(Supplementary Material)
Large Scale Multi-Illuminant (LSMI) Dataset for Developing White Balance Algorithm under Mixed Illumination

Dongyoung Kim[1], Jinwoo Kim[1], Seonghyeon Nam[2], Dongwoo Lee[1], Yeonkyung Lee[3], Nahyup Kang[3], Hyong-Euk Lee[3], ByungIn Yoo[3], Jae-Joon Han[3], Seon Joo Kim[1]

[1]Yonsei University, [2]York University, [3]Samsung Advanced Institute of Technology

Nikon D810	Single		Multi		Mixed	
	mean	median	mean	median	mean	median
[1]	3.18	2.61	4.65	4.19	3.93	3.48
HDRnet	2.76	2.43	3.20	3.01	2.99	2.61
U-Net	**1.51**	**1.14**	**2.36**	**1.84**	**1.95**	**1.45**
Sony α9	Single		Multi		Mixed	
	mean	median	mean	median	mean	median
[1]	3.25	2.62	4.38	3.93	3.86	3.19
HDRnet	2.70	**2.37**	3.65	3.33	3.21	2.89
U-Net	**2.83**	2.44	**3.04**	**2.78**	**2.94**	**2.66**

Table 1. Mean Angular Error (MAE) values on the test set captured by Nikon D810 and Sony α9.

1. LSMI Dataset Samples

Fig. 1 and Fig. 2 show the samples of LSMI dataset. Our dataset contains various real-world scenes with dynamic scene geometry and natural lighting situation. As shown in the illuminant coefficient map of each scene, the mix of illumination happens naturally, and fully reflecting the scene geometry.

2. Experiment Results

Table 1 shows the pixel-level MAE, each models are trained with respect to Nikon D810 and Sony α9 data subset of LSMI. As shown in the table, HDRnet and U-Net outperforms patch-based method [1], despite no additional modification of the model structure for illumination estimation. This indicates that the LSMI dataset gives the model a good intuition for white balancing at the data level. Fig. 3 shows qualitative results of white balanced images under multi-illuminated settings. It can be easily seen that the pixel-level methods, HDRnet and U-Net work better than the patch-based model [1] on multi-illumination images. In more detail, we observed that there is a huge gap in performance near the window between the patch-based method and the pixel-level methods.

3. Pixel-Level Relighting Results

Since multi-illuminant scene of LSMI dataset provides pixel-level mixture ratio of each illuminant, we can freely manipulate the color of lights through pixel-level relighting process. In addition, we can also freely relight a single-illuminant subset of LSMI. Since relighting of a single-illuminant scene is done by multiplying all pixels by the same chromaticity value, here we explains how to do pixel-level relighting in a multi-illuminant scene.

We already have ground truth illumination map and white balanced image of each scene, so the only thing we need to do is to relight the original multi-illuminant image. First, we sample HSV color vectors within the range H[0,1], S[0.2,0.8], and V=1. The number of HSV color vector is determined by the number of illuminants in the scene, 2 or 3 here. When sampling the Hue values to be applied to each light, the absolute value of the difference between the hue values is set to be at least 0.3. The Sampled HSV vector are converted into RGB vectors and normalized so that G=1. We have mixture ratio map with respect to each scene, so sampled RGB vectors are properly interpolated while reflecting the scene geometry. Fig. 4 shows pixel-level relighted samples of LSMI dataset. We speculate that almost all walls are white, and making the wall color neutral would have been a favorable learning direction for the model.

4. Failure Cases

Since we use the default architecture design for HDRnet and U-Net and there is no structural modification, the results of pixel-level white balancing contain some failure cases.

For the HDRnet, we observed that the model is not good

MAE	Single		Multi	
	mean	median	mean	median
FC4 [2]	**1.21**	**0.84**	4.45	3.57
U-Net	2.95	1.86	**2.35**	**2.00**

Table 2. Mean Angular Error (MAE) values on the test set of LSMI by using FC4 and U-Net (R3).

at scenes with colored illuminations like the second and the last row in Fig. 5. We think that the rare combination of the illuminations in the scene which has not been seen frequently in the train set triggers this problem. The augmentation techniques that provide more various combinations of the colors of illuminations could address this problem.

We found that the U-Net sometimes makes green objects achromatic, as shown in the first row of Fig. 5. The second and the third row of Fig. 5 shows the additional failure cases related to the color of the wall. Images show that U-Net sometimes produce results with slightly lower saturation than the original saturation of the wall. We speculate that almost all walls are white, and making the wall color neutral would have been a favorable learning direction for the model.

5. Comparison with global WB methods

In Table 2, we additionally compared [2], a state-of-the-art DNN-based single illumination WB model on the mean angular error (MAE) metric. The FC4 model is trained on a single illuminant training subset of LSMI, and tested on the multi illuminant test subset of LSMI.

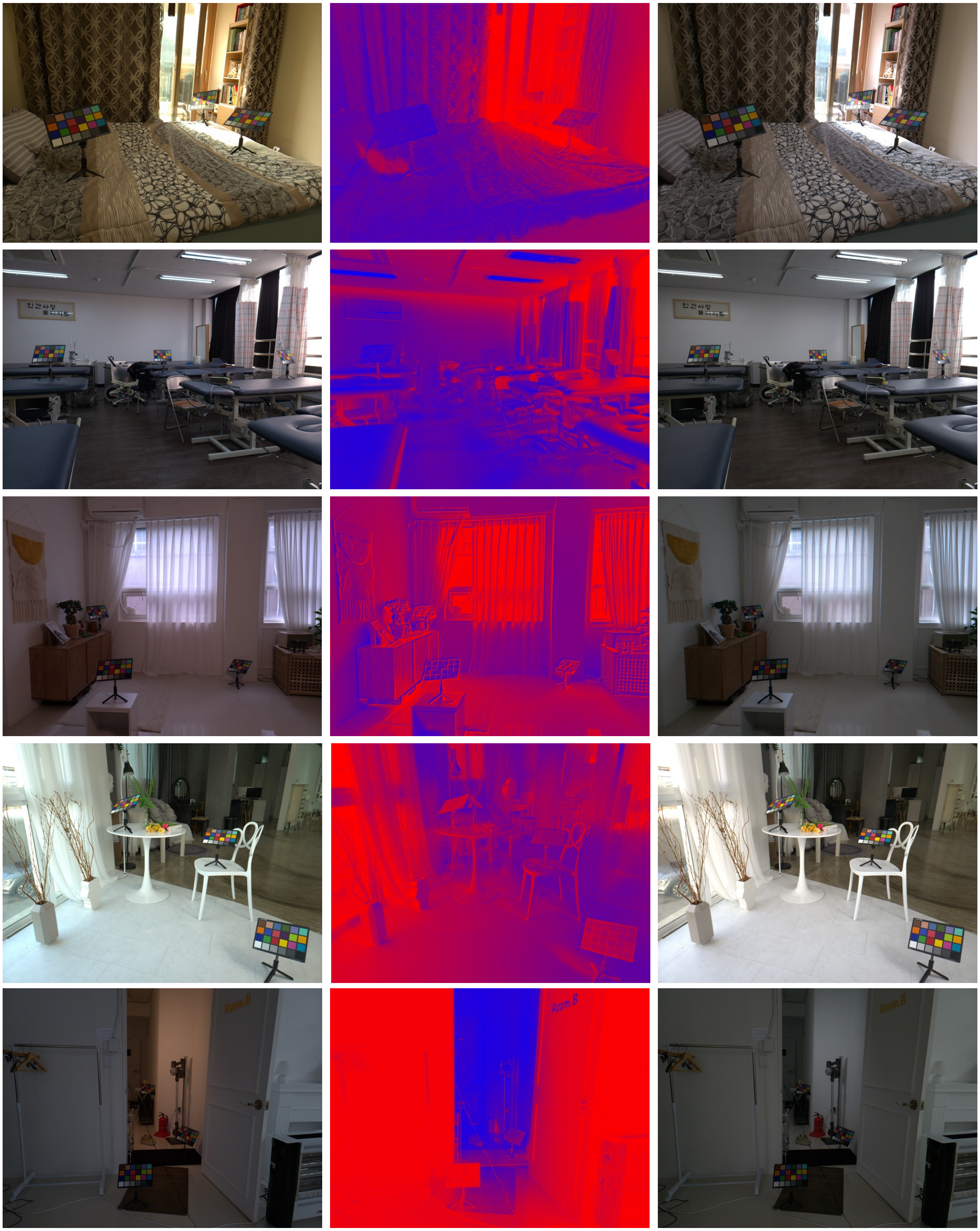
6. More details of user study

Fig. 6 shows an example of our survey. In the first question, participants picked the photo which is most likely captured under white light. In the second, participants chose the most preferred photo. Our results were preferred by the majority of users as shown in the main paper.

References

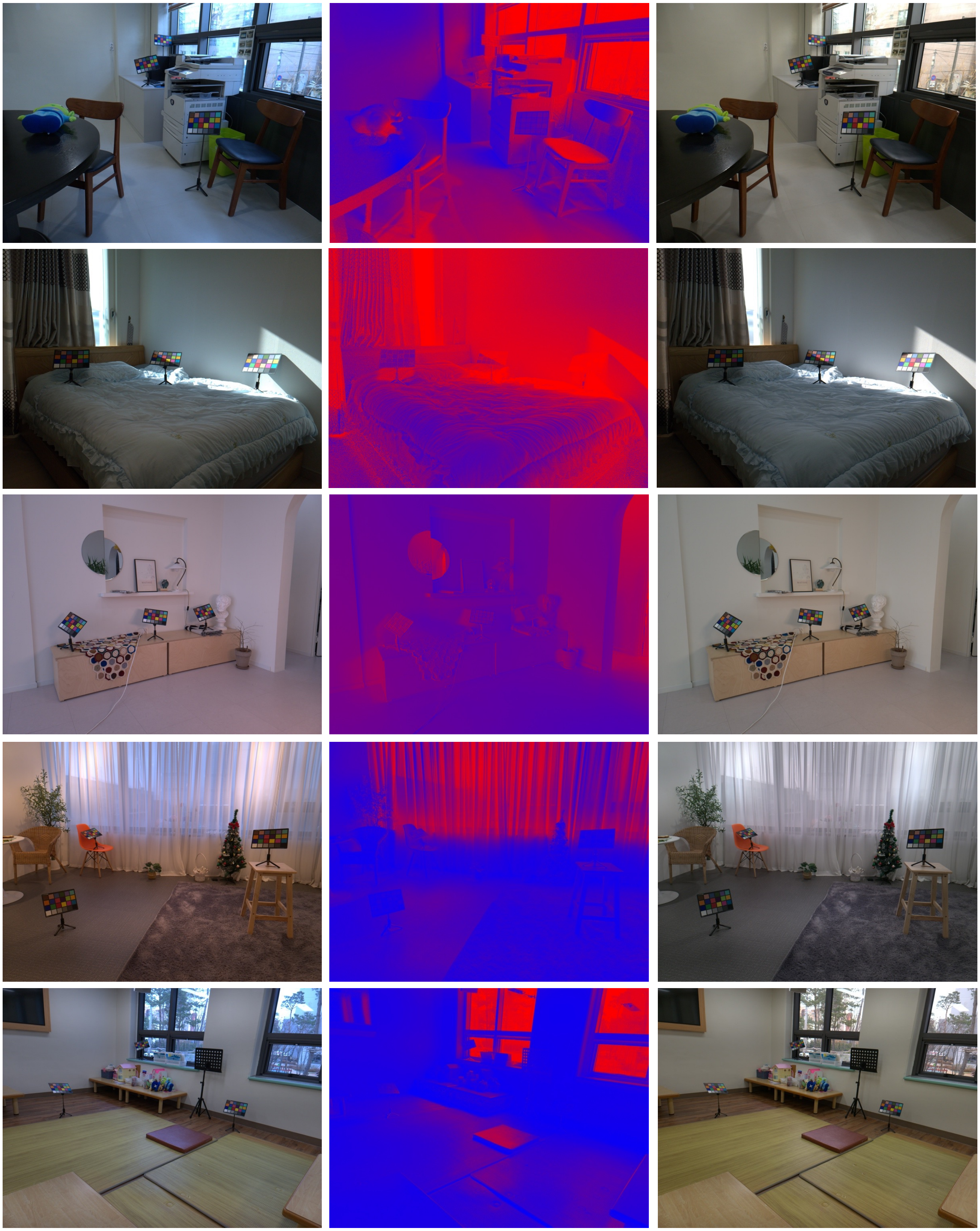
[1] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, 26(9):4347–4362, 2017. 1

[2] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085–4094, 2017. 2



[Multi-illuminant scene] [Illuminant coefficient map] [Pixel-level WB Image]

Figure 1. Samples of LSMI dataset. Multi-illuminant scene and pixel-level white balanced images are converted into sRGB space for visual convenience.



[Multi-illuminant scene] [Illuminant coefficient map] [Pixel-level WB Image]

Figure 2. Samples of LSMI dataset. Multi-illuminant scene and pixel-level white balanced images are converted into sRGB space for visual convenience.

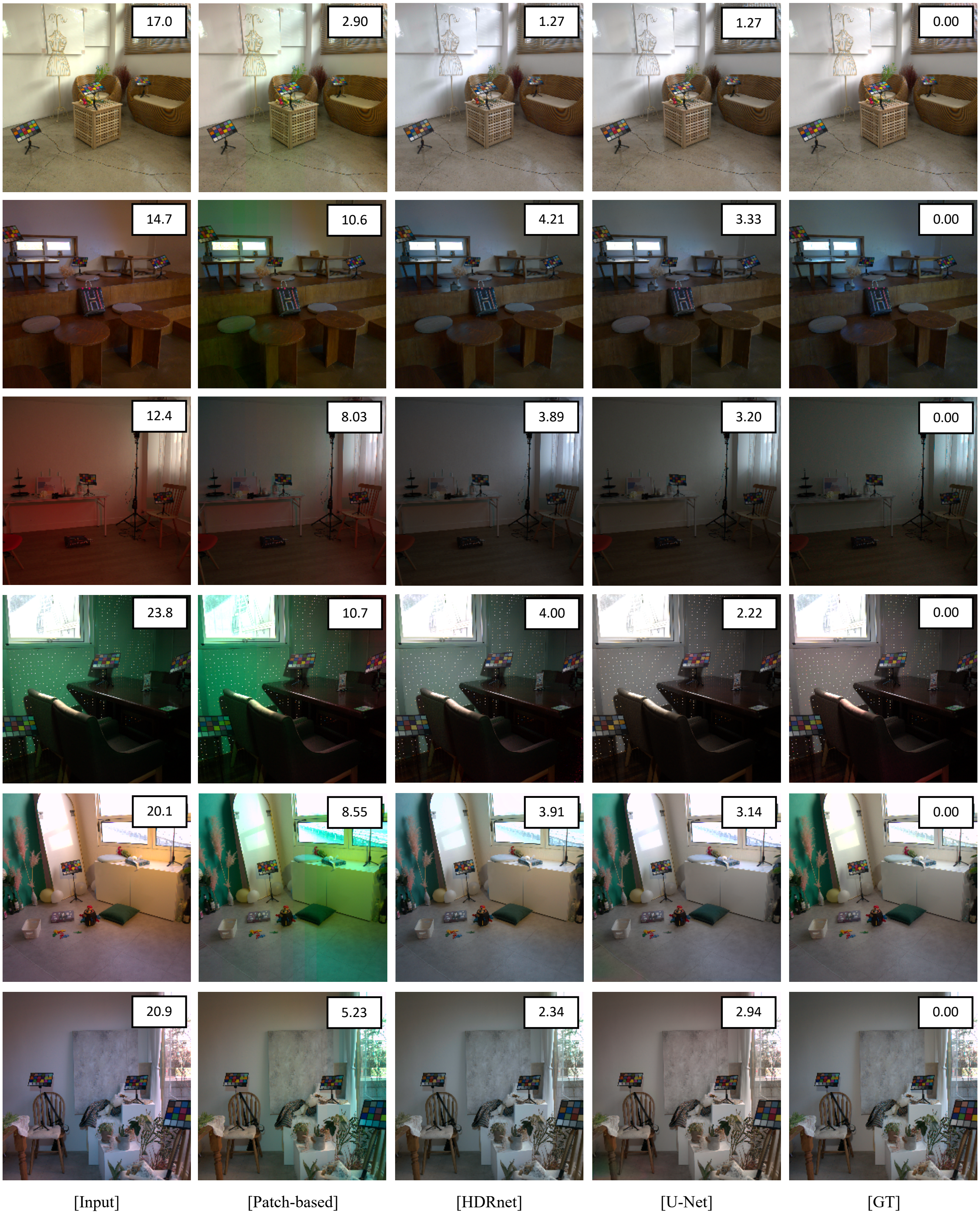
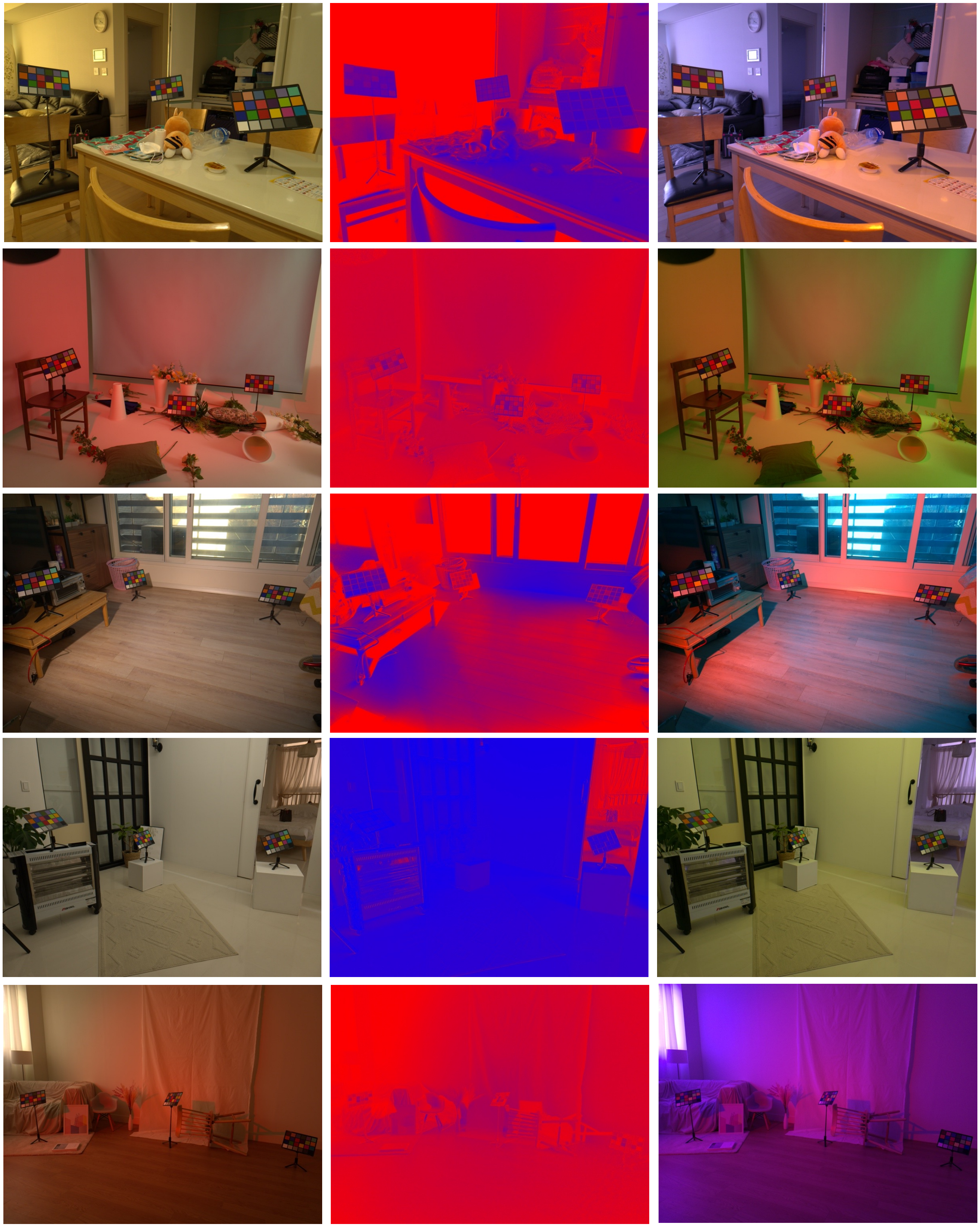


Figure 3. Visualization of various white balanced results using the patch-based model (second column), HDRnet (third column), and U-Net (fourth column). We also add input (first column) and ground truth white balanced images (last column). The first, middle, and last two rows contain images captured by Samsung Galaxy Note 20 Ultra, Sony $\alpha 9$, and Nikon D810. Illumination mean angular errors (MAE) are provided for the reference.

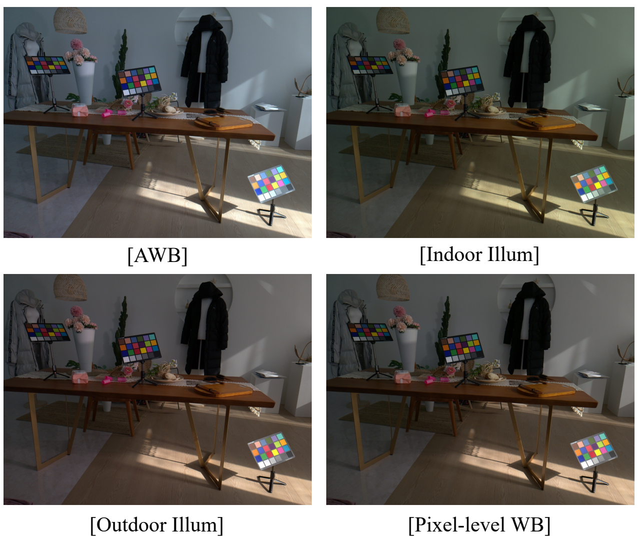


[Original scene] [Illuminant coefficient map] [Pixel-level relighted scene]

Figure 4. Pixel-level religted images. Since LSMI supports pixel-level mixture ratio, which named illuminant coefficient map, we can freely manipulate the color of light sources while maintaining and reflecting the scene geometry.

[Input] [HDRnet] [U-Net] [GT]

Figure 5. Failure cases of U-Net and HDRnet. Since we use basic architecture without any structural modifications, there are some failure cases.



Q1. *Choose the image that looks to be taken under white illumination.*

Q2. *Choose the image that you prefer the most based on your aesthetic criteria.*

Figure 6. Sample images used in the Amazon Mechanical Turk user study and the survey questions.