Efficient View Synthesis and 3D-based Multi-Frame Denoising with Multiplane Feature Representations

Supplementary Materials

Thomas Tanay Aleš Leonardis Matteo Maggioni
Huawei Noah's Ark Lab
{thomas.tanay,ales.leonardis,matteo.maggioni}@huawei.com

A. Network architecture

The architecture of our Multiplane Feature Encoder-Renderer (MPFER) is described in the main paper and illustrated in Figure 3. The Encoder and the Renderer consist in two identical Unets with a base of 64 channels, illustrated in more details in Figure S2.

B. Average metrics

In Section 4.1, Table 2, We compare our MPFER model to various 2D-based video restoration methods for denoising of synthetic noise on the Spaces dataset. Two of the methods we consider, BPN [4] and DeepRep [1], are burst denoising methods producing only one denoised output for the entire set of noisy inputs. By default, we chose this output to be frame number 6 at the center of the camera rig and compared the performances of all the methods on that frame. However, MPFER as well as BasicVSR [2] and BasicVSR++ [3], are multi-frame denoising methods producing one denoised output per noisy input. We compare their average performances over the 16 frames of the validation sequences in Table S1. We see that the overall performances are comparable to those on frame 6 from Table 2. In particular, MPFER outperforms all other methods by large margins on all noise levels. To qualitatively evaluate the cross-view consistency of different methods, we also plot V×W slices computed on scene 52 in Figure S1. We run BPN and DeepRep 16 times (once per frame) to obtain these profiles. Our method qualitatively matches the ground-truth better than other methods.

Figure S1. V×W slices computed on scene 52 of Spaces.

C. Ablations

Our MPFER method depends on three hyperparameters: the number of depth planes D, the number of channels in the multiplane representation C, and the upscaling factor of the PSV/MPF representation s. In Table 2 of the main paper, we evaluated the influence of model size by varying these three hyperparameters simultaneously. We now evaluate the influence of each hyperparameter independently in Table S2. We see that the performance of the method increases with D, C and s, and so does the computational complexity. Interestingly, the performance improvement is higher when C increases from 4 to 16 (+0.4dB at Gain 20), than when D increases from 16 to 64 (+0.23dB at Gain 20), while the increase in computational complexity is significantly lower ($\times 1.33$ vs $\times 2.97$ respectively). This observation confirms that multiplane features are inherently more powerful representations than multiplane images, allowing to perform efficient 3D-based video restoration with fewer depth planes.

D. Qualitative evaluations

We consider 4 experimental setups in the main paper: (1) Novel View Synthesis on Spaces, (2) Denoising on Spaces, (3) Denoising on the Real Forward Facing dataset, and (4) Novel View Synthesis under Noise Conditions on the Real Forward Facing dataset. We present some visual comparisons with state-of-the-art methods for the setups (2) and (3) in Figure 4, and for the setup (4) in Figure 1. We present some additional visual comparisons for the setup (1) in Figure S3 here.

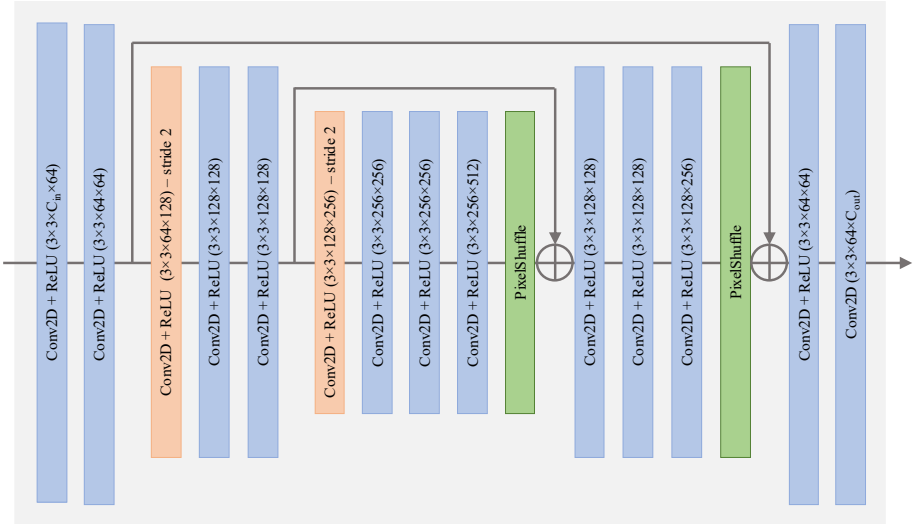

Figure S2. Unet architecture used for the Encoder and for the Renderer in all our MPFER experiments.

	Gain 4			Gain 8			Gain 16			Gain 20			GFlops@
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	500×800
VBM4D	32.30	0.90	n/a	30.12	0.849	n/a	27.53	0.763	n/a	26.58	0.723	n/a	n/a
VNLB	33.41	0.917	0.089	30.31	0.869	0.144	25.79	0.794	0.279	23.58	0.746	0.363	n/a
BasicVSR	36.86	0.957	0.029	34.45	0.935	0.052	31.62	0.895	0.099	30.59	0.875	0.124	2090
BasicVSR++	36.81	0.957	0.030	34.39	0.934	0.051	31.62	0.895	0.091	30.60	0.875	0.111	4300
UNet-SF	35.15	0.942	0.042	32.67	0.910	0.074	29.86	0.857	0.134	28.87	0.834	0.160	**440**
UNet-BR	35.23	0.943	0.040	32.72	0.912	0.070	29.97	0.861	0.124	29.02	0.840	0.148	**470**
UNet-BR-OF	36.37	0.955	0.029	34.18	0.934	0.049	31.65	0.896	0.091	30.71	0.878	0.112	710
MPFER-16	37.20	0.965	0.021	35.37	0.952	0.033	33.22	0.927	0.055	32.41	0.915	0.067	**470**
MPFER-32	37.52	0.967	**0.020**	35.69	0.954	**0.030**	33.50	0.931	0.051	32.66	0.919	0.063	1210
MPFER-64	**37.60**	**0.968**	**0.020**	**35.78**	**0.955**	**0.030**	**33.58**	**0.932**	**0.050**	**32.74**	**0.920**	**0.061**	1810

Table S1. Denoising on Spaces. Average metrics over the 16 frames in the validation sequences. Best results in **bold**, second best underlined.

	Gain 4			Gain 8			Gain 16			Gain 20			GFlops@
(D, C, s)	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	500×800
Influence of the number of depth planes													
$(16, 8, 1.25)$	37.75	0.969	0.019	35.98	0.957	0.028	33.83	0.934	0.049	33.00	0.922	0.061	610
$(32, 8, 1.25)$	37.86	0.970	0.018	36.11	0.958	0.027	33.95	0.935	0.047	33.10	0.924	0.059	1010
$(64, 8, 1.25)$	38.00	0.970	0.018	36.25	0.959	0.027	34.08	0.936	0.047	33.23	0.925	0.057	1810
Influence of the number of channels in the multiplane representation													
$(32, 4, 1.25)$	37.64	0.969	0.021	35.81	0.957	0.031	33.59	0.935	0.051	32.74	0.923	0.062	910
$(32, 8, 1.25)$	37.86	0.970	0.018	36.11	0.958	0.027	33.95	0.935	0.047	33.10	0.924	0.059	1010
$(32, 16, 1.25)$	37.94	0.970	0.019	36.17	0.958	0.028	33.99	0.936	0.047	33.14	0.924	0.058	1210
Influence of the upscaling factor													
$(16, 8, 1.0)$	37.56	0.968	0.020	35.80	0.955	0.030	33.70	0.933	0.051	32.89	0.921	0.063	470
$(16, 8, 1.25)$	37.75	0.969	0.019	35.98	0.957	0.028	33.83	0.934	0.049	33.00	0.922	0.061	610
$(16, 8, 1.5)$	37.86	0.969	0.019	36.08	0.957	0.029	33.92	0.935	0.050	33.08	0.923	0.061	780

Table S2. Denoising on Spaces. Influence of hyperparameters (D, C, s): number of depth planes, number of channels in the multiplane representation, upscaling factor.

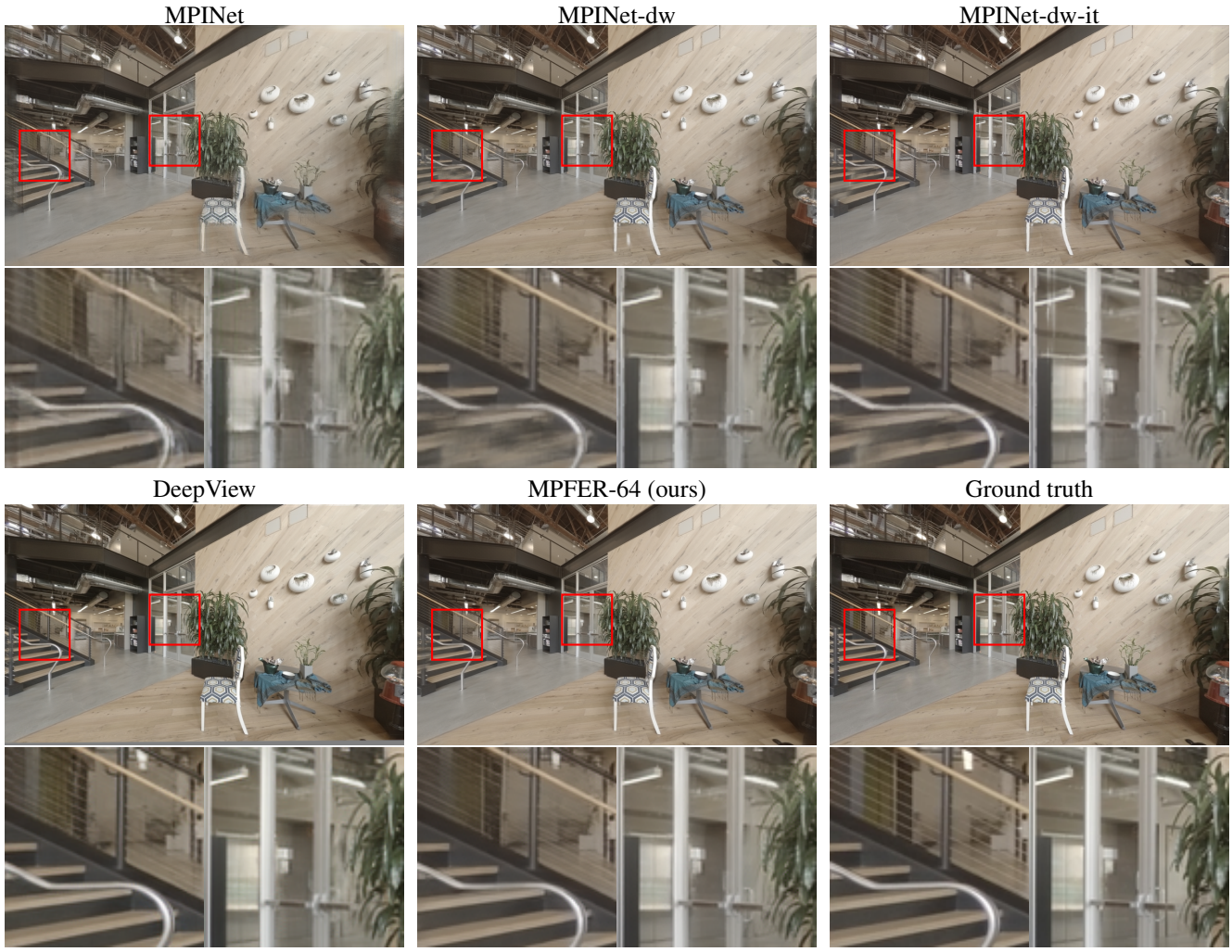

Figure S3. Qualitative evaluation for novel view synthesis on Spaces (best viewed zoomed in).

References

[1] Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. Deep reparametrization of multi-frame super-resolution and denoising. In *ICCV*, pages 2460–2470, 2021. 1

[2] Kelvin CK Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Basicvsr: The search for essential components in video super-resolution and beyond. In *CVPR*, pages 4947–4956, 2021. 1

[3] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *CVPR*, pages 5972–5981, 2022. 1

[4] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In *CVPR*, pages 11844–11853, 2020. 1