aware network and distorted perspective images. In *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 1829–1833, 2021. doi: 10.1109/ICIP42928.2021.9506233

[20] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[21] C. Ozcinar, A. Rana, and A. Smolic. Super-resolution of omnidirectional images using adversarial learning. In *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6. IEEE, 2019.

[22] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier. Deepsphere: Efficient spherical convolutional neural network with healpix sampling for cosmological applications. *Astronomy and Computing*, 27:130–146, 2019.

[23] T. D. K. Phan and T. H. Y. Tran. A space-variant nonlinear algorithm for denoising omnidirectional images corrupted by poisson noise. *IEEE Signal Processing Letters*, 27:535–539, 2020.

[24] P. J. Roddy and J. D. McEwen. Sifting convolution on the sphere. *IEEE Signal Processing Letters*, 28:304–308, 2021.

[25] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28

[26] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[27] J. P. Snyder. *Flattening the earth: two thousand years of map projections*. University of Chicago Press, 1997.

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

[29] 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.

[30] P. Tchebycheff. Sur deux théorèmes relatifs aux probabilités. pp. 305–315, 1890.

[31] E. Upenik, P. Akyazi, M. Tuzmen, and T. Ebrahimi. Inpainting in omnidirectional images for privacy protection. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2487–2491. IEEE, 2019.

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

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

[34] M. Yu, H. Lakshman, and B. Girod. A framework to evaluate omnidirectional video coding schemes. In *2015 IEEE international symposium on mixed and augmented reality*, pp. 31–36. IEEE, 2015.

[35] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.

[36] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.