Supplemental Material: Robust Lensless Image Reconstruction via PSF Estimation

Joshua D. Rego
Arizona State University

Karthik Kulkarni
Arizona State University

Suren Jayasuriya
Arizona State University

1. Network Architecture Details

For the network architectures, we use the same Generator and discriminator architecture for both the reconstruction and PSF estimation networks. For the generator network architecture, detailed in Table 1, we use five encoding layers, five decoding layers and a convolution layer in the center and output. Skip connections are also used to transfer high-level features to the decoding layers. Each encoding layer uses two convolution layers and then performs batch normalization and a ReLU activation function. Each decoding layer uses an upsampled output from the previous layer and concatenates it with the corresponding encoding layer output before using three convolution layers. A kernel size, **k**, of 3×3 is used for all layers except the output layer and a stride, **s**, of 1 is used for all layers.

Table 1: Network architecture parameters for the generator

Generator					
Input	Layer	k	s	Channels in/out	Activation
image	enc1	3	1	3 / 24	ReLU
enc1	enc2	3	1	24 / 64	ReLU
enc2	enc3	3	1	64 / 128	ReLU
enc3	enc4	3	1	128 / 256	ReLU
enc4	enc5	3	1	256 / 512	ReLU
enc5	conv1	3	1	512 / 512	-
conv1'+enc5	dec5	3	1	512 / 256	ReLU
dec5'+enc4	dec4	3	1	256 / 128	ReLU
dec4'+enc3	dec3	3	1	128 / 64	ReLU
dec3'+enc2	dec2	3	1	64 / 24	ReLU
dec2'+enc1	dec1	3	1	24 / 24	ReLU
dec1	conv2	1	1	24 / 3	-

For the discriminator architecture, detailed in Table 2, we use four convolution layers with the LeakyReLU activation function and an output layer that uses a Linear activation function. Each layer uses a kernel size of 3×3 and a stride of 2 except the output layer that uses a stride of 1. For training we take the mean of the output logits from the discriminator as the critics score for the generator's output.

Table 2: Network architecture parameters for the discriminator

Discriminator					
Input	Layer	k	s	Channels in/out	Activation
image	conv1	3	2	3 / 64	LeakyReLU
conv1	conv2	3	2	64 / 128	LeakyReLU
conv2	conv3	3	2	128 / 256	LeakyReLU
conv3	conv4	3	2	256 / 512	LeakyReLU
conv4	out	-	-	512 / 1	Linear

2. Additional Results

In the next few pages, we show additional results for our single PSF reconstruction, multiple PSF reconstruction, and real data captured from the hardware prototype.

References

[1] Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Learned reconstructions for practical mask-based lensless imaging. *Opt. Express*, 27(20):28075–28090, 09 2019.

Figure 1: **Comparison on DiffuserCam dataset [1]**: We perform single PSF reconstruction on a set of 9 test images using different models. Note that output of our model is very close to the Ground Truth and performs on-par with the state-of-the-art Le-ADMM-U model.

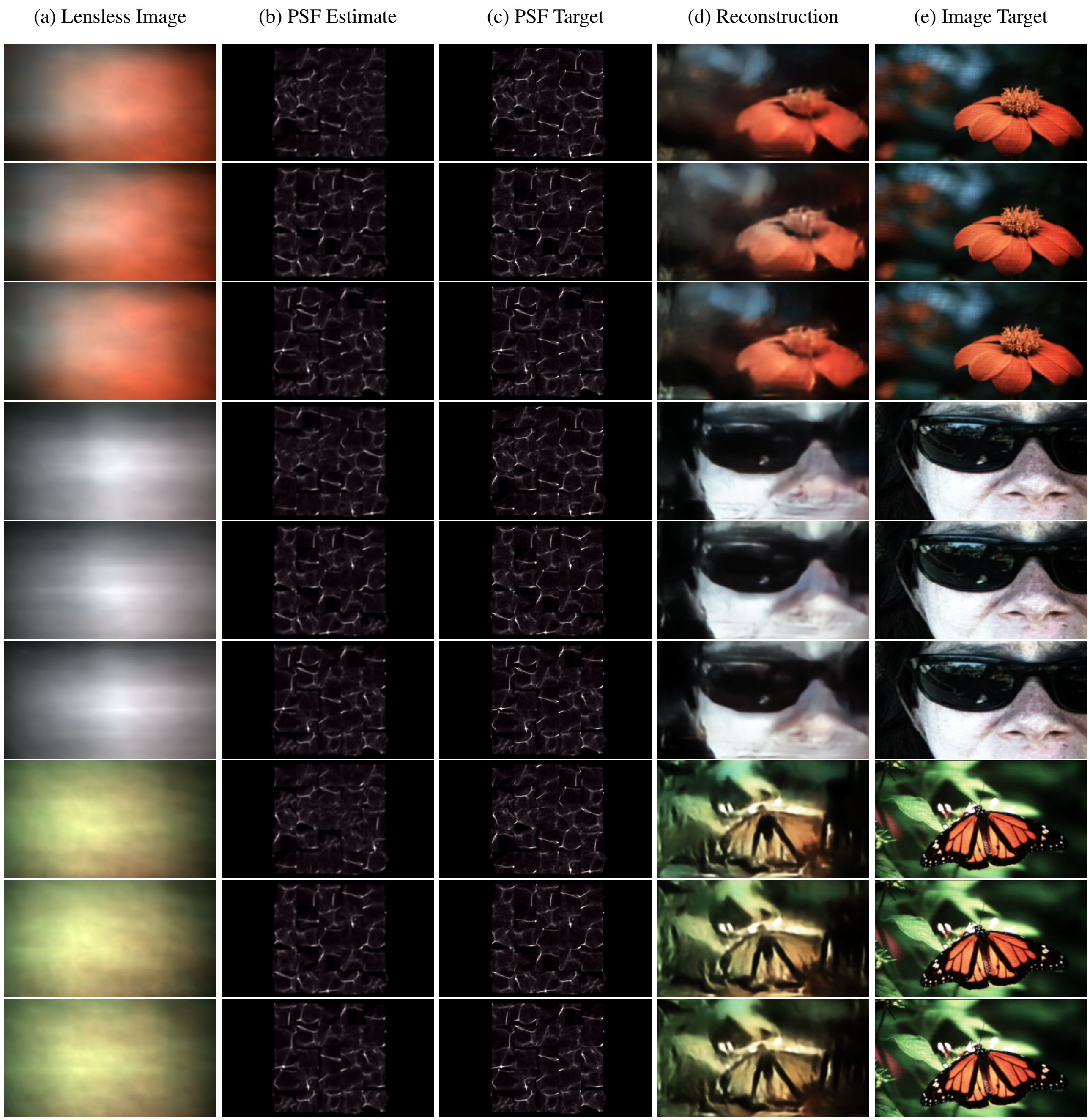

Figure 2: **Multiple PSF Reconstruction:** We display the PSF estimated as well as the reconstructed images for three different simulated PSFs on the DiffuserCam dataset [1]. Our method is able to generalize and reconstruct multiple PSFs, although the quality of the reconstruction is missing some high frequency information. However, the resulting reconstructions are uniform in quality across the three different PSFs. This shows the potential for generalization of our network for multiple lensless camera without training a separate model for each one.

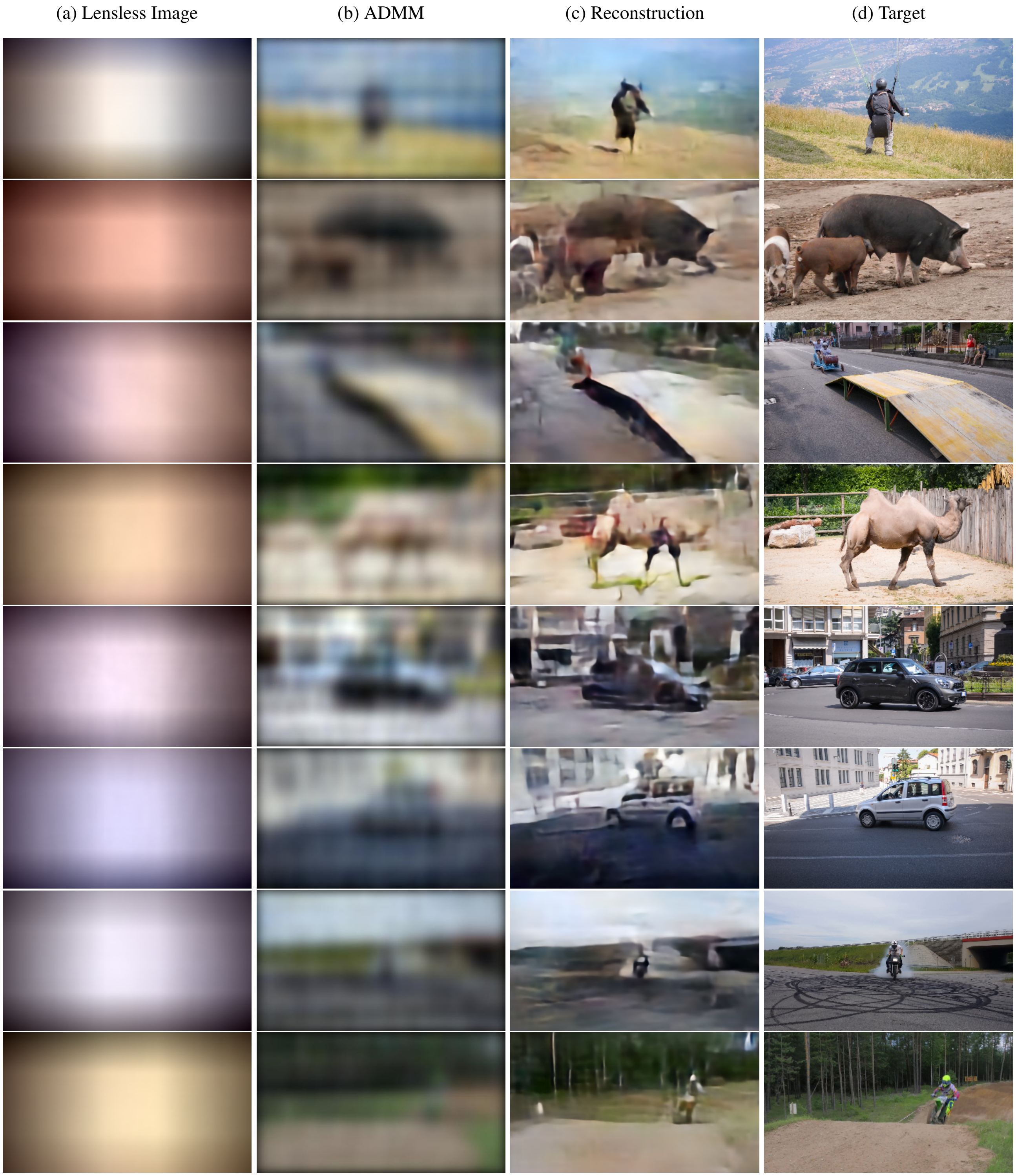

Figure 3: **Real Data with Multiple PSFs:** Example reconstructions of real image data captured with three different PSFs. Here reconstruction performance was not as visually pleasing, primarily due to the real PSFs quality in captured caustics. Despite this limitation, the network is still able to invert the meaurements to achieve plausible reconstructions.