

Supplementary Material

1 SUPPLEMENTARY TABLES AND FIGURES

1.1 Figures



Figure S1. Average PSNR (dB) on the testing dataset case 1 as a function of the number of input LR frames (A), the attention head spatial window size (B), the number of videos used for training relative to the total number of training videos (C), and the model size and architecture parameters (D).



Figure S2. Average PSNR (dB) on the testing dataset case 2 as a function of the number of input LR frames (A), the attention head spatial window size (B), the number of videos used for training relative to the total number of training videos (C), and the model size and architecture parameters (D).



Figure S3. Reconstructed HR frame PSNR (dB) (row 1), AAD (row 2), and SSIM (row 3) each as a function of the LR image PSNR that was used to generate the Poisson noise in the input LR frames, based on 3, 5, 7, and 9 input LR frames. Compared to the LR image sequence, the input HR image sequence was temporally down-sampled by a factor of 2. At a given number of input LR frames, the model was trained with the PyTorch DDP on 24 nodes to normalize potential variation in the training. Results were evaluated on case 1 (column A) and case 2 (column B) and presented as box plots.



Figure S4. Reconstructed HR frame PSNR (dB) (row 1), AAD (row 2), and SSIM (row 3) each as a function of the LR image PSNR that was used to generate the Poisson noise in the input LR frames, based on 3, 5, 7, and 9 input LR frames. Compared to the LR image sequence, the input HR image sequence was temporally down-sampled by a factor of 20. At a given number of input LR frames, the model was trained with the PyTorch DDP on 24 nodes to normalize potential variation in the training. Results were evaluated on case 1 (column A) and case 2 (column B) and presented as box plots.