BNUDC: A Two-Branched Deep Neural Network for Restoring Images from Under-Display Cameras
Supplementary Material

Jaihyun Koh[1,2], Jangho Lee[1] and Sungroh Yoon[1,3,4]

[1] Department of Electrical and Computer Engineering, Seoul National University

[2] Samsung Display Corporation [3] Interdisciplinary Program in AI, Seoul National University

[4] AIIS, ASRI, INMC and ISRC, Seoul National University

{satyricon, ubuntu, sryoon}@snu.ac.kr

A. Inverse Color Filtering for POLED dataset

We have discussed the inverse color filtering for images from the POLED dataset as a pre-process reversing the color shift in the training dataset. The measured average X, Y, and Z filter values on the training images with respect to the pixel position $\eta_{\mathrm{xyz}}(m, n)$ are obtained by

$$\eta_{\mathrm{xyz}}(m, n) = \frac{1}{N} \sum_{i=1}^{N} \frac{y^i_{\mathrm{xyz}}(m, n)}{x^i_{\mathrm{xyz}}(m, n)}, \tag{1}$$

and its graphical results are shown in Fig. A1. This suggests that the transmission properties of wavelengths in the OLED layers are spatially variant. We have two options to obtain the inverse color filter: one is measuring pixel-wise filters $\eta^l_{\mathrm{xyz}} \in \mathbb{R}^{h \times w \times 3(XYZ)}$ by (1), and the other is obtaining a global filter $\eta^g_{\mathrm{xyz}} \in \mathbb{R}^{3(XYZ)}$ which can be expressed as

$$\eta^g_{\mathrm{xyz}} = \frac{1}{h \times w} \sum_{m=1}^{h} \sum_{n=1}^{w} \eta^l_{\mathrm{xyz}}(m, n), \tag{2}$$

which is equivalent to equation (9) in the main body. We performed two experiments applying local η^l_{xyz} and global η^g_{xyz} inverse filtering, and empirically found that the performance of global filtering is better. Local inverse filtering can remove the pixel artifact which occurs at static pixel position and can significantly improve initial color differences. However, it tends to remove edge information from images, resulting in low image quality in terms of both per-pixel differences and perceptual quality. A further examples of inverse filtering are shown in Fig. A2.

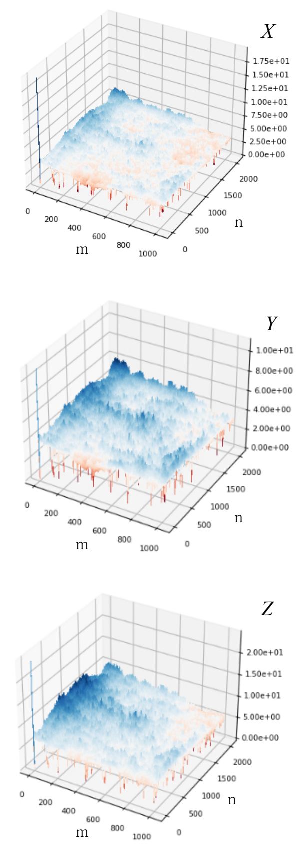

Figure A1. Average XYZ of measured color filter on the pixel position (m, n) $\eta_{\mathrm{xyz}}(m, n)$ obtained by comparing UDC and ground truth images in the training images from POLED dataset (see equation (1)). It shows that the color transmission of thin-film layers in OLEDs is spatially variant.

B. Analysis with Guided Filter

Our 1D affine transform connection is associated with a local linear model in the guided filter [3, 11], which transfers the high-frequency structure from the degraded image

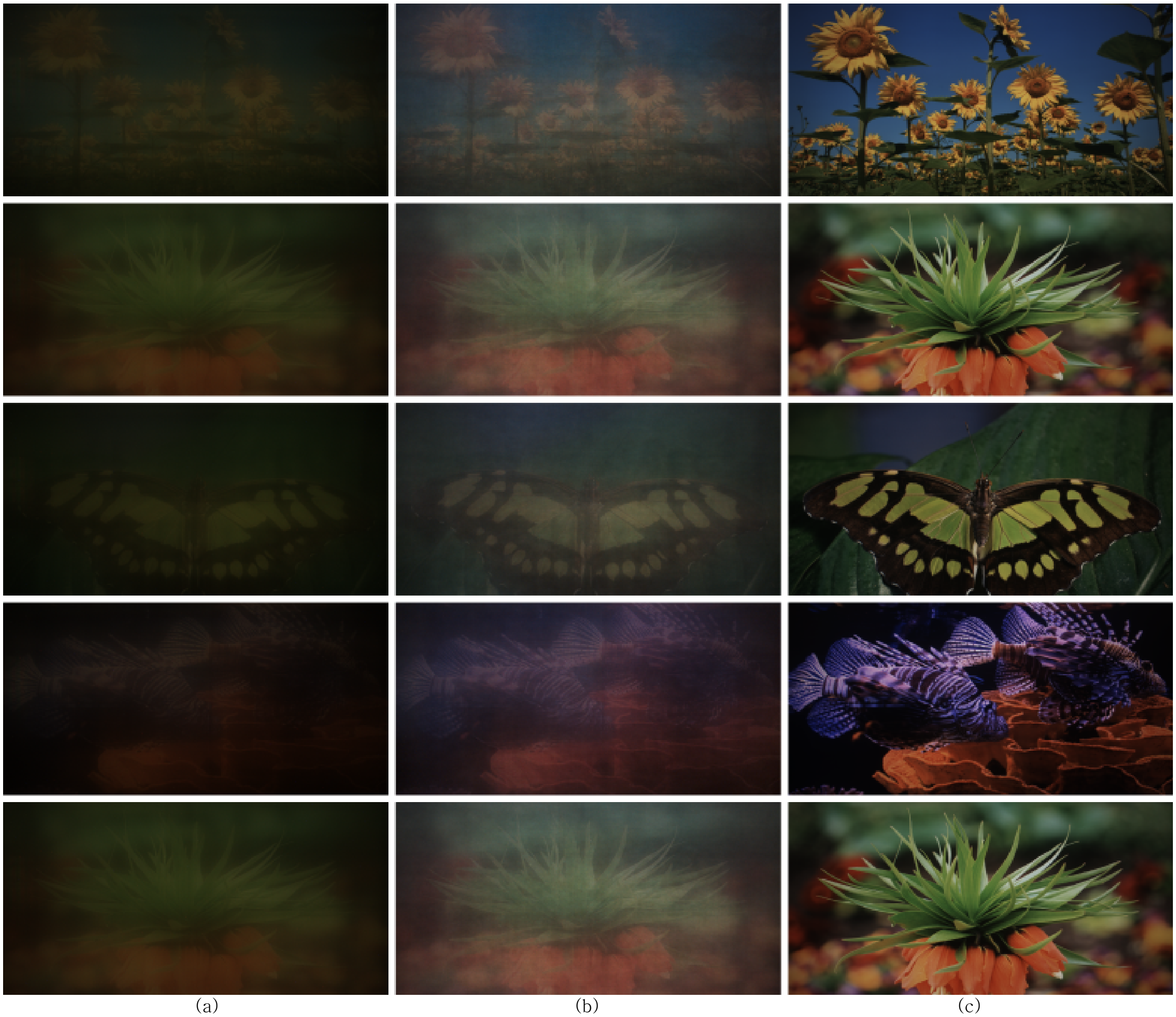

Figure A2. An example of the use of inverse color filtering to pre-process an image from the POLED dataset: (a) UDC image, (b) after inverse color filtering, and (c) ground truth.

to the output of the HFR branch. The 3D affine transform connection in the LFR branch is its extended version which constrains the solution space to color adjustment and low-frequency reconstruction, in which low-frequency features are transferred to the output. As we discussed in the main body, linear transformations give the network inductive bias to change styles or remove noise while preserving the structure of the guided image.

Analogies to the affine transform approach include the deep-learning method for style transfer [5, 10], in which adaptive instance normalization [4] changes the style of an image while preserving its structure. Similarly, SPADE [6] synthesizes a photo-realistic image with its structure taken from a segmentation map, using linear transformation normalization. Our affine transform connection can be viewed as changing a corrupted style to a clean style, while maintaining the structural information of the image.

C. Detailed Network Branches

The components in our BNUDC network are presented in Fig. A3. In the high-frequency reconstruction network, We use the flat network which maintains the resolution of the input image in the feature space, and also use a par-

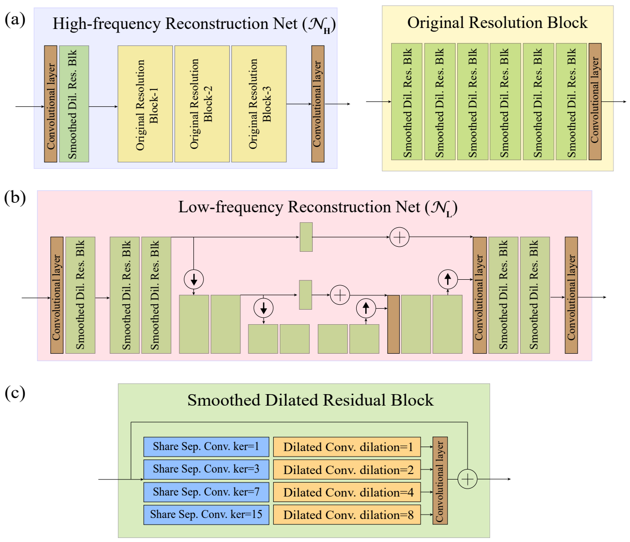

Figure A3. Network branches: (a) the HFR network $\mathcal{N}_\mathrm{H}$ composed of original-resolution blocks of flat networks; (b) the LFR $\mathcal{N}_\mathrm{L}$ which is a U-net [7]; and (c) a unit residual block which uses parallel dilated convolution layers.

allel dilated convolution residual block [1, 9]. In the low-frequency reconstruction network, we employ the U-Net with a skip up-sampling scheme.

During training the depth of the feature space in the HFR and LFR network is 72 and 36 channels respectively. The HFR network consists of three full-resolution blocks, each of which contains six smoothed dilated residual blocks. The LFR network has fifteen smoothed dilated residual blocks (see Fig. A3).

D. Additional Results

We provide additional experimental results in Fig. A4, A5, A6, A7, A8, A9 and A10, and refer to the captions for the information.

References

[1] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 1375–1383. IEEE, 2019. 3

[2] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, and Jinwei Gu. Removing diffraction image artifacts in under-display camera via dynamic skip connection network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 662–671, 2021. 8

[3] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer, 2010. 1

[4] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017. 2

[5] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020. 2

[6] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3

[8] Hrishikesh Panikkasseril Sethumadhavan, Densen Puthussery, Melvin Kuriakose, and Jiji Charangatt Victor. Transform domain pyramidal dilated convolution networks for restoration of under display camera images. In *European Conference on Computer Vision*, pages 364–378. Springer, 2020. 5

[9] Varun Sundar, Sumanth Hegde, Divya Kothandaraman, and Kaushik Mitra. Deep atrous guided filter for image restoration in under display cameras. In *European Conference on Computer Vision*, pages 379–397. Springer, 2020. 3, 5

[10] Xiaolong Wang and Abhinav Gupta. Generative image modeling using style and structure adversarial networks. In *European conference on computer vision*, pages 318–335. Springer, 2016. 2

[11] Huikai Wu, Shuai Zheng, Junge Zhang, and Kaiqi Huang. Fast end-to-end trainable guided filter. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1838–1847, 2018. 1

[12] Yuqian Zhou, Michael Kwan, Kyle Tolentino, Neil Emerton, Sehoon Lim, Tim Large, Lijiang Fu, Zhihong Pan, Baopu Li, Yang Qirui, Yihao Liu, Jigang Tang, Tao Ku, Shibin Ma, Bingnan Hu, Jiarong Wang, Densen Puthussery, Hrishikesh P S, Melvin Kuriakose, and Lianping Xing. Udc 2020 challenge on image restoration of under-display camera: Methods and results. 08 2020. 7

[13] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9179–9188, 2021. 5, 7

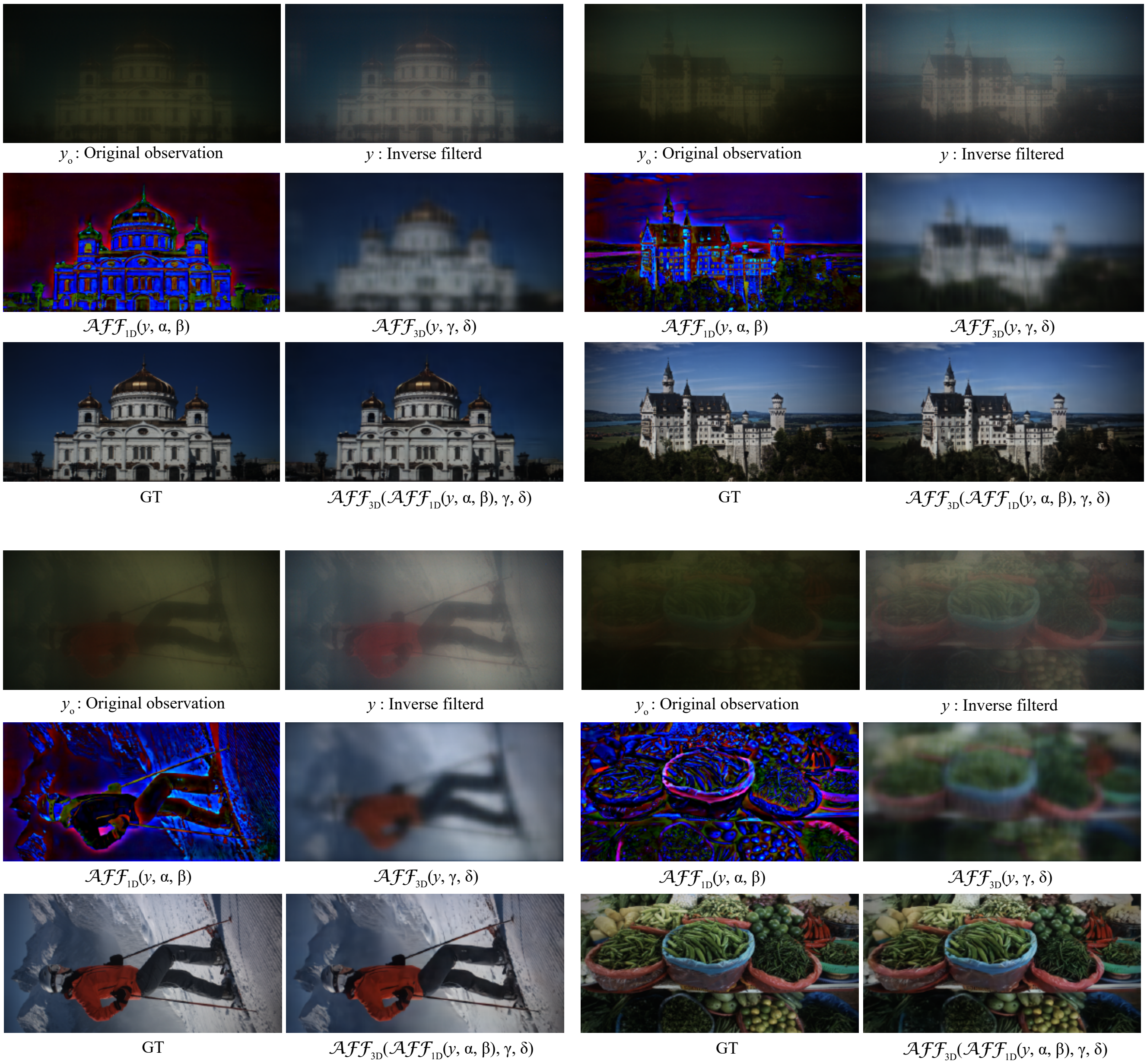

Figure A4. Visualizations of intermediate images from the HFR and LFR branches on an example from the POLED dataset. In each images group, the top row shows the UDC image (left) and the same image after pre-processing (right); the second row show the results from the HFR (left) and the LFR network (right); and the last row contains the ground-truth image (left) and the restored image (right).

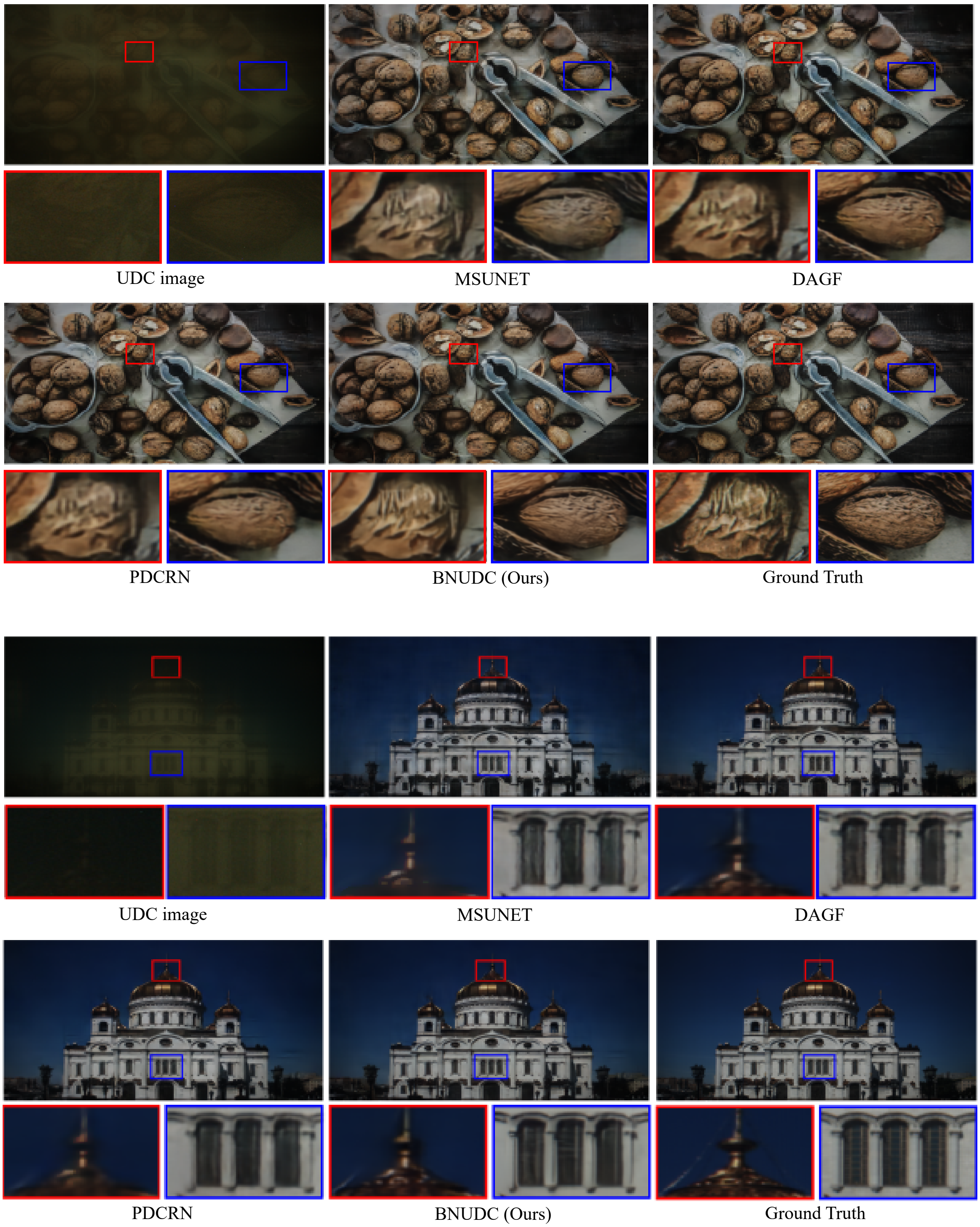

Figure A5. Example images from the POLED dataset restored by four different networks. In each image group, the first row contains the original UDC images, reconstructed images obtained using MSUNET [13] and DAGF [9]. The second row shows the restored images obtained by PDCRN [8], our BNUDC and the ground truth image.

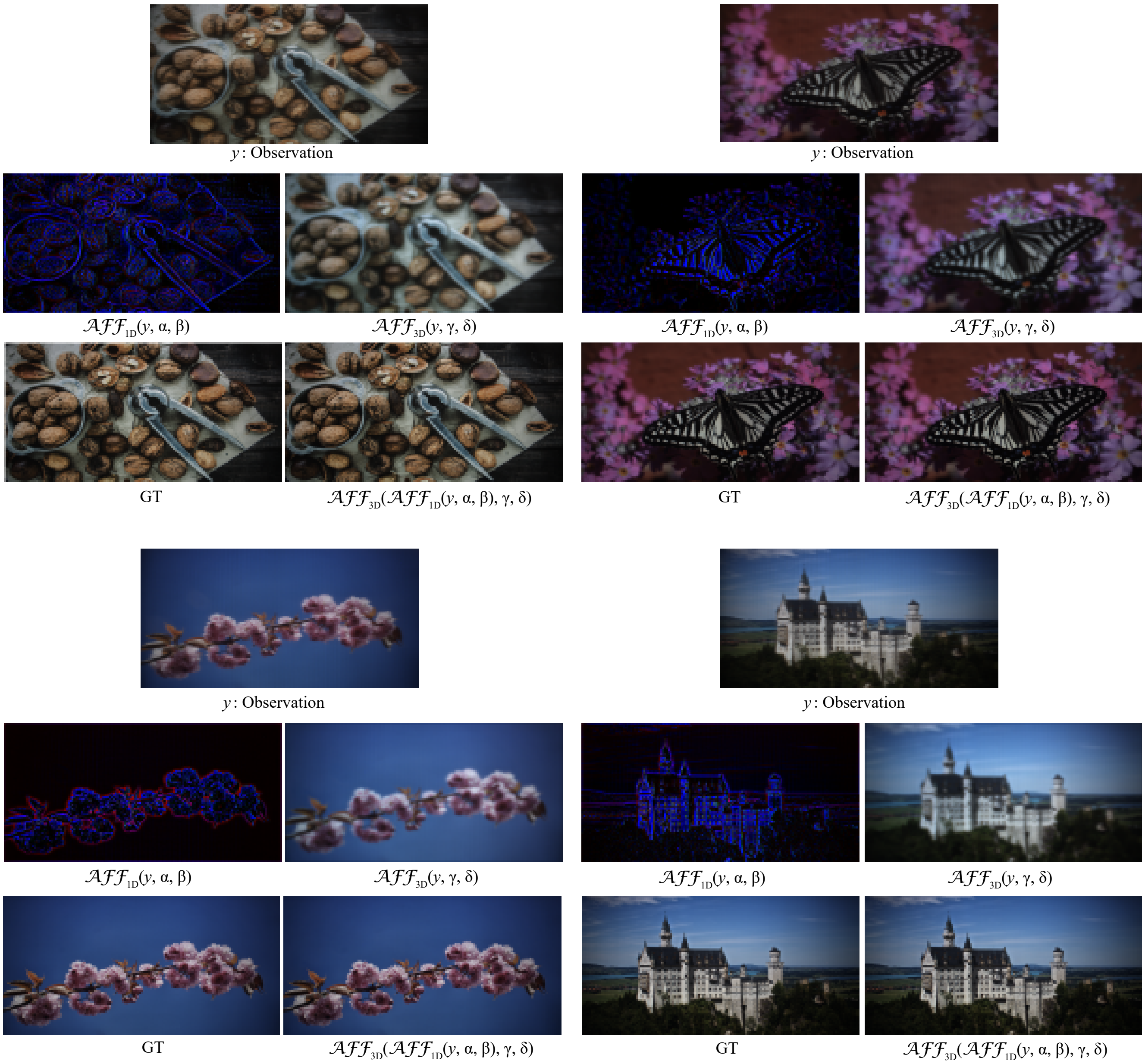

Figure A6. Visualizations of intermediate images from the HFR and LFR branches on an example from the TOLED dataset. In each images group, the top row shows the captured UDC image, the second row show the results from the HFR (left) and the LFR network (right); and the last row contains the ground-truth image (left) and the restored image (right).

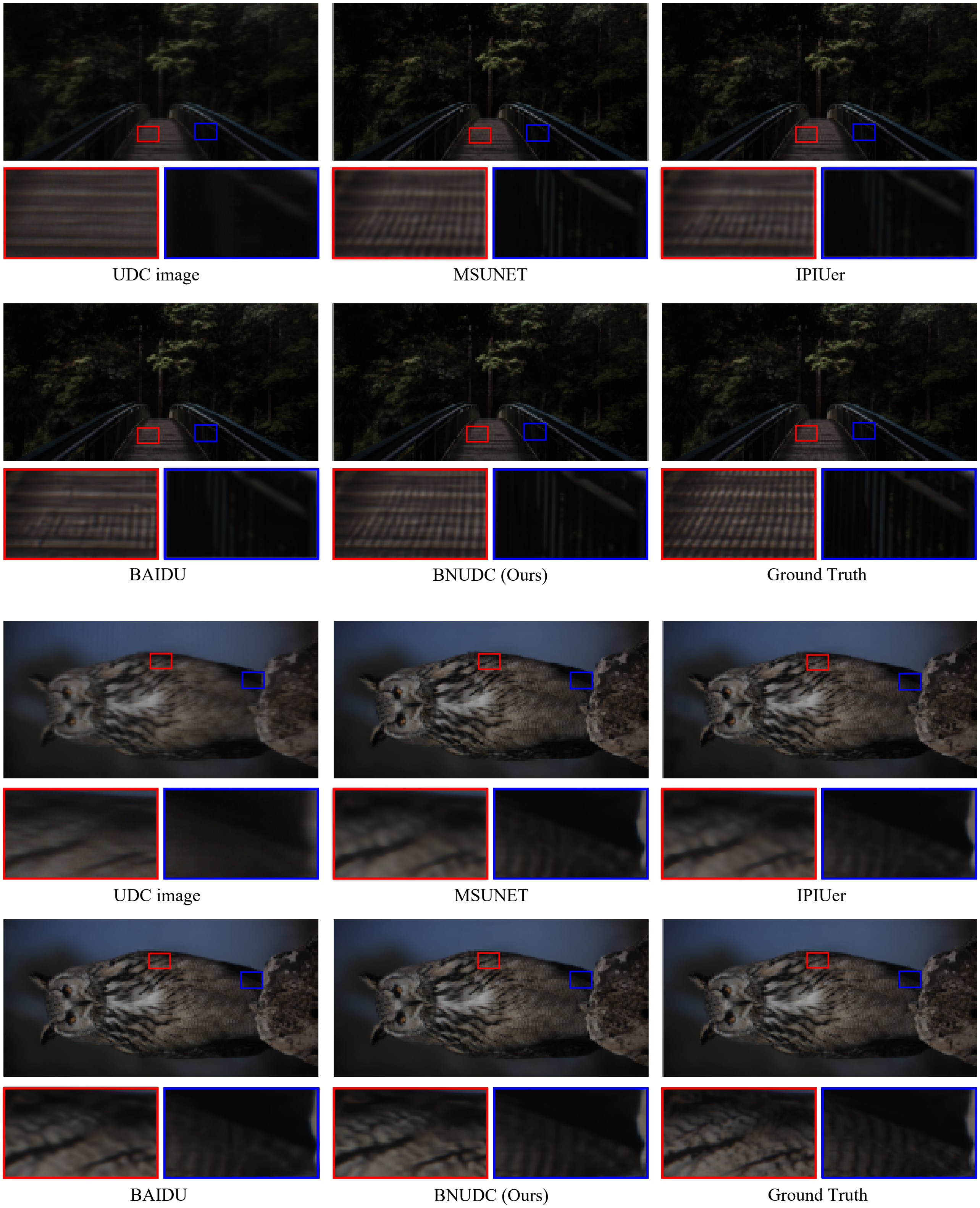

Figure A7. Example images from the TOLED dataset restored by four different networks. In each image group, the first row contains the original UDC images, reconstructed images obtained using MSUNET [13] and IPIUer [12]. The second row shows the restored images obtained by BAIDU [12], our BNUDC and the ground truth image.

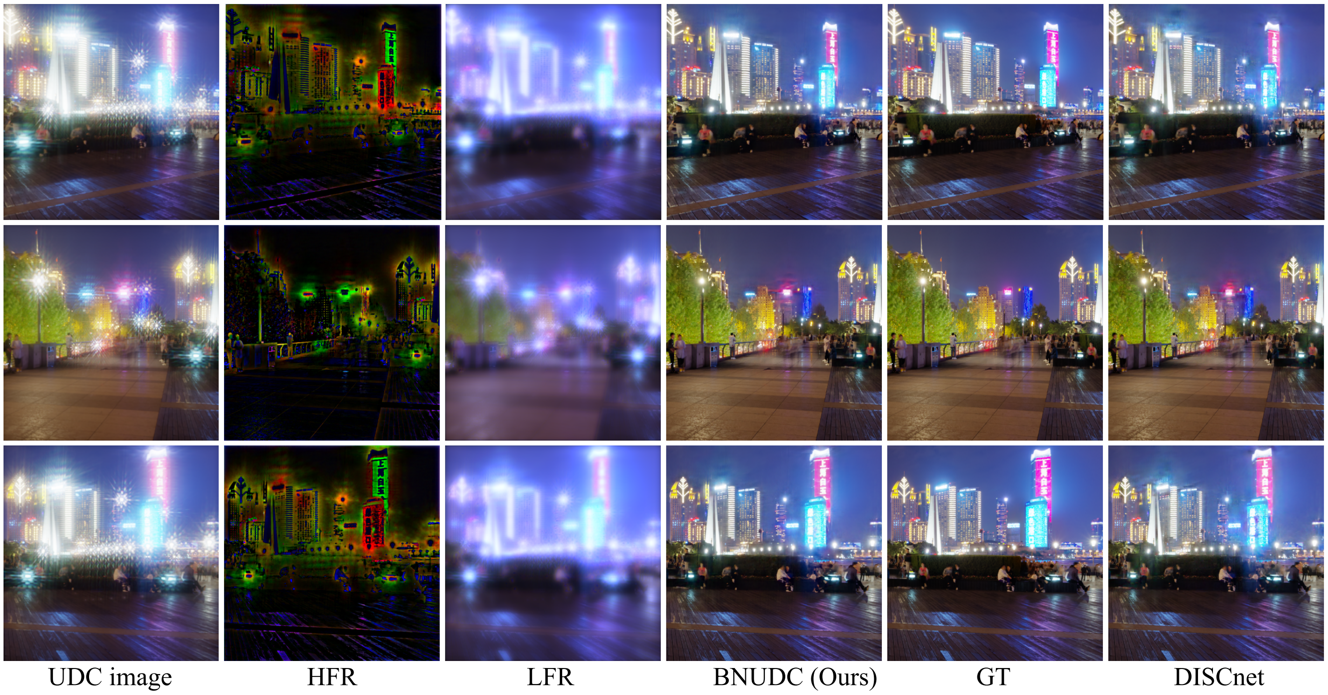

Figure A8. Six example images from the SYNTH dataset. The first column contains the original UDC images. The next two column contains the intermediate results obtained by the HFR and LFR branch. Subsequent columns shows the restored images obtained using our BNUDC, ground truth images, and restored images by DISCnet [2]. (Best viewed in digital version with zoom.)

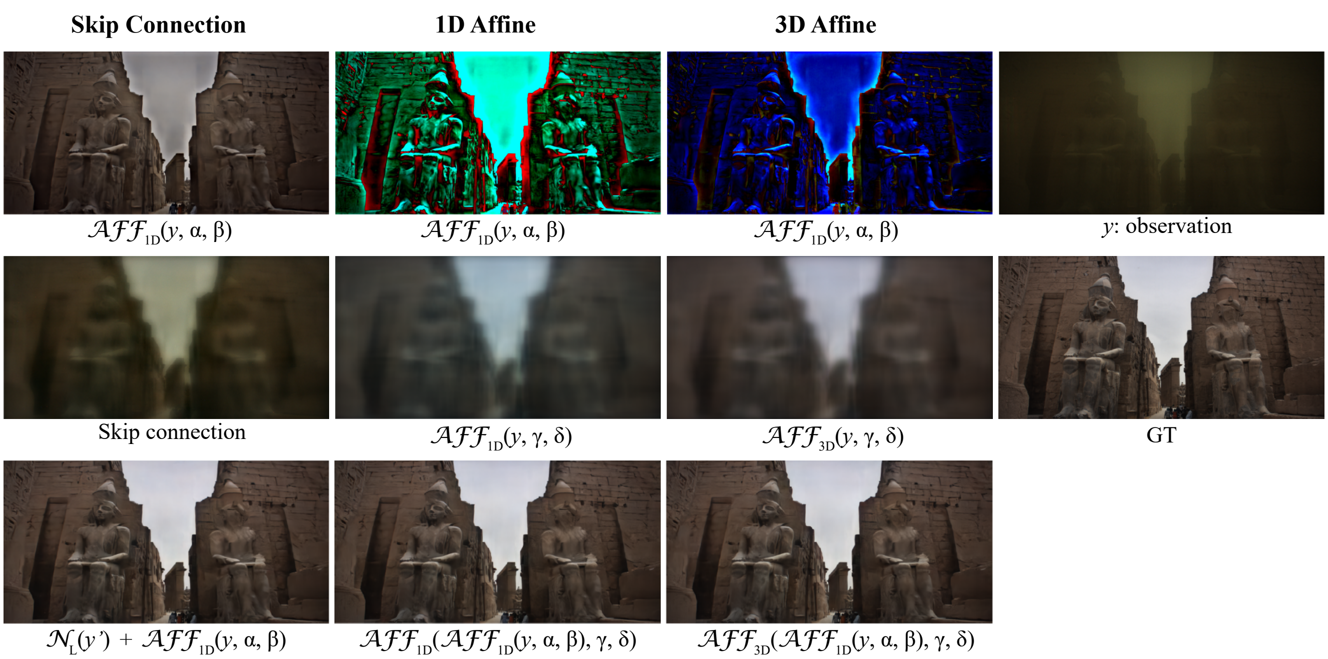

Figure A9. Task separability. The first column shows images obtained with a skip connection in the LFR branch; the second column shows images obtained using a 1D affine in the LFR branch; and the third column shows images obtained using a 3D affine transform in the LFR branch. Each column contains (from top to bottom) results from the HFR, LFR, and the final restored image.

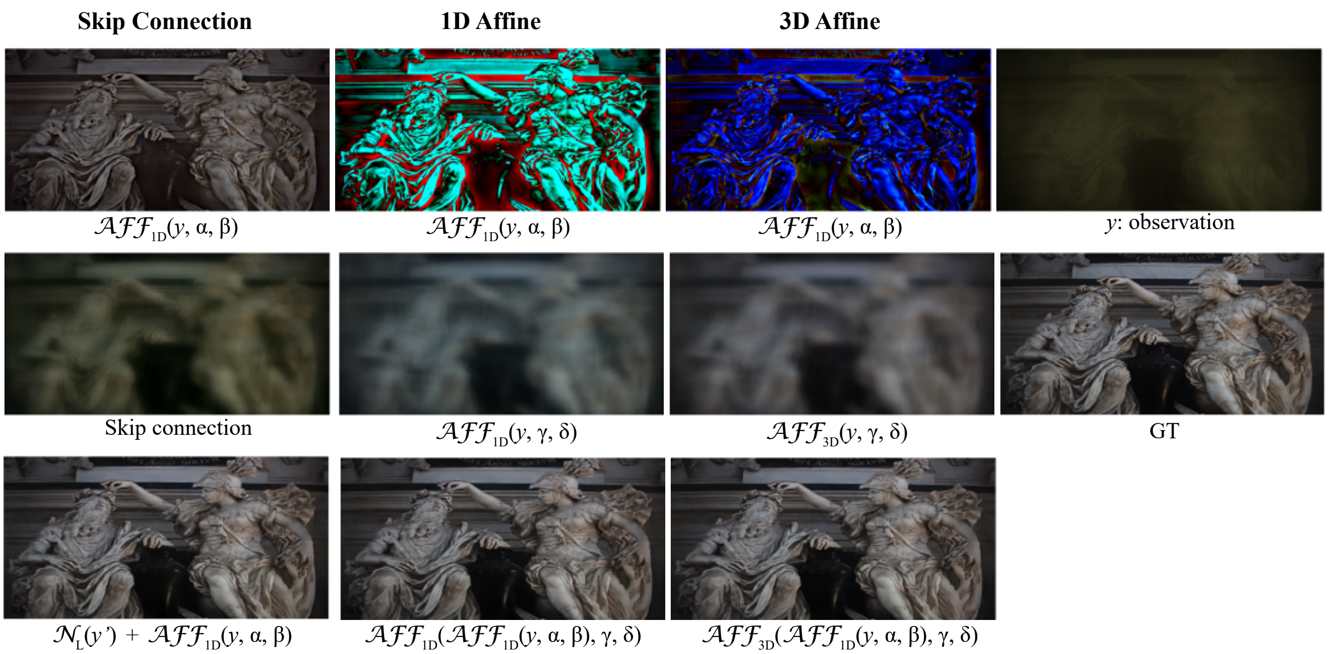

Figure A10. Task separability. The first column shows images obtained with a skip connection in the LFR branch; the second column shows images obtained using a 1D affine in the LFR branch; and the third column shows images obtained using a 3D affine transform in the LFR branch. Each column contains (from top to bottom) results from the HFR, LFR, and the final restored image.