SUPPLEMENTARY MATERIAL
S3VD SELF-SUPERVISED SPATIAL DOWNSAMPLING LOSS: A METHOD FOR TRAINING VIDEO FPN DENOISING NETWORKS

Author(s) Name(s)

Author Affiliation(s)

1. ARCHITECTURE OF THE UNET

The Unet takes 5 frames which are concatenated at the input and outputs the FPN. It encompasses 4 scales. At each scale in the encoder and decoder a residual block is applied four times. Residual block consists of two 3×3 conv kernels and a ReLU activation between them. The downsampling is implemented using strided convolution and the upsampling transposed convolution. The bottleneck at the coarsest resolution (the input size divided by 16) consists of another residual block repeated four times. The input and output layers are simply 3×3 convolutions. In the skipped connection between the encoder and the decoder, the features are added and fed to the transposed convolution, which will upsample them and divide the channel dimension by a factor of 2.

2. EXPERIMENT ON REAL DATA

For the experiment on real data, we use the multi-view infrared dataset [1] that contains real FPN. The supervised Unet is trained with synthetic FPN on a grayscale version of the REDS dataset temporally downscaled by a factor of three. To estimate the noise to have a fair comparison as possible, we estimated the noise level from the unsupervised methods. UDVD [2] cannot remove spatially correlated FPN but it can remove unstructured one. We can approximate the standard deviation of the spatially uncorelated FPN by its estimation from UDVD. For the vertical / horizontal stripes FPN, we compute the FPN with our method in the multi-view setting, then average the estimated FPN along row / column.

3. REFERENCES

[1] Arnaud Barral, Pablo Arias, and Axel Davy, "Fixed pattern noise removal for multi-view single-sensor infrared camera," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2024, pp. 1669–1678.

[2] Dev Yashpal Sheth, Sreyas Mohan, Joshua Vincent, Ramon Manzorro, Peter A. Crozier, Mitesh M. Khapra, Eero P. Simoncelli, and Carlos Fernandez-Granda, "Unsupervised deep video denoising," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021.