Classic Video Denoising in a Machine Learning World: Robust, Fast, and Controllable

Supplementary Material

1. Qualitative Results

We provide test results of various methods from the main paper, along with a comparative web demo. The demo allows users to compare different methods and adjust brightness, contrast, and saturation to achieve better comparative results. Additionally, we provide a web demo based on different denoising strengths, allowing users to explore the results of spatial and temporal denoising and compare the detail preservation capabilities under varying denoising levels. Please refer to these demos from our project page: https://srameo.github.io/projects/levd.

2. More Quantitative Results

We provide the LPIPS [17] and SSIM [15] scores of different methods on the CRVD benchmark in Tab. 3.

3. More Ablative Experiments

As summarized in Tab. 1 (top and middle), we also tried using ResNet-101 [5] and SDXL VAE [8] as the backbone as well as RAFT [13] and PWC-Net [11] as the motion estimator. These ablations do not offer too much signal other than that our initial choices for these are reasonable.

Our experimental results show that while RAFT [13] offers slightly better performance (PSNR improvement of 0.17, SSIM improvement of 0.0057), its runtime is more than 5 times that of SpyNet [9] (24.53 ms vs 4.56 ms) as shown in Tab. 2. PWC-Net [11], though faster than RAFT, is still 3 times slower than SpyNet and performs slightly worse than our method. Therefore, we chose SpyNet as our optical flow estimator, primarily considering the optimal balance between performance and speed.

For the backbone selection, as shown in Tab. 1, ConvNext [7] outperforms ResNet-101 [5] and SDXL VAE [8] across all metrics. ResNet-101 shows a decrease of 0.17 in PSNR, 0.0024 in SSIM, and an increase of 0.0109 in LPIPS; while SDXL VAE exhibits an even more significant performance drop. These results confirm the rationale behind our choice of ConvNext as the backbone.

As shown in Fig. 1, omitting H.264 transcoding from the data pipeline results in numerous temporal compression artifacts in the denoising results. These artifacts primarily manifest as inconsistencies between video frames, significantly degrading the final visual quality.

As shown in Fig. 2, the denoising results with different anchor frame choices are similar though the noise level of the anchor frames differs a lot.

	PSNR	delta	SSIM	delta	LPIPS	delta
ConvNext - Ours	36.04	–	0.9472	–	0.0763	–
ResNet-101	35.87	- 0.17	0.9448	- 0.0024	0.0872	+ 0.0109
SDXL VAE	35.79	- 0.25	0.9439	- 0.0033	0.0949	+ 0.0186
SpyNet - Ours	36.04	–	0.9472	–	0.0763	–
RAFT	36.22	+ 0.17	0.9529	+ 0.0057	0.0801	+ 0.0038
PWC-Net	35.96	- 0.09	0.9464	- 0.0007	0.0919	+ 0.0156

Table 1. Additional experiments on the CRVD (sRGB) dataset.

	Flow + Align	$\mathcal{P}(\cdot;\theta)$	Temp. Denoise	Spat. Denoise
SpyNet - Ours	4.56 ms	1.15 ms	6.20 ms	2.37 ms
RAFT	24.53 ms	1.19 ms	6.17 ms	2.32 ms
PWC-Net	14.29 ms	1.22 ms	6.13 ms	2.36 ms

Table 2. Runtime for each denoising stage at a 720p resolution.

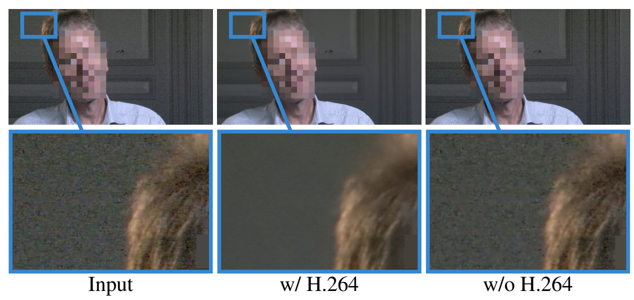

Figure 1. Our approach w/ and w/o H.264 augmentation.

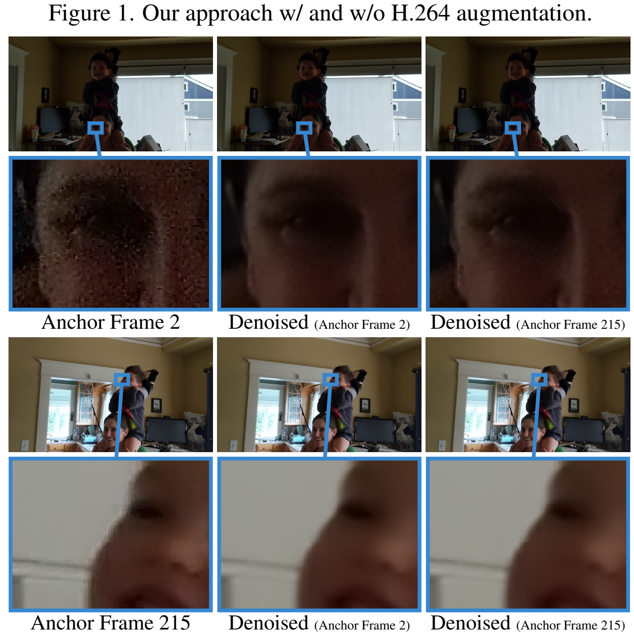

Figure 2. Denoising results with different anchor frames choices.

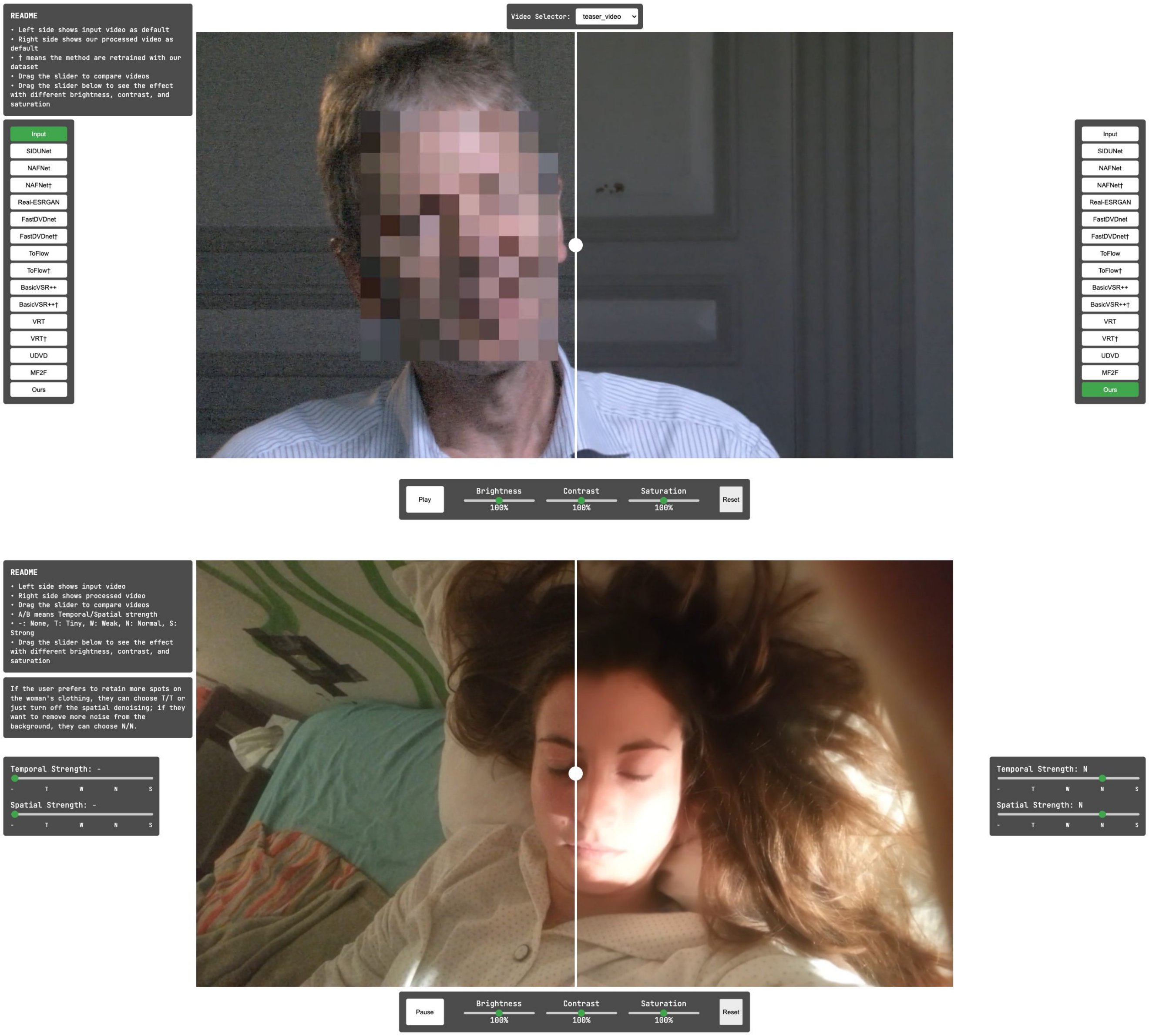

Figure 3. Screenshots of our interactive web demos. Top: a comparison interface that allows users to examine and compare results from different denoising methods side-by-side. Bottom: a control interface that enables users to interactively adjust spatial and temporal denoising strengths to explore the trade-off between detail preservation and noise reduction.

References

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

[2] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 3

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, 2022. 3

[4] Valéry Dewil, Jérémy Anger, Axel Davy, Thibaud Ehret, Gabriele Facciolo, and Pablo Arias. Self-supervised training for blind multi-frame video denoising. In *WACV*, 2021. 3

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[6] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *IEEE TIP*, 2024. 3

[7] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *CVPR*, 2022. 1

[8] Dustin Podell, Zion English, Kyle Lacey, Andreas

	ISO 1600		ISO 3200		ISO 6400		ISO 12800		ISO 25600		Overall		Speed	
	SSIM (higher SSIM is better)	rank	SSIM (higher SSIM is better)	rank	SSIM (higher SSIM is better)	rank	SSIM (higher SSIM is better)	rank	SSIM (higher SSIM is better)	rank	SSIM (higher SSIM is better)	rank	FPS (higher FPS is better)	rank
SID [2]†	0.9689	6th of 10	0.9622	6th of 10	0.9522	4th of 10	0.9289	3rd of 10	0.8162	4th of 10	0.9257	3rd of 10	6.95	3rd of 10
NAFNet [3]†	0.9727	3rd of 10	0.9663	3rd of 10	0.9580	1st of 10	0.9345	2nd of 10	0.8531	2nd of 10	0.9369	2nd of 10	1.69	7th of 10
Real-ESRGAN [14]	0.8906	10th of 10	0.8612	10th of 10	0.8542	10th of 10	0.8541	9th of 10	0.8521	3rd of 10	0.8624	9th of 10	0.24	8th of 10
FastDVDNet [12]†	0.9712	5th of 10	0.9651	4th of 10	0.9510	5th of 10	0.9135	5th of 10	0.7685	7th of 10	0.9139	5th of 10	5.72	4th of 10
TOFlow [16]†	0.9636	8th of 10	0.9557	8th of 10	0.9408	7th of 10	0.9000	6th of 10	0.7573	8th of 10	0.9035	6th of 10	2.84	6th of 10
BasicVSR++ [1]†	0.9721	4th of 10	0.9664	2nd of 10	0.9425	6th of 10	0.8568	8th of 10	0.6285	9th of 10	0.8733	8th of 10	7.41	2nd of 10
VRT [6]†	0.9730	2nd of 10	0.9644	5th of 10	0.9287	8th of 10	0.8133	10th of 10	0.5601	10th of 10	0.8479	10th of 10	0.05	10th of 10
UDVD [10]	0.9461	9th of 10	0.9352	9th of 10	0.9147	9th of 10	0.8819	7th of 10	0.7831	6th of 10	0.8922	7th of 10	0.16	9th of 10
MF2F [4]	0.9657	7th of 10	0.9612	7th of 10	0.9537	3rd of 10	0.9271	4th of 10	0.7960	5th of 10	0.9207	4th of 10	4.62	5th of 10
Ours	0.9740	1st of 10	0.9665	1st of 10	0.9560	2nd of 10	0.9381	1st of 10	0.9013	1st of 10	0.9472	1st of 10	31.66	1st of 10

	ISO 1600		ISO 3200		ISO 6400		ISO 12800		ISO 25600		Overall		Speed	
	LPIPS (lower LPIPS is better)	rank	LPIPS (lower LPIPS is better)	rank	LPIPS (lower LPIPS is better)	rank	LPIPS (lower LPIPS is better)	rank	LPIPS (lower LPIPS is better)	rank	LPIPS (lower LPIPS is better)	rank	FPS (lower FPS is better)	rank
SID [2]†	0.0307	5th of 10	0.0358	7th of 10	0.0558	4th of 10	0.0901	3rd of 10	0.2767	4th of 10	0.0978	3rd of 10	6.95	3rd of 10
NAFNet [3]†	0.0240	3rd of 10	0.0353	6th of 10	0.0475	2nd of 10	0.0886	2nd of 10	0.2495	2nd of 10	0.0890	2nd of 10	1.69	7th of 10
Real-ESRGAN [14]	0.1466	10th of 10	0.1808	10th of 10	0.1964	10th of 10	0.2077	8th of 10	0.2622	3rd of 10	0.1987	10th of 10	0.24	8th of 10
FastDVDNet [12]†	0.0309	6th of 10	0.0349	5th of 10	0.0577	5th of 10	0.1315	5th of 10	0.3628	7th of 10	0.1236	5th of 10	5.72	4th of 10
TOFlow [16]†	0.0393	8th of 10	0.0489	8th of 10	0.0811	6th of 10	0.1604	6th of 10	0.4048	8th of 10	0.1469	6th of 10	2.84	6th of 10
BasicVSR++ [1]†	0.0264	4th of 10	0.0349	4th of 10	0.0870	7th of 10	0.2078	9th of 10	0.4812	9th of 10	0.1675	8th of 10	7.41	2nd of 10
VRT [6]†	0.0229	2nd of 10	0.0347	2nd of 10	0.1034	8th of 10	0.2584	10th of 10	0.5405	10th of 10	0.1920	9th of 10	0.05	10th of 10
UDVD [10]	0.0750	9th of 10	0.0877	9th of 10	0.1222	9th of 10	0.1738	7th of 10	0.3593	6th of 10	0.1636	7th of 10	0.16	9th of 10
MF2F [4]	0.0347	7th of 10	0.0349	3rd of 10	0.0449	1st of 10	0.0952	4th of 10	0.3522	5th of 10	0.1124	4th of 10	4.62	5th of 10
Ours	0.0198	1st of 10	0.0304	1st of 10	0.0513	3rd of 10	0.0861	1st of 10	0.1940	1st of 10	0.0763	1st of 10	31.66	1st of 10

Table 3. Video denoising results on the CRVD (sRGB) benchmark. The resutls demonstrate that our approach not only achieves the best overall performance but is also four times faster than the second-fastest method. For detailed PSNR metrics on the CRVD dataset, please refer to Tab. 2 of the main paper.

Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *ICLR*, 2024. 1

[9] Anurag Ranjan and Michael J. Black. Optical flow estimation using a spatial pyramid network. In *CVPR*, 2017. 1

[10] Dev Yashpal Sheth, Sreyas Mohan, Joshua L Vincent, Ramon Manzorro, Peter A Crozier, Mitesh M Khapra, Eero P Simoncelli, and Carlos Fernandez-Granda. Unsupervised deep video denoising. In *ICCV*, 2021. 3

[11] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *CVPR*, 2018. 1

[12] Matias Tassano, Julie Delon, and Thomas Veit. Fastdvdnet: Towards real-time deep video denoising without flow estimation. In *CVPR*, 2020. 3

[13] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *ECCV*, 2020. 1

[14] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV Workshop*, 2021. 3

[15] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 2004. 1

[16] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *IJCV*, 2019. 3

[17] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 1