Supplementary Materials for 3RE-Net: Joint Loss-REcovery and Super-REsolution Neural Network for REal-time Video

1 More Experiment Results with Confidence Intervals.

Table. 1 shows an extended version of Table. 1 in the main paper using Vid4 Dataset, where the 95% confidence intervals are also shown. Our 3RE-Net improves the quality of the frame significantly (in terms of PSNR and SSIM) and outperforms all benchmarks.

Schemes	PSNR	SSIM
RAW (Bicubic)	17.7606 ±0.1461	0.4889 ±0.0062
(B1) DAIN + RRN	18.3051 ±0.1383	0.4712 ±0.0067
(B2) RRN + DAIN	18.0939 ±0.1631	0.5000 ±0.0046
(B3) DAIN + RBPN	18.8285 ±0.1344	0.5141 ±0.0079
(B4) RBPN + DAIN	18.8222 ±0.1178	0.5442 ±0.0051
(B5) BMBC + RRN	18.0185 ±0.1212	0.5093 ±0.0061
(B6) RRN + BMBC	19.4391 ±0.1446	0.5688 ±0.0062
(B7) BMBC + RBPN	18.6883 ±0.1194	0.5370 ±0.0059
(B8) RBPN + BMBC	19.2415 ±0.1233	0.5477 ±0.0044
(B9) RRN + DFC-Net	18.1021 ±0.1426	0.5074 ±0.0063
(B10) RBPN + DFC-Net	18.6556 ±0.1300	0.5318 ±0.0042
(B11) RRN + VI-Net	19.5097 ±0.1416	0.5460 ±0.0067
(B12) RBPN + VI-Net	19.5979 ±0.1776	0.5376 ±0.0058
3RE-Net (ours)	**21.2680 ±0.1238**	**0.6435 ±0.0060**

Table 1: Average PSNR (in dB) and SSIM using Vid4 dataset with 95% confidence intervals.

2 Ablation Studies

We conduct a series of ablation studies to analyze the effectiveness of each component in the 3RE-Net. Unless otherwise mentioned, we use the same training and testing settings introduced in the experiments section.

2.1 Choice of the Number of Supporting Frames

We alter N (number of supporting frames) and test the performance. With $N \in \{2, 3, 4\}$, we re-train the motion synthesis network while keeping the rest of the pipeline unchanged. We then evaluate their delay and PSNR. Results are shown in Fig. 1.

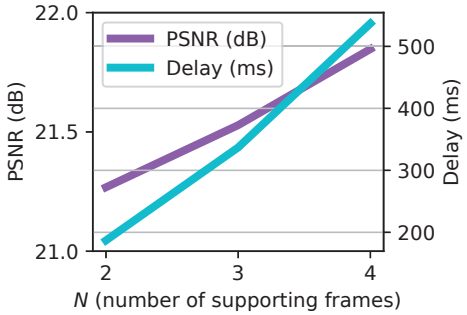

Fig. 1: Delay and PSNR on different number of N.

We observe that larger N brings higher quality, but is more time-consuming. By increasing N from 2 to 3, we sacrifice 150 ms delay while the PSNR gain is 0.25 dB. By further increasing N to 4, we sacrifice 350 ms delay while the PSNR gain is 0.8 dB. Under the current hardware (NVIDIA GeForce RTX 3090 GPU), $N = 2$ achieves desired high quality with a reasonably small delay. However, in the future, as the improvement of computational capabilities of the hardware, larger N will also be useful for realizing better quality. This is the reason that we do not limit $N = 2$ throughout this paper.

2.2 Motion Group Options in Step A.

To demonstrate the effectiveness of the two groups of motions in Step A (motion extraction and prediction), we also experiment with only one group of motions. We re-train the network using the same warping and synthesis modules but with less candidates. The results are shown in Table 2. It is demonstrated that using motion groups $\mathbb{V}$ and $\mathbb{U}$ outperforms using one motion group only.

Motion Group(s)	PSNR (dB)
$\mathbb{V}$ only	20.24
$\mathbb{U}$ only	20.52
$\mathbb{V}$ and $\mathbb{U}$ **(used by 3RE-Net)**	**21.27**

Table 2: Ablation study on motion groups selection in Step A.

2.3 Kernel Size in Step D.

To demonstrate the effectiveness of kernel size selection in the motion synthesis network (Step D), we train the network using various kernel sizes. The results are shown in Table. 3. It is demonstrated that 5×5 outperforms other kernel sizes, while other kernel sizes also yield acceptable performance.

Kernel size	PSNR
3×3	21.19
5×5 **(used by 3RE-Net)**	**21.27**
7×7	21.24

Table 3: Ablation study on kernel size in Step D.

2.4 Alignment Options in Step E.

To demonstrate the effectiveness of the feature-level alignment in Step E, we also experiment with image-level alignment and no alignment. The results are shown in Table 4. In these three experiments, feature-level alignment outperforms image-level alignment and no alignment.

Methods	PSNR (dB)
No alignment	20.08
Image-level alignment	21.02
Feature-level alignment **(used by 3RE-Net)**	**21.27**

Table 4: Ablation study on alignment methods in Step E.

3 Inference Delay Breakdown

We measure the inference delay breakdown of all modules in 3RE-Net. The results are shown in Fig. 2. The results are averaged over 1000 times of experiments using Nvidia GeForce RTX 3090 GPU.

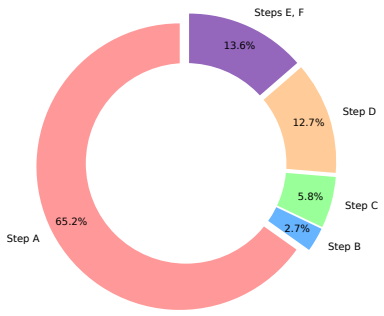

Fig. 2: Delay breakdown of all modules (Step A – F) in 3RE-Net .

Among all modules in 3RE-Net, Step A (motion extraction and prediction) causes the largest delay (65.2% of the overall delay). This is because the motion extraction and prediction module performs pixel-wise estimation among all frames, and outputs a large number of motions. However, this will benefit all subsequent modules, and the delays caused by Steps B – F are much smaller. In sum, the overall delay of all components (187 ms) is much improved compared with other benchmarks.

4 More Visualized Results.

Fig. 3 shows the reconstructed loss region under all schemes, supplementing Fig. 5 in the main paper. Our 3RE-Net improves the quality of the frame significantly and outperforms all benchmarks.

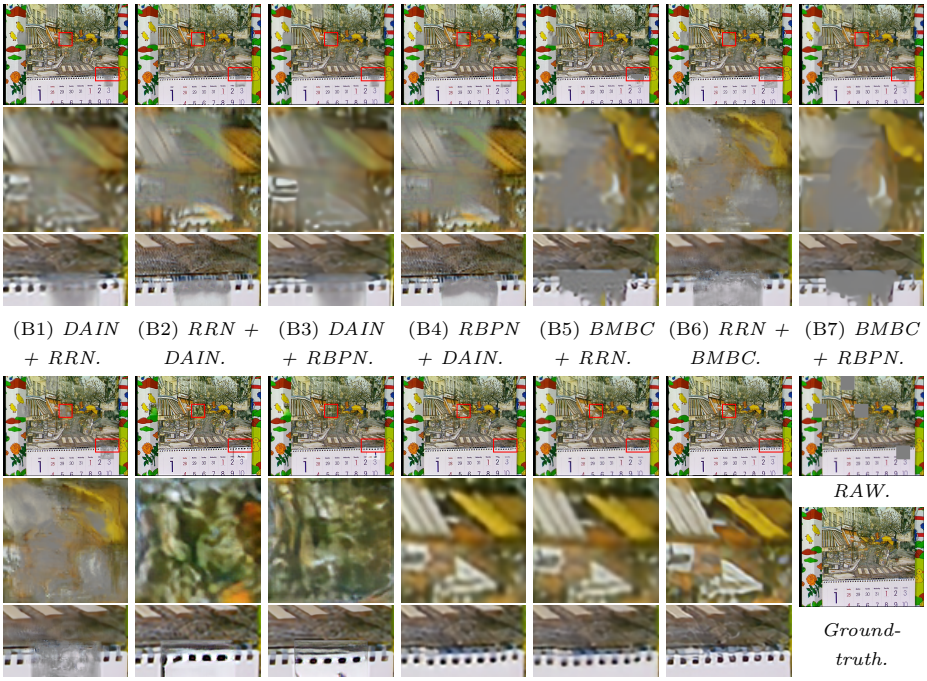

Fig. 3: Qualitative comparison of the reconstructed loss region. Two regions bounded in red are zoomed in. The second row shows one of the reconstructed regions. The third row shows the border between the damaged and undamaged parts.