EVRNet: Efficient Video Restoration on Edge Devices - Supplementary

Sachin Mehta*
University of Washington
USA

Amit Kumar
Facebook Inc.
USA

Fitsum Reda†
Google
USA

Varun Nasery
Facebook Inc.
USA

Vikram Mulukutla
Facebook Inc.
USA

Rakesh Ranjan
Facebook Inc.
USA

Vikas Chandra
Facebook Inc.
USA

ACM Reference Format:
Sachin Mehta, Amit Kumar, Fitsum Reda, Varun Nasery, Vikram Mulukutla, Rakesh Ranjan, and Vikas Chandra. 2021. EVRNet: Efficient Video Restoration on Edge Devices - Supplementary. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3474085.3475477

1 ABLATIONS

Effect of different CUs: Table 1 studies the effect of single- and multi-scale convolutional units (CUs) with and without SE unit. Multi-scale CU units with SE help improve the performance in case of AWGN denoising while no gain was observed in case of deblocking and super-resolution. We hypothesize that this is because compression happens at macro-block level, and both single and multi-scale blocks are able to effectively remove compression artifacts. Unlike macro-block compression, AWGN noise is identically distributed in the frames and kernels at different scales helps learn better representations and remove noisy artifacts (see gray color row in Table 1b).

Effect of the depth of alignment, differential, and fusion modules: Table 2 studies EVRNet with different values of N_A, N_D, and N_F. We are interested in efficient networks for edge devices, therefore, we studied only those combinations that satisfies this criteria: $N_A + N_D + N_F = 9$. Similar to the effect of different CUs, we did not observe much gains when varying the depth of alignment, differential, and fusion modules for the task of deblocking and super-resolution. However, for denoising, we found that deeper alignment modules delivers the best trade-off between performance and MACs. Therefore, in our main experiments, we used $N_A = 5$, $N_D = 2$, and $N_F = 2$ (see gray color row in Table 2).

2 QUALITATIVE RESULTS ON THE VIMEO-90K DATASET

2.1 Deblocking

Figures 1, 2, and 3 demonstrate EVRNet's ability in deblocking videos at different compression factors in diverse environments (Q; lower value of Q means higher compression). For example, in Figure 1b, EVRNet is able to remove the macro-block artifacts even under high compression ($Q = 15$) around objects (e.g., hand, vegetables, and mixing bowl).

2.2 Denoising

Figures 4, 5, 6, 7, and 8 demonstrates EVRNet's ability in denoising different types of noise in diverse scenes. For example, in Figure 4c, EVRNet is able to remove the noise and restore videos with high-quality.

2.3 Video super-resolution ($4\times$)

Figure 9 and 10 shows that EVRNet is effective in restoring the details for $4\times$ video super-resolution in diverse settings. For example, in Figure 10a, EVRNet is able to restore fine details (e.g., hair strands) which are hard to restore with bicubic interpolation.

*Work completed during internship at Facebook Inc..
†Work done while working at Facebook Inc..

MM '21, October 20–24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8651-7/21/10…$15.00
https://doi.org/10.1145/3474085.3475477

				RGB		Y-Channel	
CU Type	SE Unit	MACs	# Params	PSNR	SSIM	PSNR	SSIM
Single	✗	9.85 G	68.15 K	36.358	0.948	38.477	0.961
Single	✓	9.85 G	72.95 K	36.323	0.948	38.403	0.961
Multi	✗	10.79 G	73.91 K	36.297	0.947	38.363	0.961
Multi	✓	10.79 G	78.71 K	36.334	0.948	38.478	0.962

(a) Deblocking ($Q = 40$)

				RGB		Y-Channel	
CU Type	SE Unit	MACs	# Params	PSNR	SSIM	PSNR	SSIM
Single	✗	9.85 G	68.15 K	31.207	0.868	32.650	0.886
Single	✓	9.85 G	72.95 K	32.006	0.896	33.365	0.914
Multi	✗	10.79 G	73.91 K	29.026	0.875	30.247	0.895
Multi	✓	10.79 G	78.71 K	32.370	0.900	33.679	0.916

(b) AWGN Denoising ($\sigma^2 = 0.001$)

				RGB		Y-Channel	
CU Type	SE Unit	MACs	# Params	PSNR	SSIM	PSNR	SSIM
Single	✗	9.90 G	68.33 K	37.406	0.962	38.042	0.966
Single	✓	9.90 G	73.14 K	37.318	0.962	37.955	0.965
Multi	✗	10.84 G	74.10 K	37.181	0.962	37.868	0.966
Multi	✓	10.84 G	78.91 K	37.378	0.962	38.002	0.966

(c) Super-resolution (2×)

Table 1: Effect of different CU units. Multi-scale blocks are effective in restoring fine-grained details (e.g., noise) while both single- and multi-scale blocks are effective in restoring block-level artifacts (e.g., compression). Here, we used $N_A = N_D = N_F = 3$.

Module depth					RGB		Y-Channel	
N_A	N_D	N_F	MACs	# Params	PSNR	SSIM	PSNR	SSIM
1	1	7	11.44 G	78.71 K	36.320	0.948	38.411	0.961
1	7	1	11.44 G	78.71 K	36.356	0.948	38.450	0.962
7	1	1	9.47 G	78.71 K	36.334	0.948	38.472	0.961
2	2	5	11.11 G	78.71 K	36.200	0.946	38.297	0.960
2	5	2	11.11 G	78.71 K	36.327	0.948	38.412	0.962
5	2	2	10.13 G	78.71 K	36.307	0.947	38.403	0.961
3	2	4	10.77 G	78.71 K	36.359	0.948	38.451	0.962
3	4	2	10.77 G	78.71 K	36.307	0.947	38.390	0.961
4	3	2	10.46 G	78.71 K	36.287	0.948	38.405	0.961
3	3	3	10.79 G	78.71 K	36.334	0.948	38.478	0.962

(a) Deblocking ($Q = 40$)

Module depth					RGB		Y-Channel	
N_A	N_D	N_F	MACs	# Params	PSNR	SSIM	PSNR	SSIM
1	1	7	11.44 G	78.71 K	31.605	0.887	32.913	0.905
1	7	1	11.44 G	78.71 K	31.753	0.884	32.951	0.901
7	1	1	9.47 G	78.71 K	30.859	0.871	32.139	0.890
2	2	5	11.11 G	78.71 K	32.139	0.901	33.477	0.919
2	5	2	11.11 G	78.71 K	32.057	0.891	33.445	0.908
5	2	2	10.13 G	78.71 K	32.403	0.903	33.884	0.921
3	2	4	10.77 G	78.71 K	31.690	0.890	33.047	0.908
3	4	2	10.77 G	78.71 K	30.785	0.874	32.193	0.896
4	3	2	10.46 G	78.71 K	31.416	0.877	32.690	0.895
3	3	3	10.79 G	78.71 K	32.370	0.900	33.679	0.916

(b) AWGN Denoising ($\sigma^2 = 0.001$)

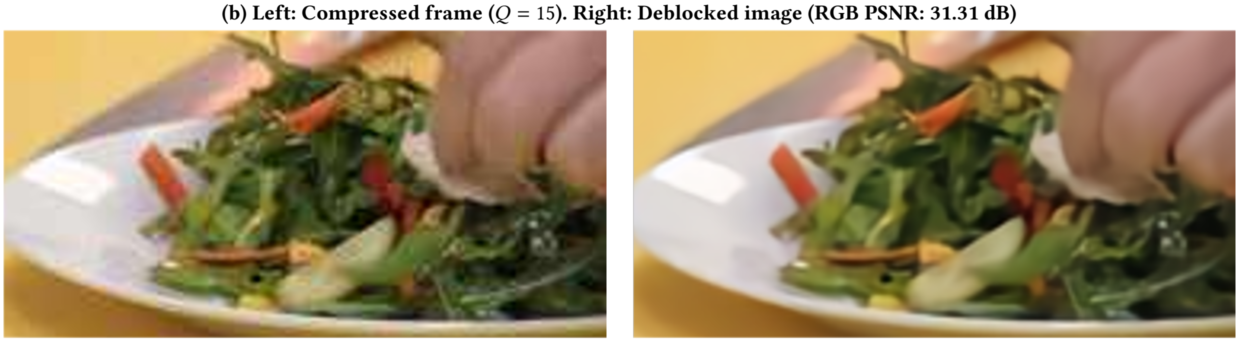
Module depth					RGB		Y-Channel	
N_A	N_D	N_F	MACs	# Params	PSNR	SSIM	PSNR	SSIM
1	1	7	11.50 G	78.91 K	37.071	0.961	37.742	0.965
1	7	1	11.50 G	78.91 K	37.136	0.961	37.774	0.965
7	1	1	9.52 G	78.91 K	37.176	0.961	37.868	0.965
2	2	5	11.17 G	78.91 K	37.072	0.961	37.740	0.965
2	5	2	11.17 G	78.91 K	37.102	0.961	37.776	0.965
5	2	2	10.18 G	78.91 K	37.196	0.961	37.855	0.965
3	2	4	10.84 G	78.91 K	37.227	0.962	37.902	0.965
3	4	2	10.84 G	78.91 K	37.071	0.961	37.740	0.965
4	3	2	10.51 G	78.91 K	37.173	0.961	37.877	0.965
3	3	3	10.84 G	78.91 K	37.378	0.962	38.002	0.966

(c) Super-resolution (2×)

Table 2: Effect of the depth of alignment, differential, and fusion modules in the EVRNet. Overall, EVRNet with deeper alignment modules provides the best trade-off between performance and number of multiplication-addition operations (MACs). In all these models, the depth of the network is fixed, i.e., $N_A + N_D + N_F = 9$.

(a) Original

(b) Left: Compressed frame (Q = 15). Right: Deblocked image (RGB PSNR: 31.31 dB)

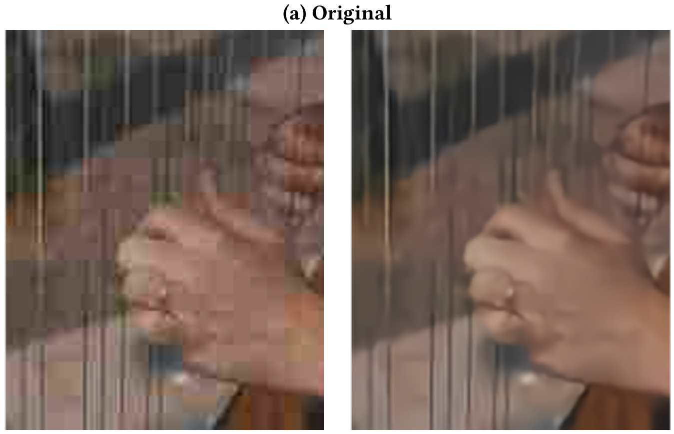


(c) Left: Compressed frame (Q = 45). Right: Deblocked image (RGB PSNR: 34.79 dB)

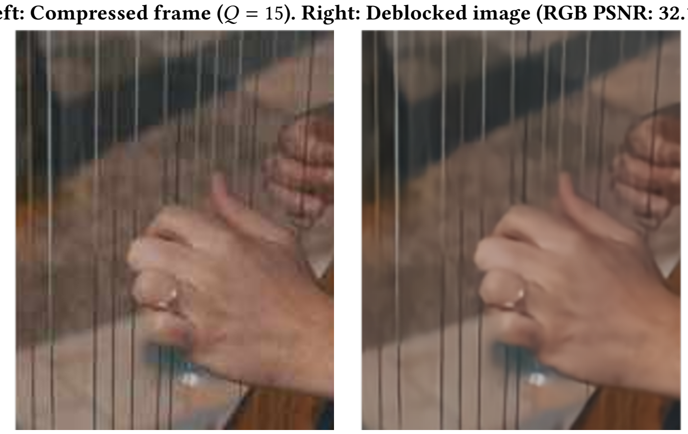
(d) Left: Compressed frame (Q = 75). Right: Deblocked image (RGB PSNR: 36.00 dB)

Figure 1: Deblocking example at different values of Q. Note that lower value of Q means higher compression.

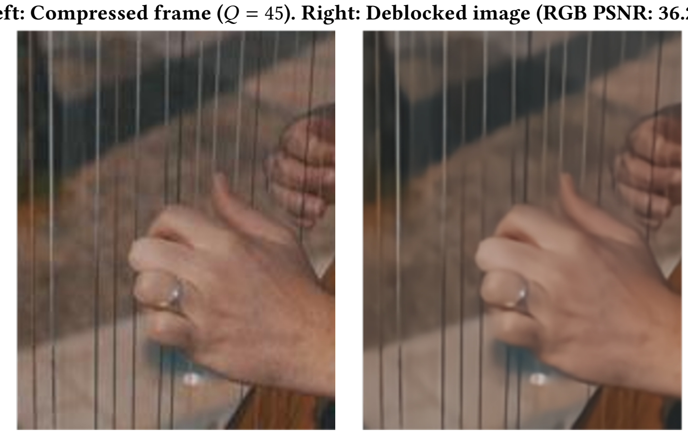
(a) Original



(b) Left: Compressed frame (Q = 15). Right: Deblocked image (RGB PSNR: 32.11 dB)

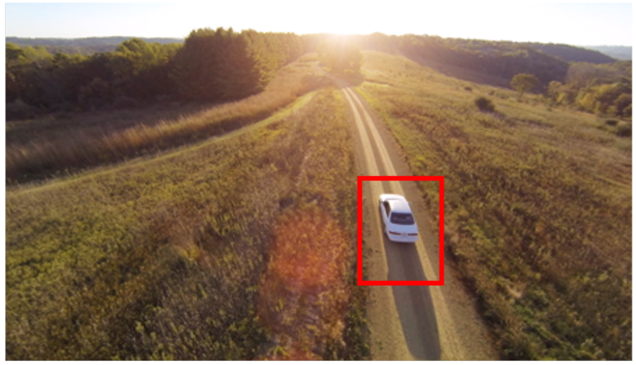


(c) Left: Compressed frame (Q = 45). Right: Deblocked image (RGB PSNR: 36.21 dB)

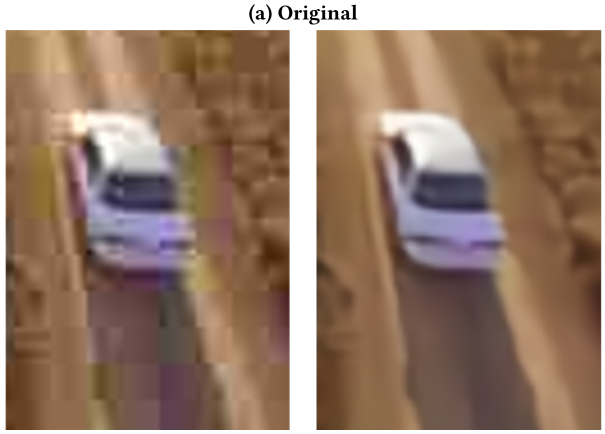


(d) Left: Compressed frame (Q = 75). Right: Deblocked image (RGB PSNR: 37.56 dB)

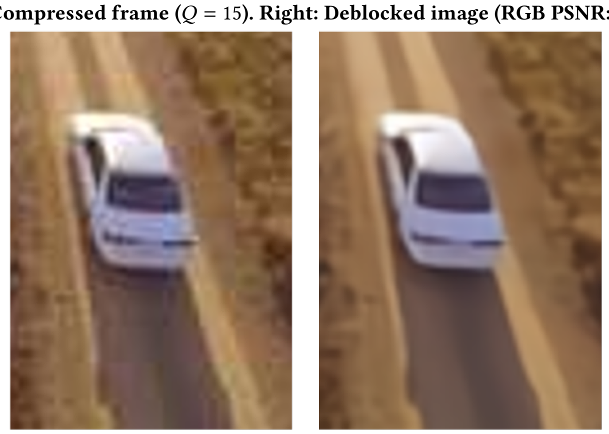
Figure 2: Deblocking example at different values of Q. Note that lower value of Q means higher compression.



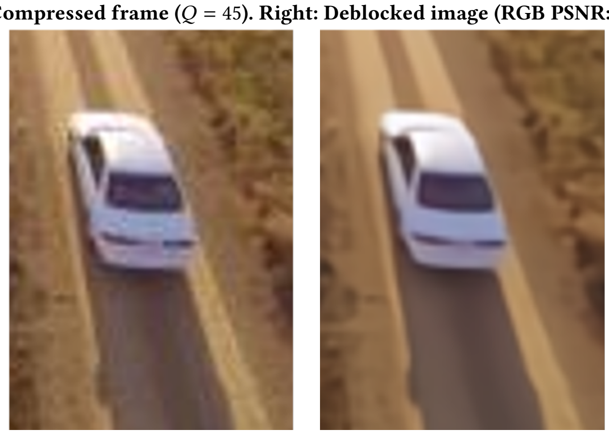
(a) Original



(b) Left: Compressed frame (Q = 15). Right: Deblocked image (RGB PSNR: 30.23 dB)

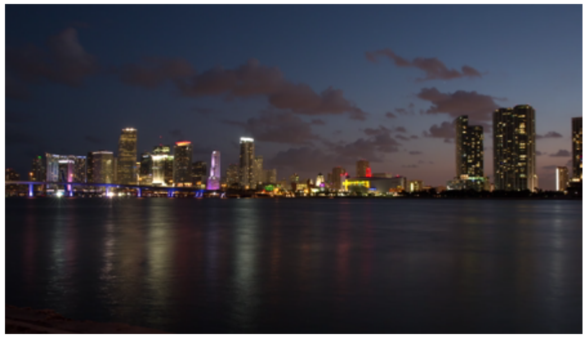


(c) Left: Compressed frame (Q = 45). Right: Deblocked image (RGB PSNR: 33.02 dB)



(d) Left: Compressed frame (Q = 75). Right: Deblocked image (RGB PSNR: 34.37 dB)

Figure 3: Deblocking example at different values of Q. Note that lower value of Q means higher compression.



(a) Original

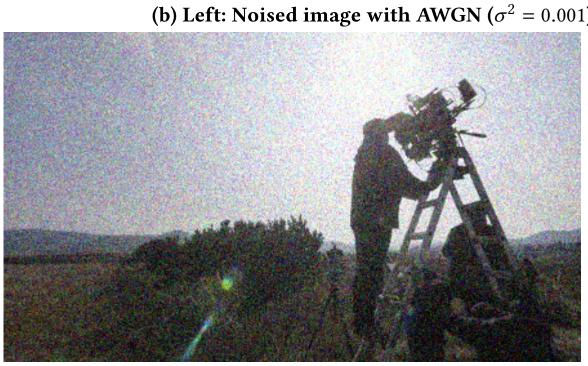
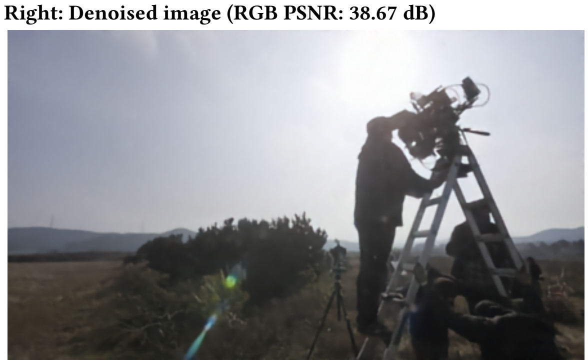
(b) Left: Noised image with AWGN ($\sigma^2 = 0.001$). Right: Denoised image (RGB PSNR: 34.57 dB)

(c) Left: Noised image with AWGN ($\sigma^2 = 0.01$). Right: Denoised image (RGB PSNR: 33.94 dB)

Figure 4: AWGN Denoising Example

(a) Original


(b) Left: Noised image with AWGN ($\sigma^2 = 0.001$). Right: Denoised image (RGB PSNR: 38.67 dB)

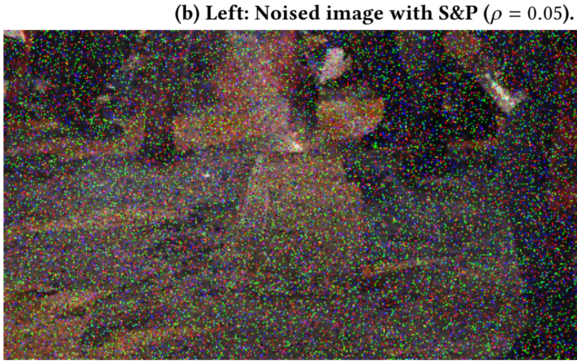



(c) Left: Noised image with AWGN ($\sigma^2 = 0.01$). Right: Denoised image (RGB PSNR: 37.11 dB)

Figure 5: AWGN Denoising Example

(a) Original

(b) Left: Noised image with S&P ($\rho = 0.05$). Right: Denoised image (RGB PSNR: 38.17 dB)



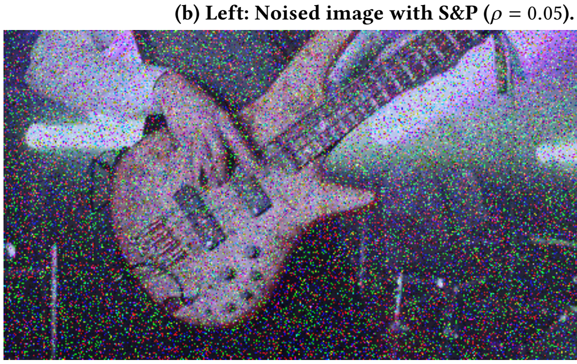
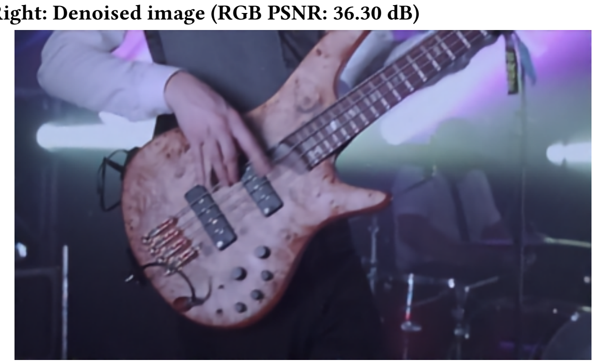
(c) Left: Noised image with S&P ($\rho = 0.15$). Right: Denoised image (RGB PSNR: 37.28 dB)

Figure 6: Salt & Pepper Denoising Example

(a) Original

(b) Left: Noised image with S&P ($\rho = 0.05$). Right: Denoised image (RGB PSNR: 36.30 dB)

(c) Left: Noised image with S&P ($\rho = 0.15$). Right: Denoised image (RGB PSNR: 35.50 dB)

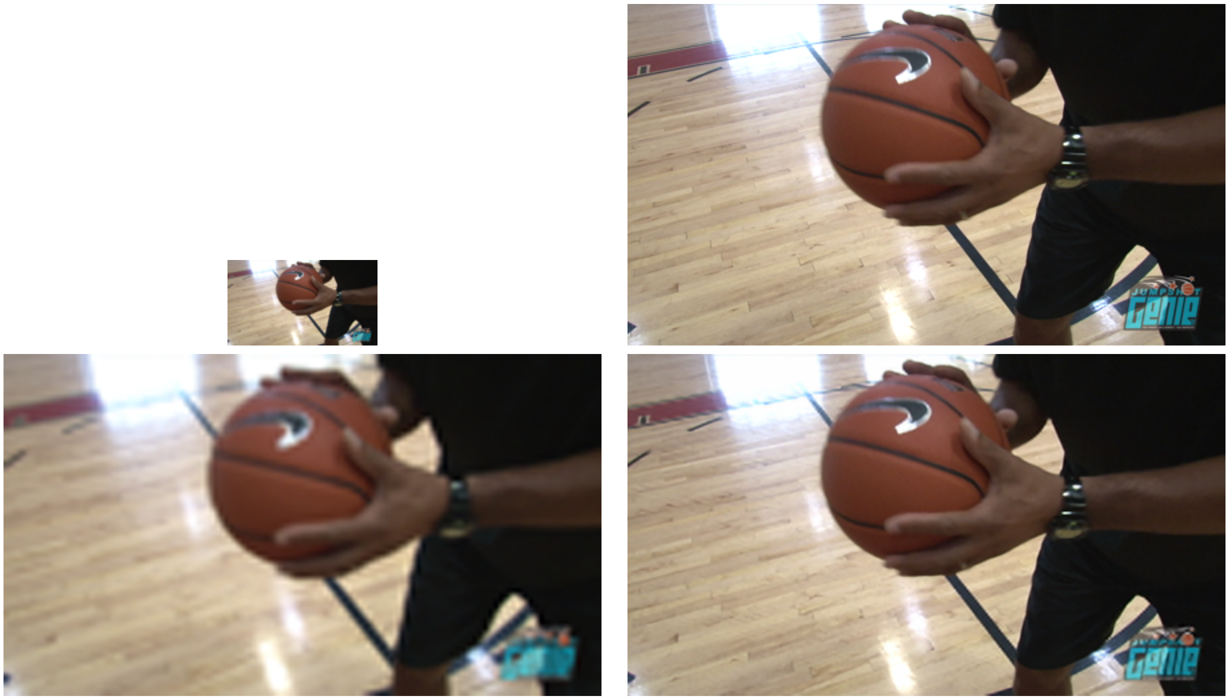
Figure 7: Salt & Pepper Denoising Example

(a) Original images

(b) Noised images with AWGN ($\sigma^2 = 0.001$) and S&P ($\rho = 0.1$)

(c) Denoised images

Figure 8: Denoising example with mixed noise

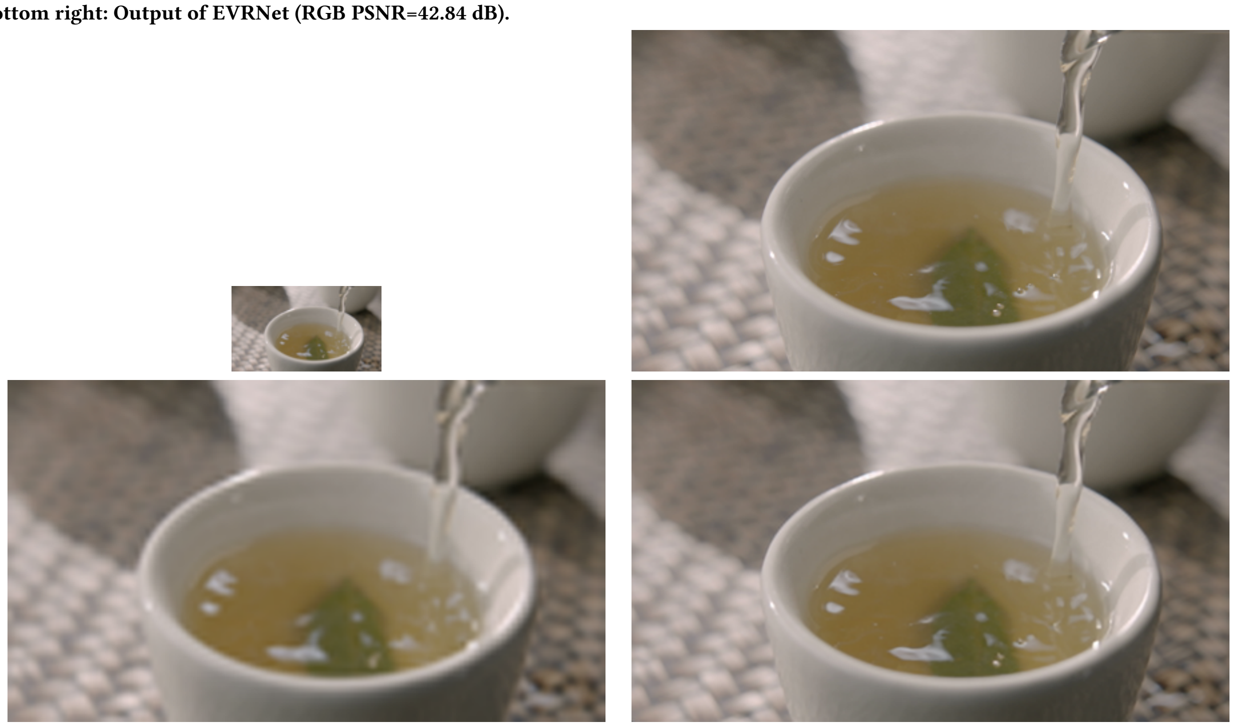


(a) Top left: Input low-resolution frame. Top right: Ground truth. Bottom left: Output of bicubic up-sampling (RGB PSNR: 28.59 dB) Bottom right: Output of EVRNet (RGB PSNR=34.76 dB).

(b) Top left: Input low-resolution frame. Top right: Ground truth. Bottom left: Output of bicubic up-sampling (RGB PSNR: 27.56 dB) Bottom right: Output of EVRNet (RGB PSNR=38.41 dB).

Figure 9: $4\times$ **Video super-resolution examples.**

(a) Top left: Input low-resolution frame. Top right: Ground truth. Bottom left: Output of bicubic up-sampling (RGB PSNR: 36.84 dB) Bottom right: Output of EVRNet (RGB PSNR=42.84 dB).



(b) Top left: Input low-resolution frame. Top right: Ground truth. Bottom left: Output of bicubic up-sampling (RGB PSNR: 36.21 dB) Bottom right: Output of EVRNet (RGB PSNR=43.97 dB).

Figure 10: $4\times$ **Video super-resolution examples.**