Basis Prediction Networks for Effective Burst Denoising with Large Kernels: Supplementary Material

Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, Ayan Chakrabarti

A. Architecture

We now provide a detailed description of the architecture of our basis prediction network, which includes a shared encoder and two decoders. The shared encoder network consists of five down-sampling blocks and encodes noisy frames into a shared feature space. The coefficients decoder decodes these features into a set of per-pixel mixing coefficients. Coefficients for each output pixel are normalized with a softmax function separately. The basis decoder first reduces these features to a 1D vector, and then decodes them to an output of shape $K \times K \times TB$, representing a burst-specific set of B basis kernels, each of shape $K \times K \times T$ ($K = 15$ and $B = 90$ for our model). Each 3D basis element is normalized with a softmax function so that the basis kernel sums up to 1. We include regular skip connections from the encoder to the coefficient decoder, and pooled-skip connections to the basis decoder. The entire architecture is illustrated in Figure 6. (For color burst denoising, the decoder output has shape $K \times K \times 3TB$, representing a set of B basis kernels, with each 4D basis element being $K \times K \times 3 \times T$.)

In our ablation study, we considered versions of our network that produced smaller kernels with $K = 5$ and $K = 9$. For $K = 5$, we used one less up-sampling block in the basis decoder than for the $K = 15$ case shown in Figure 6, and added one $3{\times}3$ convolutional layer with valid padding before the layer $\star$ in Figure 6. For $K = 9$, we also used one less up-sampling block in the basis decoder, but in this case, replaced the layer $\star$ in with a $2{\times}2$ transpose convolution layer with valid padding (i.e., one that does a $2{\times}2$ "full" convolution).

While previous works [30, 8] learns a global set of kernel for all images and apply light-weight local processing to hard select a kernel for each pixel, our method is more powerful as it can output any kernel of larger size by predicting per-burst basis and per-pixel coefficients.

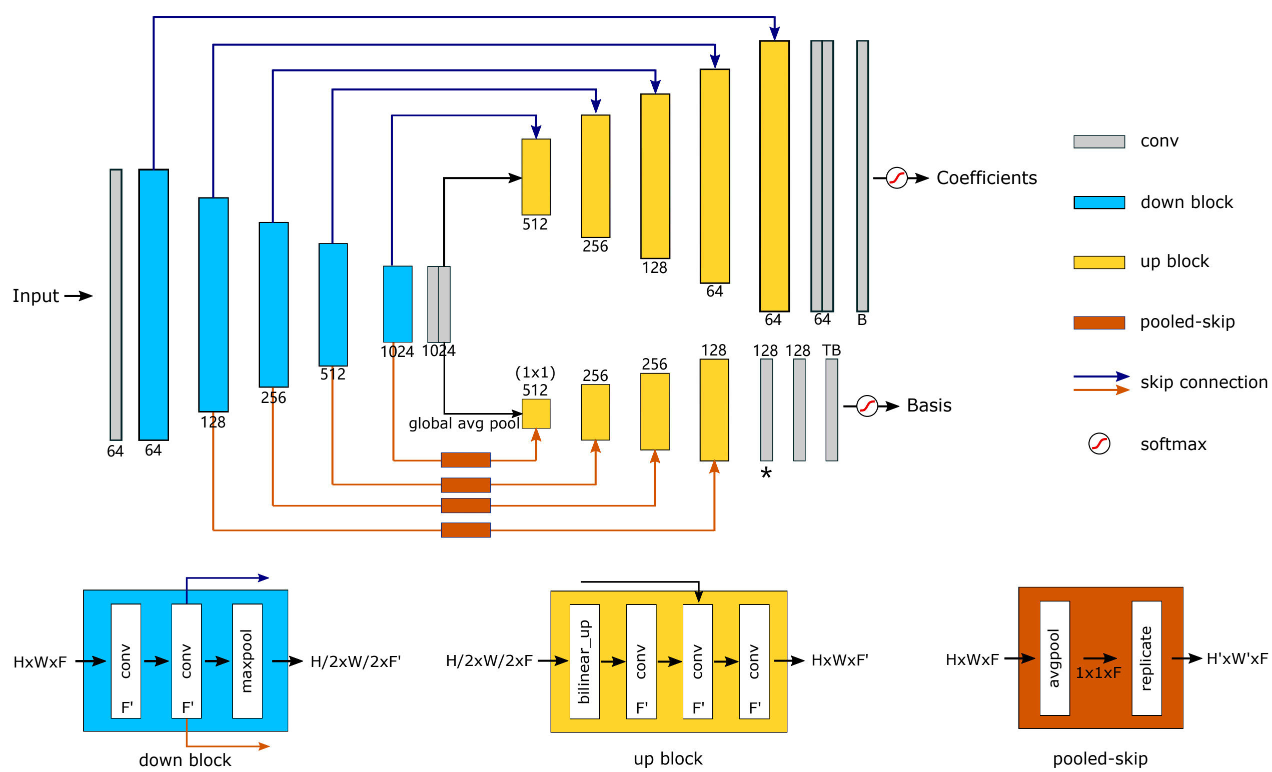

Figure 6: Details for our basis prediction network. All convolutional layers are 3x3 convolutional layers with same padding, except $\star$ which is a 2x2 convolutional layer with valid padding. The down-sampling block reduces the spatial size of the input by 2 with a 2x2 stride 2 max-pooling layer. The up-sampling block has a bilinear up-sampling layer with a factor of 2. The pooled skip connection first applies a global spatial average pooling of the encoder's activations and replicate the average vector to the resolution of the decoder layer.

Our fixed basis ablation replaced the entire decoder branch with just a learned tensor, of size $K \times K \times (TB)$, to serve as the basis. However, we still retain the encoder and coefficient decoder, and the weights of these networks are learned jointly with the fixed basis tensor. Note that in this case, the denoising kernels at each pixel are formed as a linear combination of this fixed set of basis kernels, based on the coefficients predicted from the input burst by the decoder at each location (this is different from approaches that select one kernel at each location from a fixed kernel set [30, 8]). As our ablation showed, having a fixed basis set yields worse denoising performance than an adaptive basis.

B. Additional Results

We show additional denoising results below: for color burst denoising in Fig. 7 and for grayscale denoising in Figs. 8-11.

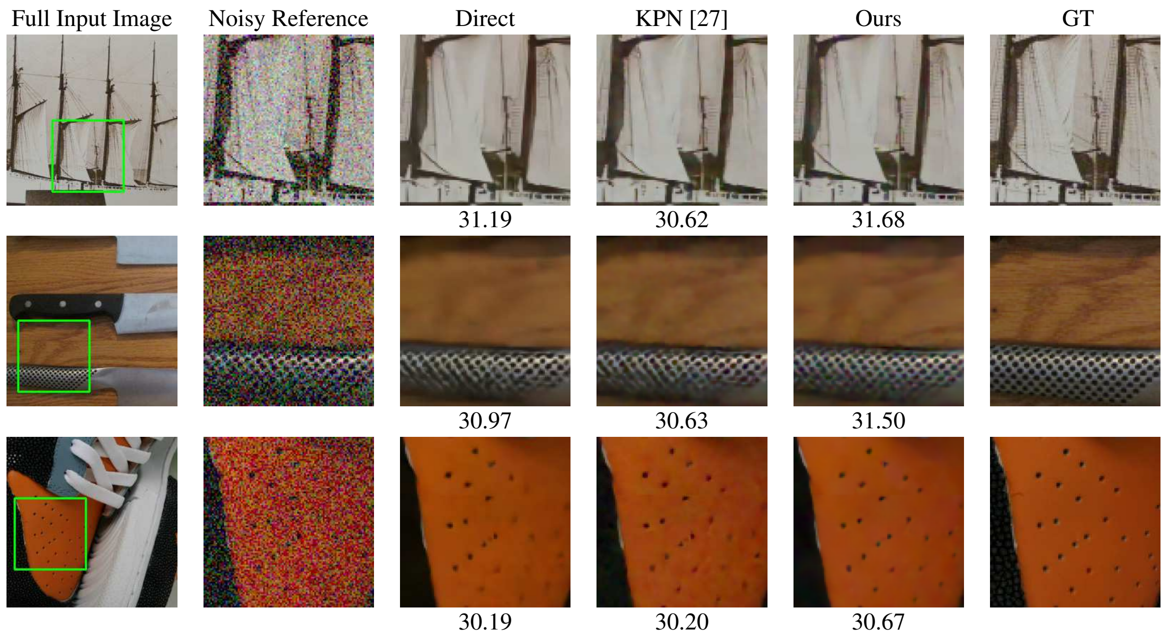

Figure 7: Additional color results on our synthetic color test set. Numbers refer to PSNR (dB) on the full image.

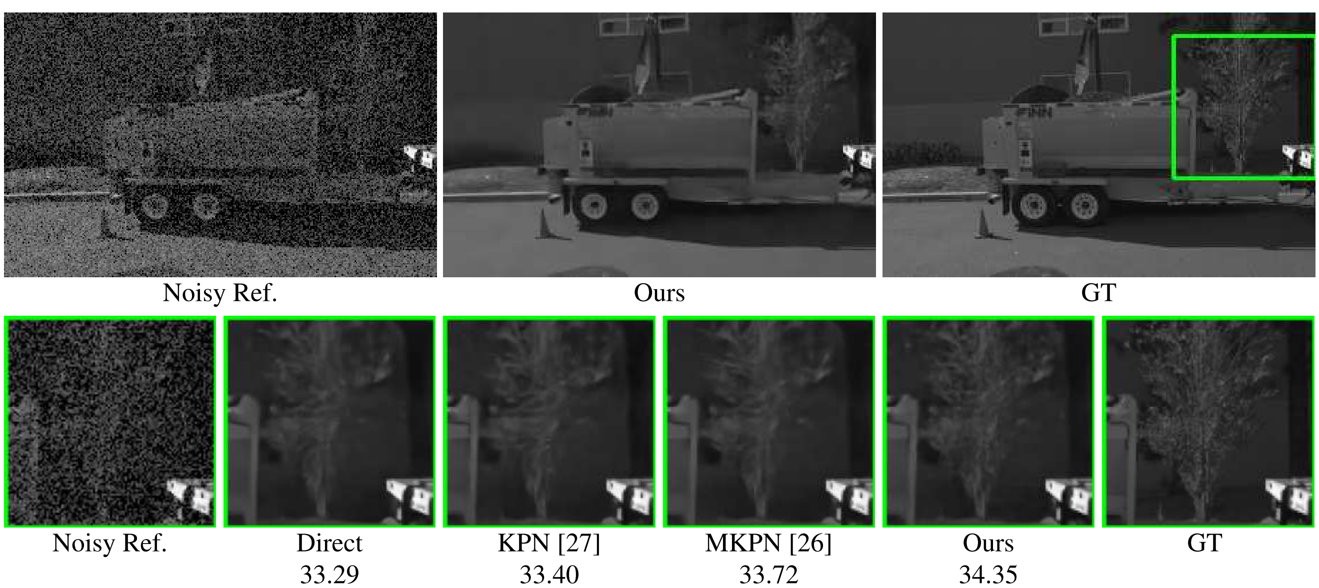

Figure 8: Additional grayscale results on benchmark test set [27]. Numbers refer to PSNR (dB) on the full image.

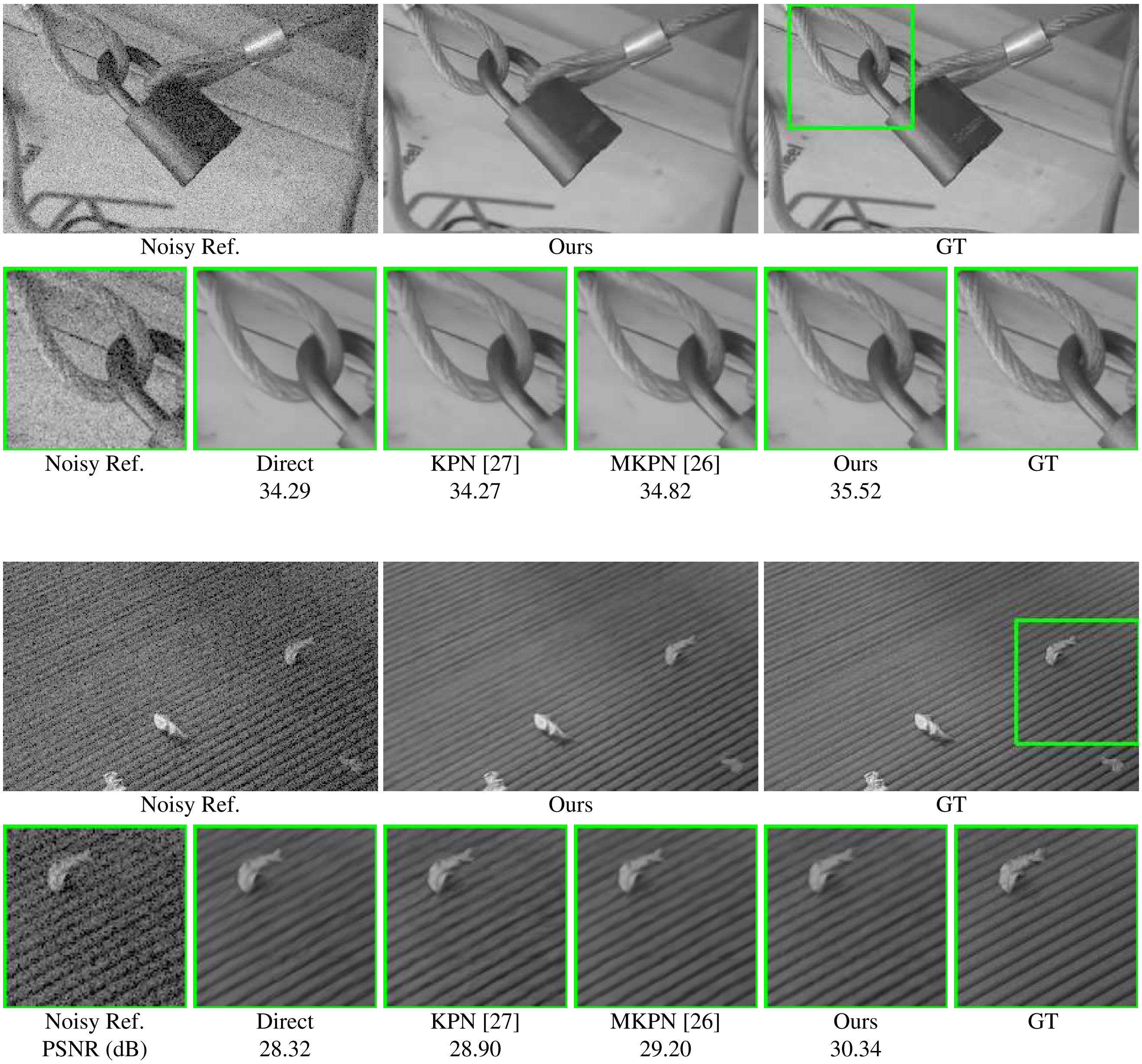

Figure 9: Additional grayscale results on benchmark test set [27]. Numbers refer to PSNR (dB) on the full image.

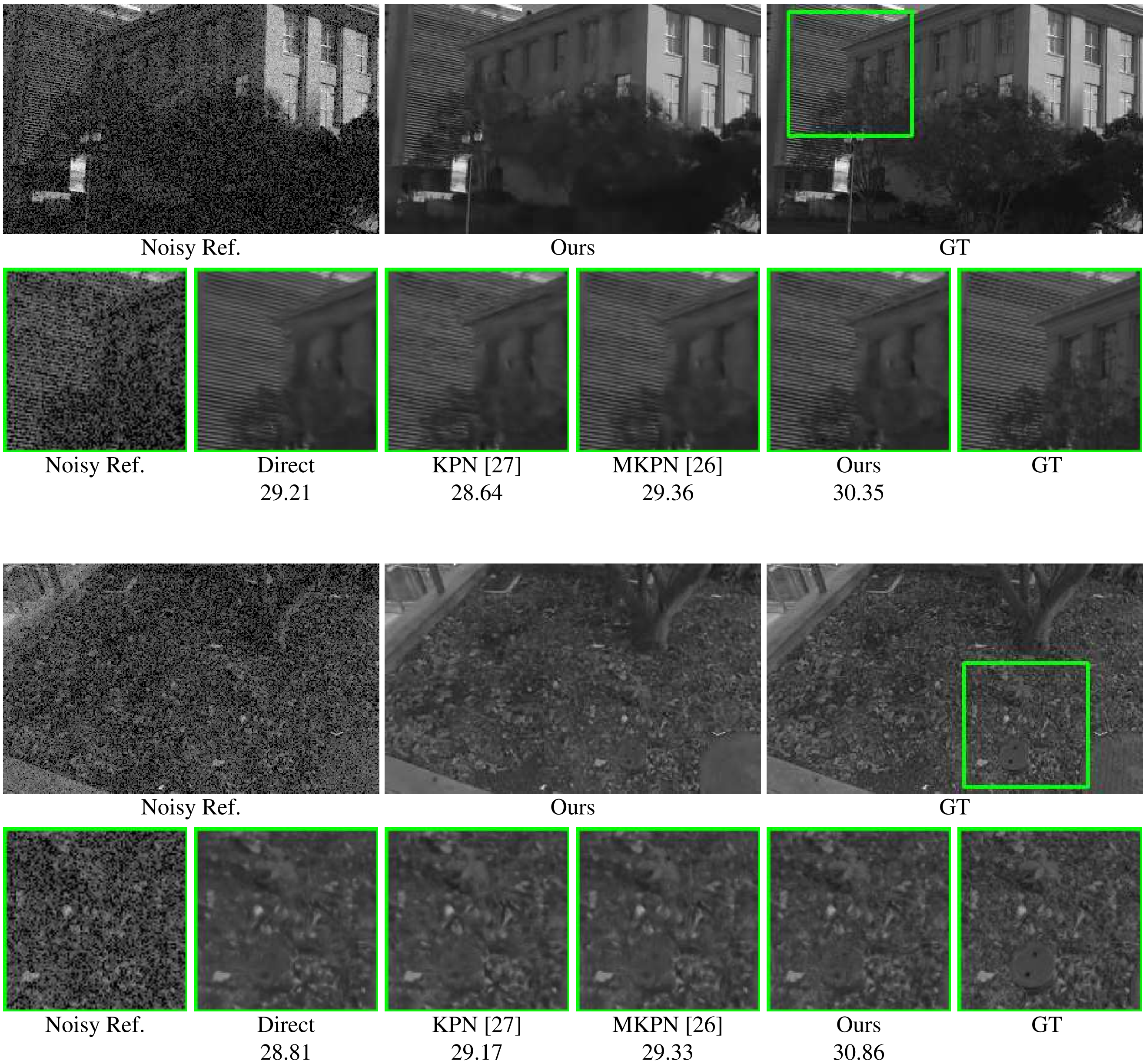

Figure 10: Additional grayscale results on benchmark test set [27]. Numbers refer to PSNR (dB) on the full image.

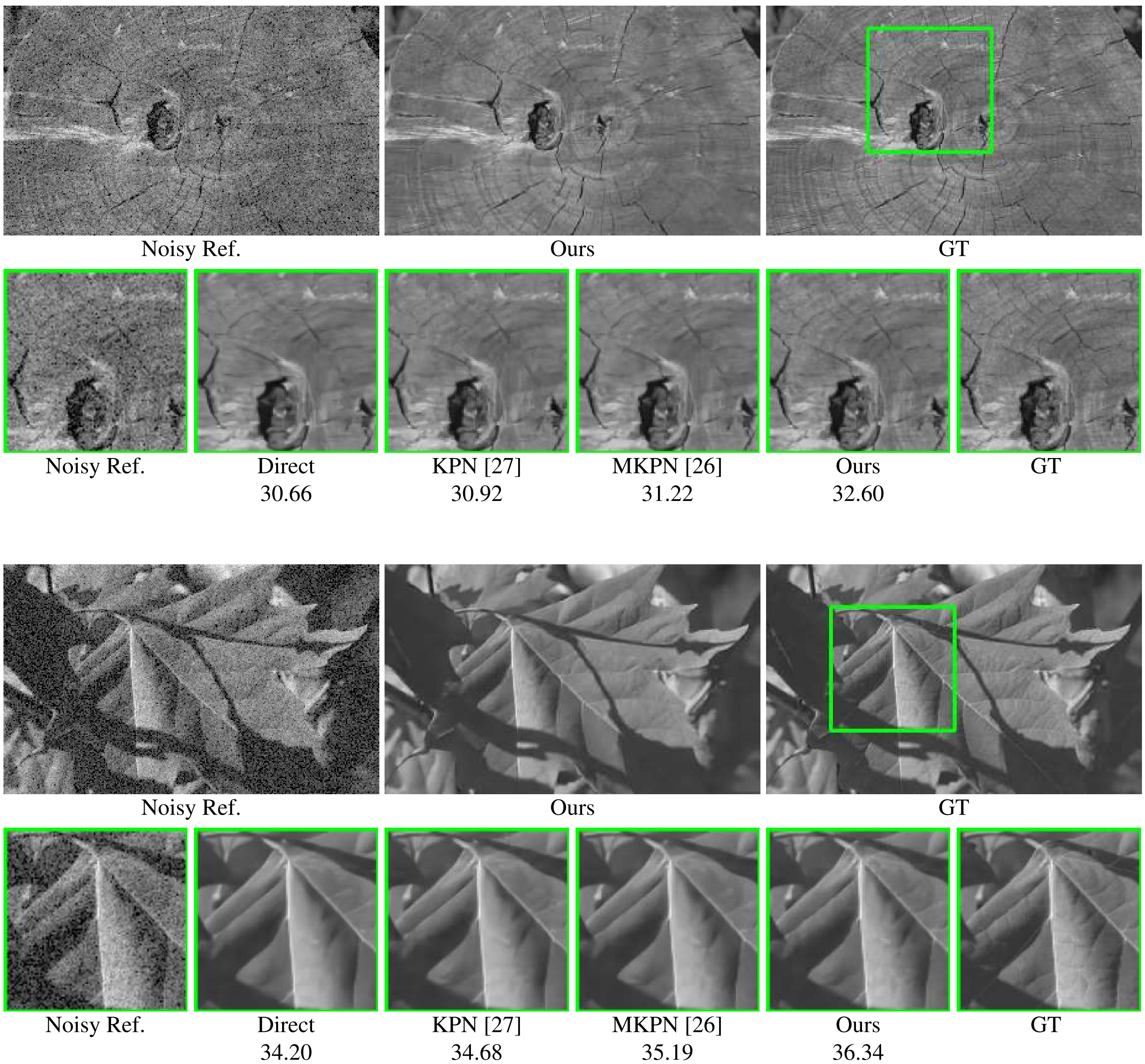

Figure 11: Additional grayscale results on benchmark test set [27]. Numbers refer to PSNR (dB) on the full image.