Supplementary Material *for* CVF-SID: Cyclic multi-Variate Function for Self-Supervised Image Denoising by Disentangling Noise from Image

Reyhaneh Neshatavar[1*] Mohsen Yavartanoo[1*] Sanghyun Son[1] Kyoung Mu Lee[1,2]
[1]Dept. of ECE & ASRI, [2]IPAI, Seoul National University, Seoul, Korea
{reyhanehneshat,myavartanoo,thstkdgus35,kyoungmu}@snu.ac.kr

S1. Training on a single image

We test our proposed CVF-SID on a practical case that uses only a single noisy image. Specifically, we train our method in a self-supervised manner and apply it to a real-world input. Figure S1 demonstrates that our CVF-SID can learn to denoise without any other external examples. During the training, we randomly crop patches to construct a mini-batch as we describe in Section 4.2 in our main manuscript. The denoising result shows that the proposed approach does not rely on a large-scale dataset but can be learned to denoise from a single image.

S2. Network architecture details

Our clean image generator consists of 16 sequentially 3×3 convolutional layers with the sizes of 64 and the padding size 1. Each convolution layer is followed by ReLU non-linear activation function. Finally, a 1×1 convolutional layer generates the RGB clean image.

Our noise generator includes ten 3×3 convolutional layers with the sizes of 64 and padding size 1, followed by the ReLU activation function. Then each branch of signal-dependent and signal-independent noise generators containing three 3×3 convolutional layers with ReLU activation function and one 1×1 convolutional layers is applied separately to generate the signal-dependent and signal-independent noise maps $\hat{N}_d$ and $\hat{N}_i$, respectively.

All convolutional weights and biases are initialized with Xavier uniform and constant 0, respectively. To reduce the padding effects, we apply reflection padding of size 20 to each side of an input image and crop the output image to obtain the image with the original size. Moreover, to guarantee that the noise maps are zeros-mean, we subtract them with their channel-wise average.

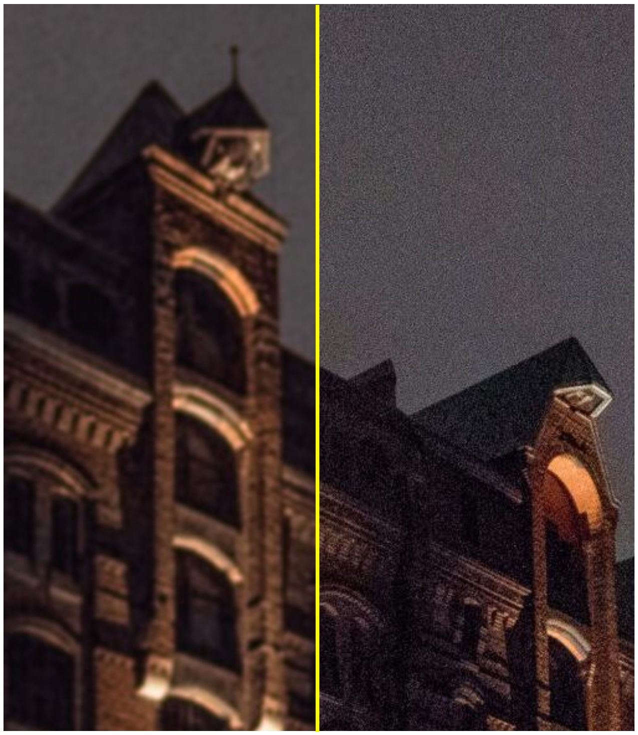
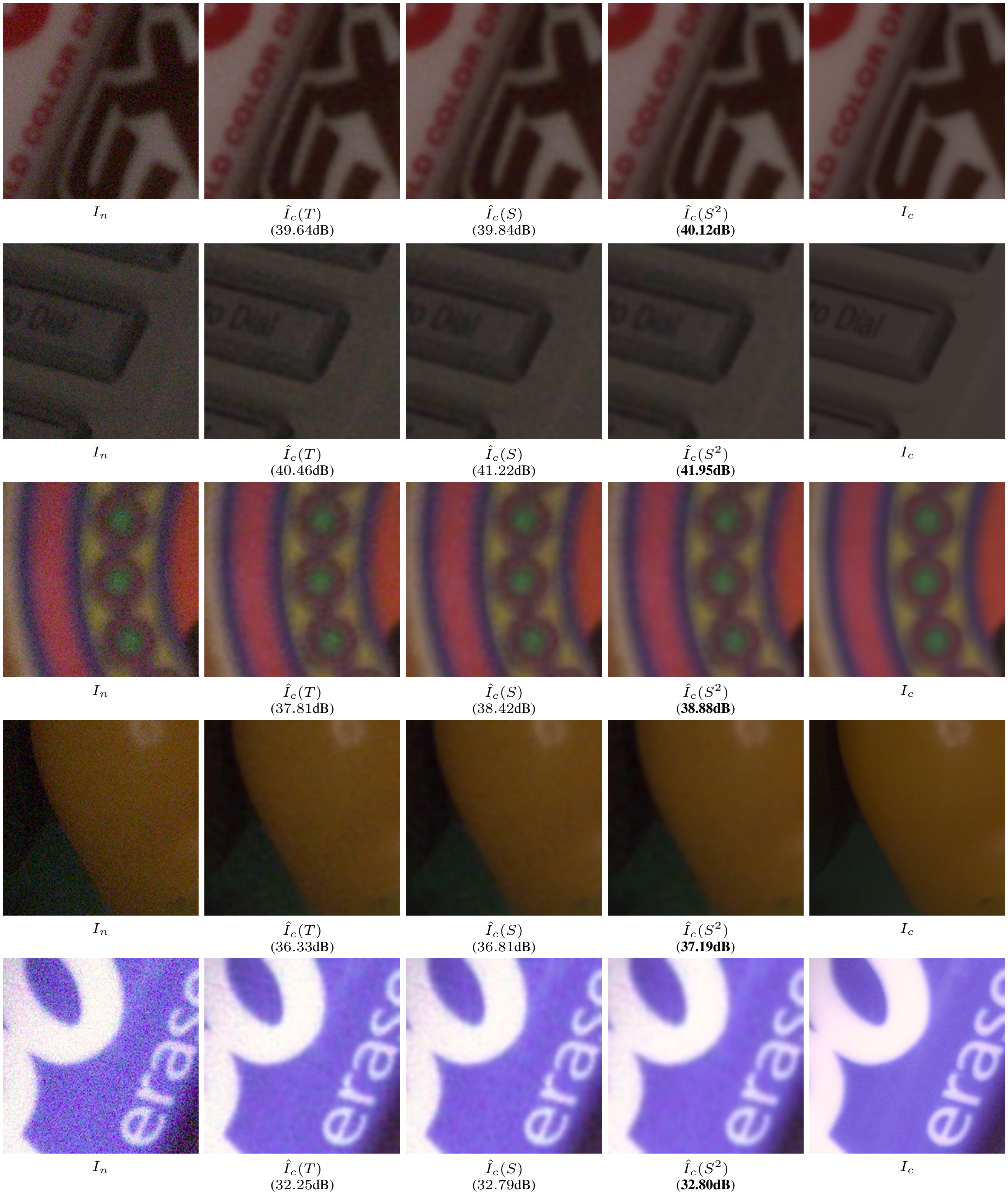


Figure S1. **Restored clean image with only a single noisy image.**

S3. Training schemes

Figure S2 provides visual comparisons between different training schemes **T**, **S**, and $\mathbf{S}^2$ used to train our CVF-SID method. Results demonstrate that our method achieves slightly better results with training scheme **S** than **T** for all images because the discrepancy between training and test images can be minimized as discussed in Section 4.3 in our main manuscript. Furthermore, we show that in some cases, the training scheme **S** is sufficient to remove the noise, and we do not need to double-train model ($\mathbf{S}^2$) to further achieve better performance, as shown in fifth row of Figure S2.

[*] equal contribution



I_n $\hat{I}_c(T)$ (39.64dB) $\hat{I}_c(S)$ (39.84dB) $\hat{I}_c(S^2)$ (**40.12dB**) I_c

I_n $\hat{I}_c(T)$ (40.46dB) $\hat{I}_c(S)$ (41.22dB) $\hat{I}_c(S^2)$ (**41.95dB**) I_c

I_n $\hat{I}_c(T)$ (37.81dB) $\hat{I}_c(S)$ (38.42dB) $\hat{I}_c(S^2)$ (**38.88dB**) I_c

I_n $\hat{I}_c(T)$ (36.33dB) $\hat{I}_c(S)$ (36.81dB) $\hat{I}_c(S^2)$ (**37.19dB**) I_c

I_n $\hat{I}_c(T)$ (32.25dB) $\hat{I}_c(S)$ (32.79dB) $\hat{I}_c(S^2)$ (**32.80dB**) I_c

Figure S2. **Visual comparisons of the predicted clean images from the SIDD validation dataset between different training schemes (T, S, and S^2).** We also provide PSNR w.r.t. ground-truth images.

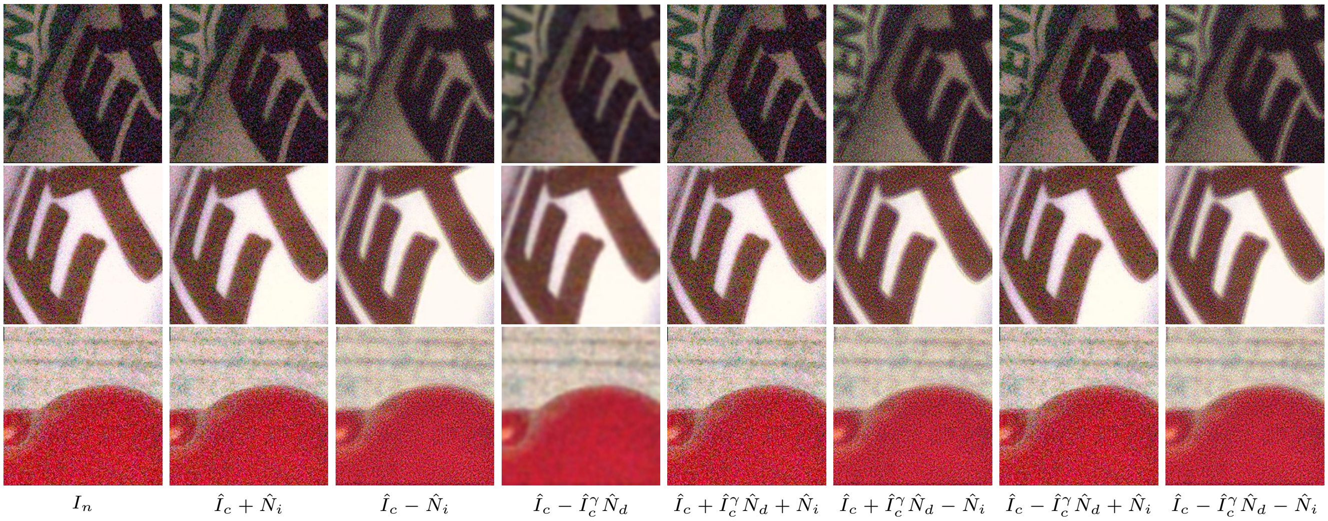

I_n
$\hat{I}_c + \hat{N}_i$
$\hat{I}_c - \hat{N}_i$
$\hat{I}_c - \hat{I}_c^\gamma \hat{N}_d$
$\hat{I}_c + \hat{I}_c^\gamma \hat{N}_d + \hat{N}_i$
$\hat{I}_c + \hat{I}_c^\gamma \hat{N}_d - \hat{N}_i$
$\hat{I}_c - \hat{I}_c^\gamma \hat{N}_d + \hat{N}_i$
$\hat{I}_c - \hat{I}_c^\gamma \hat{N}_d - \hat{N}_i$

Figure S3. **Generated synthetic noisy images by our proposed self-supervised augmentation strategy on the SIDD validation dataset.**

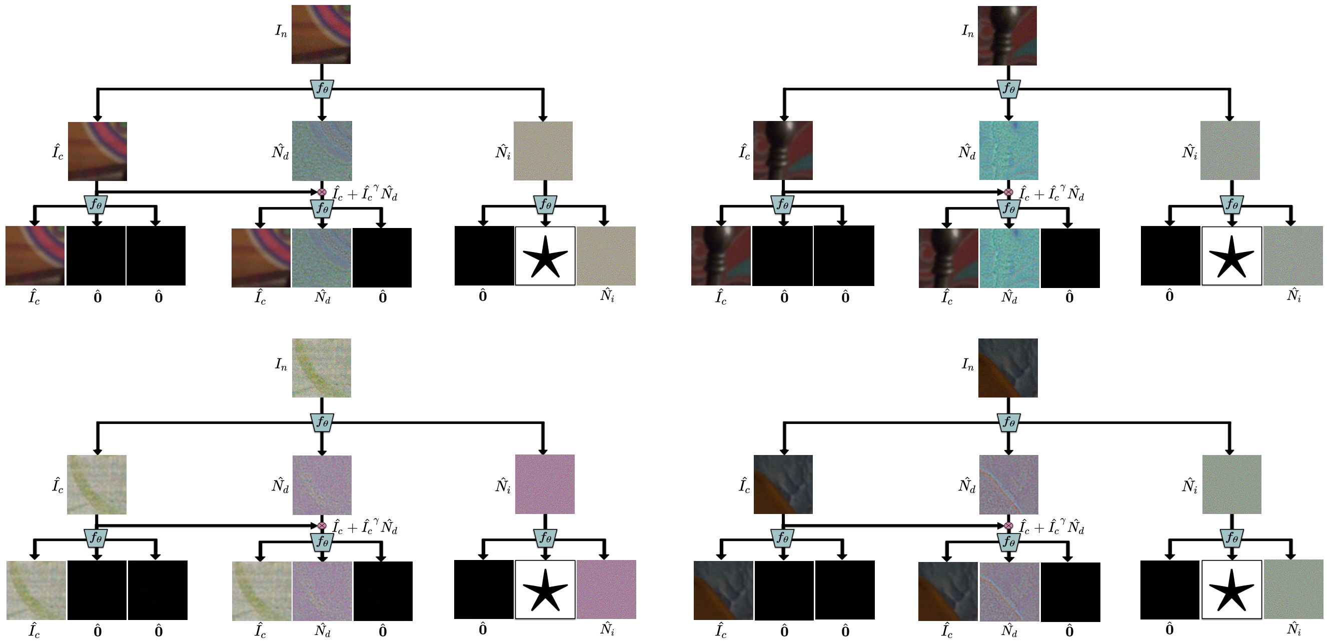


Figure S4. **Decomposition results of our CVF-SID method on the SIDD validation dataset with training scheme (S).**

S4. Cascading additional models

We further apply three ($\mathbf{S}^3$) and four ($\mathbf{S}^4$) CVF-SID, successively, to the original noisy images from the SIDD validation dataset. The results show the improvement of model performance to some extent but saturate at $\mathbf{S}^3$ as follows:

Method	**S**	**S²**	**S³**	**S⁴**
PSNR/SSIM	34.67/0.943	34.81/**0.944**	**34.84**/0.943	**34.84**/0.942

Table S1. **Quantitative comparison on the number of cascading models on sRGB images in the SIDD validation dataset.**

S5. Augmentations

In Figure S3, we visualize various synthesized images from our self-supervised augmentation strategy described in Section 3.2 in our main manuscript. We note that our augmentation strategy can generate several real-world noisy images from only a single noisy image without requiring additional information.

We also analyze the effect of the hyperparameter λ_{aug} on the SIDD validation dataset as follows:

λ_{aug}	0	0.01	0.1	1
PSNR/SSIM	34.43/0.942	34.53/0.942	**34.67/0.943**	34.55/0.936

Table S2. **Quantitative comparison on hyperparameter $\lambda_{\mathbf{aug}}$ on sRGB images in the SIDD validation dataset.**

S6. Decomposition results

Figure S4 shows how our CVF-SID can effectively disentangle the clean image, signal-dependent, and signal-independent noises from a noisy input image. The results demonstrate that CVF-SID is successfully learned to satisfy our constraints on its outputs, which are described in Section 3.2. For example, when we feed the initially predicted noise-free image $\hat{I}_c$ again to the network f_θ, we can get the same clean image for $f_\theta^{\text{clean}}(\hat{I}_c)$ and zeros for the corresponding noise maps. One limitation of our method is that the predicted clean image $\hat{I}_c$ and the signal-dependent noise map $\hat{N}_d$ are not completely independent, which is opposed to our assumption mentioned in Section 3.2. This contradiction can be due to considering a fixed correlation parameter γ, which may vary per image for a real-world scenario.