Illumination Spectrum Estimation for Multispectral Images via Surface Reflectance Modeling and Spatial-Spectral Feature Generation

Supplementary Material

S1. Comprehensive evaluation on the BeyondRGB dataset

We provide detailed experimental results presented in Sec. 4.2. Following standard practices in color constancy research, we evaluate the 25th, 50th (median), and 75th percentile for each result, along with the mean and standard deviation. These statistics are reported in Tab. S1 for PWIR, BeyondRGB, and the proposed technique. The error rates in different percentile values indicate that our model achieves stable and reliable illumination estimation overall.

Dataset	Method	$\Delta A_{MS} \downarrow$ mean	25%	median	75%	std
Lab	PWIR [54]	27.10	18.55	23.91	35.68	10.86
	BeyondRGB [14]	5.92	4.04	5.39	8.01	2.92
	Ours	**3.65**	**1.73**	**2.70**	**4.97**	**2.72**
Field	PWIR [54]	16.31	11.61	15.17	21.82	6.07
	BeyondRGB [14]	7.22	3.31	6.14	9.89	5.54
	Ours	**5.97**	**3.22**	**5.04**	**7.39**	**3.90**

Table S1. Performance evaluation on BeyondRGB dataset.

S2. Further ablation study on network architecture

We developed and tested several modules thoroughly to achieve a better trade-off between accuracy and complexity for the implementation of our approach.

Implementation of learnable SU. We present comparative experiments with a fully learnable model using a Transformer and deeper CNN layers as in Zeng et al. [S1]. The results showed that, while the deeper structure slightly reduced AE, it demanded substantially higher computational complexity. In fact, Zeng's method was developed for HSI denoising to preserve spatial details. These validated our design choices for the ISE.

Learnable SU model	mean-ΔAE (°)	GPU Memory (GB)	# of Param.	inf. time (s)
Transformer SU	4.22	18.77	7,329,238	0.027
Ours	4.40	14.96	6,795,365	0.022

Table S2. Comparison of learnable SU.

Cross-attention (CA) in feature fusion. On top of the standard CA, we achieved more effective feature integration for the interaction between data-driven and physics-driven features through the GRU. This is accomplished by adding abundance (U) to the attention as in Eq. (9). This showed improvements in mean-ΔA_{MS} compared to when U was removed in Eq. (9) (Ours: 4.40°, wo/ U: 4.92°), supporting our contribution of the spatial-spectral features.

Such consideration in task-specific designs enhanced the trade-off.

S3. Ablation study on endmembers

We evaluated the performance with varying numbers of endmembers on the BeyondRGB dataset. The SU block learns K unique spectral patterns inherent in the image, while increasing K would result in longer inference time. The inference time was computed based on the first batch with a batch size of 50. The results are presented in Sec. S3. In the BeyondRGB dataset, K = 5 provides the optimal balance. Although K = 7 shows marginally better performance with a 0.03 improvement, K = 5 reduces computational time by 22% (from 3.27s to 2.54s) while maintaining nearly equivalent performance. When K exceeds to 7, the performance drops to 4.71. In KAUST dataset, we observed lower K values exhibited better performance. The optimal K can be selected based on the spectral characteristics of the target dataset.

Dataset	# of Endmembers (K)					
BeyondRGB	3	4	5	6	7	8
mean-$\Delta A_{MS} \downarrow$	4.72	4.76	4.40	4.83	4.37	4.71
Inf. Time (s)	2.37	2.50	2.54	2.84	3.27	3.42

Table S3. Performance comparisons and inference time for different numbers of endmembers (K) in our SU block.

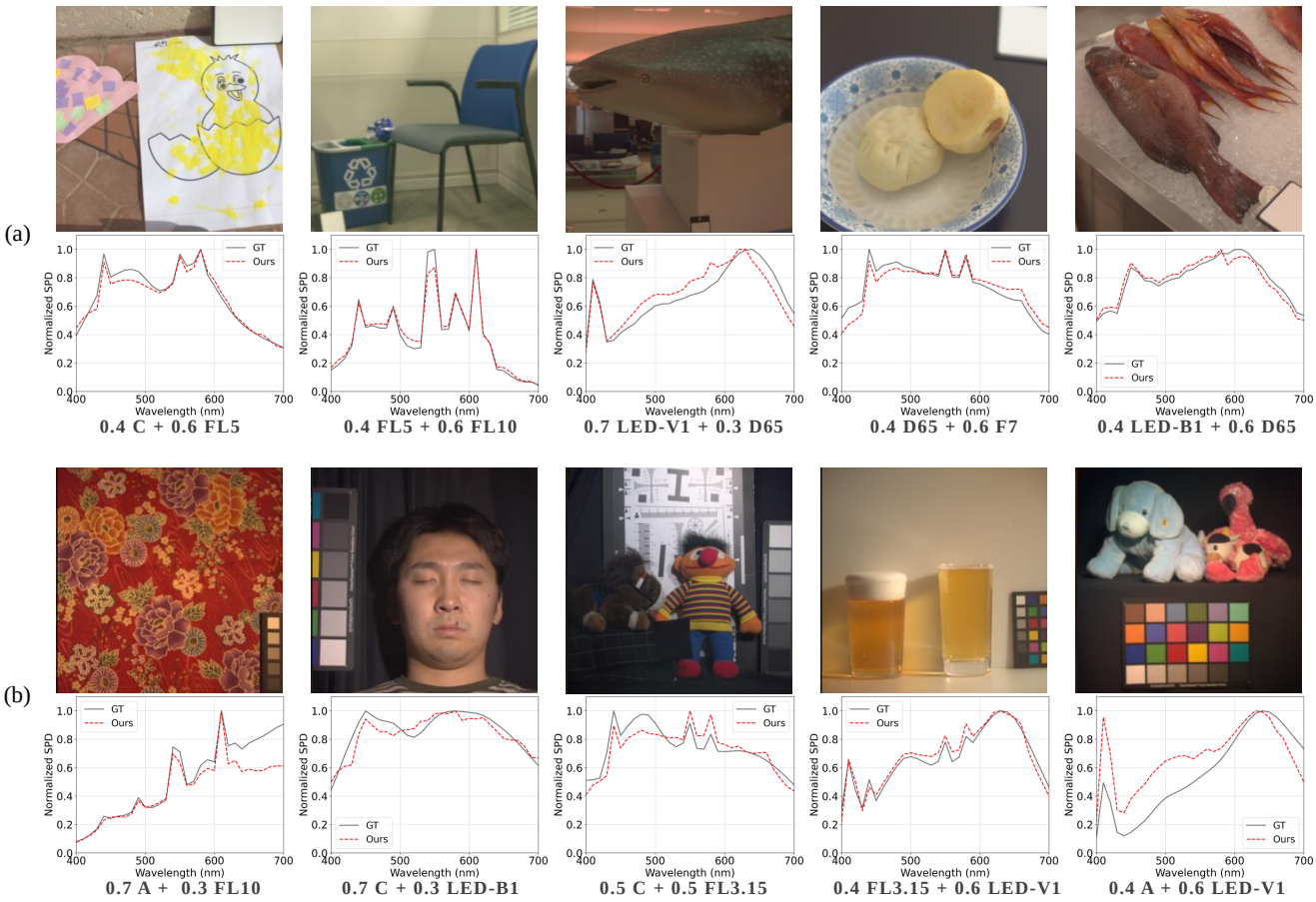

Figure S1. Visualization of estimated illumination spectrum for KAUST and CAVE datasets under random illumination conditions. (a) and (b) show results from the KAUST and CAVE datasets, respectively. The top row shows rendered input RGB images, and the bottom row presents the corresponding illumination spectra. Gray and red lines represent the ground truth and our estimated spectra, respectively. The two illuminants used and their mixing ratios are indicated below each illumination plot.

S4. Evaluation on random mixed illuminants

We demonstrate the superior performance of the proposed technique compared to other methods under various illuminations. For this, we use synthetic illuminations on the KAUST and CAVE reflectance dataset. Specifically, we equally mix two standard illuminants during training, and, for testing, we generate more diverse lighting conditions using varying mixing ratios (0.3-0.7). This setup allows us to assess how effectively our model can adapt to different illumination combinations. CAVE dataset is used only for the tests. As shown in Tab. S4, we observe slight increase in ΔA_{MS} values compared to standard illumination results (from 5.55 to 6.25 for KAUST and from 5.93 to 6.52 for CAVE, increases of 0.70 and 0.59, respectively), the proposed model provides stable performance even under more diverse illumination conditions. The qualitative results are illustrated in Fig. S1 for KAUST dataset.

Dataset	Illumination	$\Delta A_{MS} \downarrow$ mean	25%	median	75%	std
KAUST	Standard (Sec. 4.1)	5.55	2.89	5.01	7.51	3.29
	Random	6.25	4.35	5.85	7.88	2.77
CAVE	Standard (Sec. 4.1)	5.93	2.88	5.13	8.21	3.90
	Random	6.52	4.29	5.99	8.30	3.02

Table S4. Performance of estimated illumination spectrum for KAUST and CAVE dataset under random illumination conditions.

S5. Results of white balancing

As a downstream task, we performed white balancing (WB) on BeyondRGB dataset using the estimated illumination spectra from two different methods, PWIR and Ours. Since our task is bound to illumination estimation for MS images, both the input images and illumination are in spectral form. To properly handle the spectral nature of both the images and illumination, the white balancing process was conducted in the XYZ color space, which provides a device-

independent representation and maintains a direct relationship with the spectral power distributions through the CIE standard observer color matching functions (CMFs). This color space enables accurate chromatic adaptation while preserving the physical meaning of the spectral data, followed by visualization in sRGB format. For color space conversion, the MS images were transformed to XYZ and RGB domains following the conversion methodology proposed in the BeyondRGB dataset paper. The illumination spectra were converted to XYZ values using CMF their corresponding RGB values were obtained using the Python colour package.

The white balance results with the estimated illumination are shown in Fig. S2 for Lab images and Fig. S3 for Field images. The closer the color appearance is to the GT image, the better the illuminant estimation. Across diverse scenes, the proposed technique consistently achieves color correction results that most closely align with the GT images. Despite the absence of RGB-specific constraints in the training process, our method demonstrates high accuracy in the RGB domain. These findings suggest that precise illumination estimation in the spectral domain can effectively extend to improved white balancing performance in RGB images.

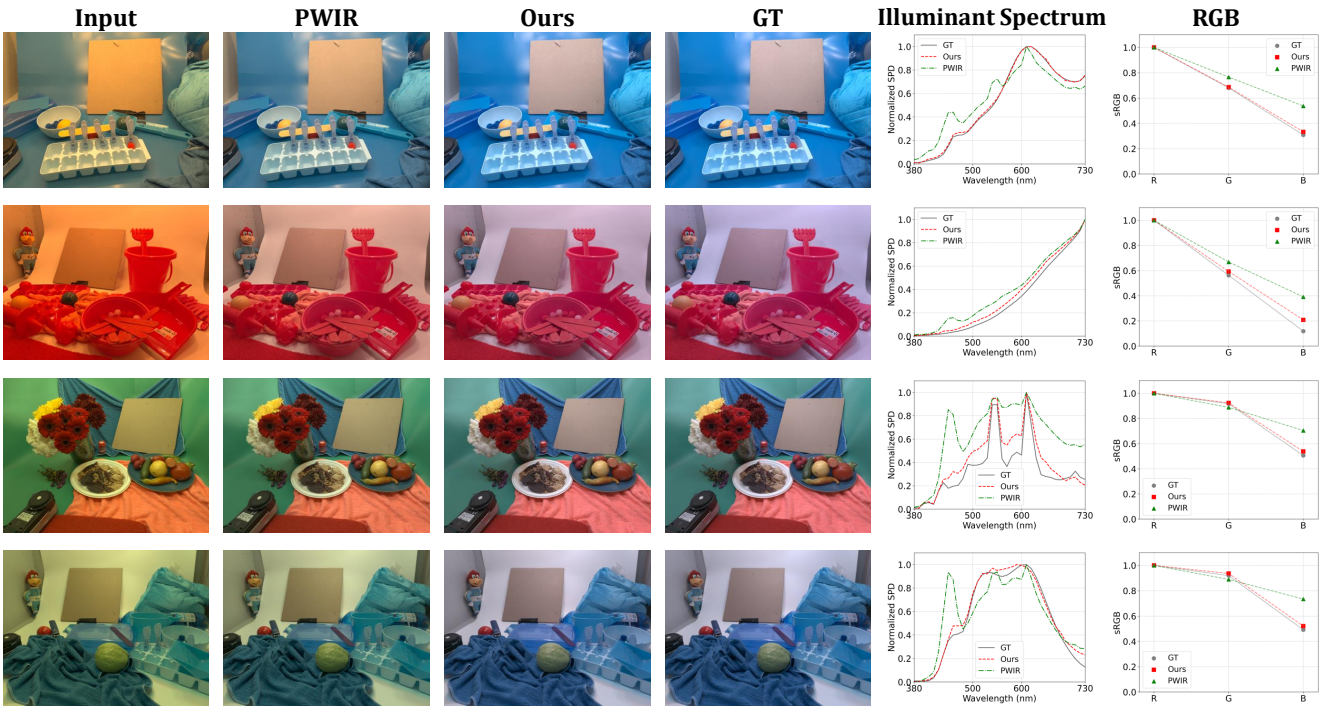

Figure S2. White balancing results on Lab images. The columns from left to right present: input images (1st), white balanced results using illuminants estimated by PWIR (2nd) and our proposed method (3rd), white balanced results using ground truth illumination (4th, GT). The rightmost columns (5th, 6th) display the spectral power distributions of ground truth and estimated illuminants from PWIR(Green) and Ours(Red) methods along with their corresponding RGB representations.

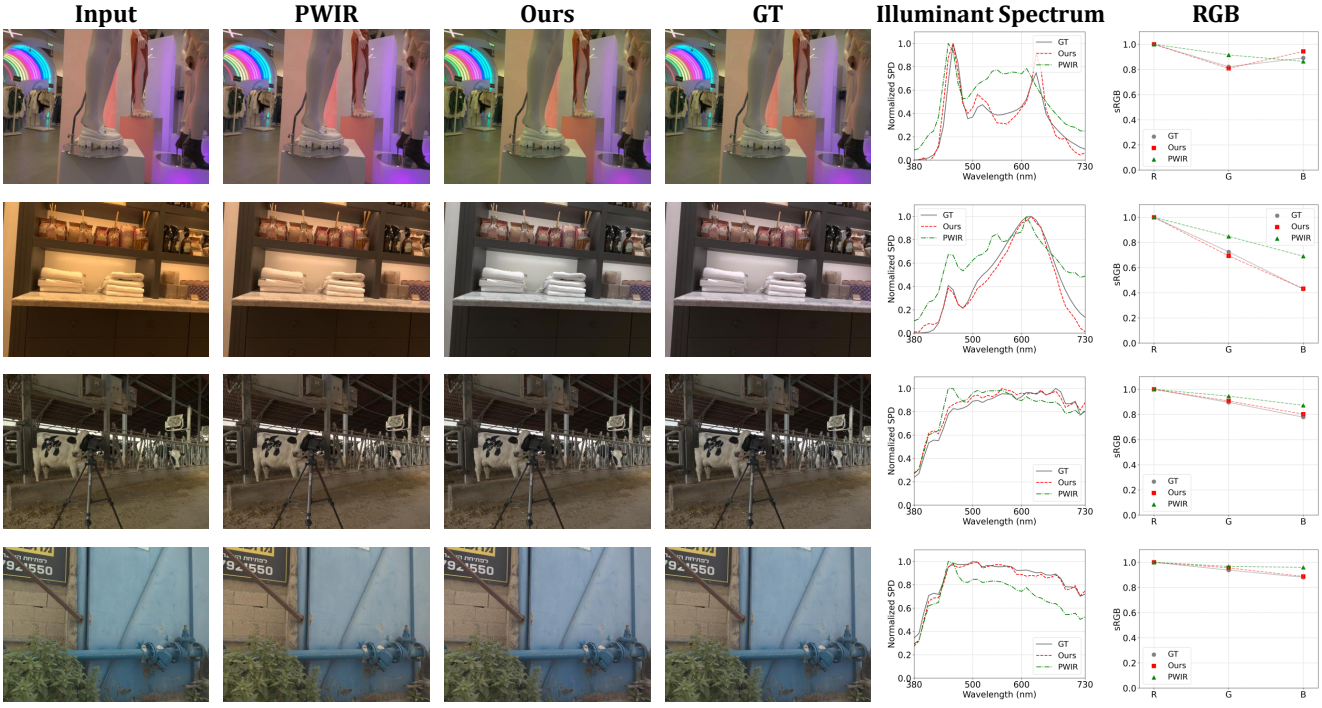

Figure S3. White balancing results on Field images. The column layout follows the same structure as Fig. S2.

Supplementary References

S1 Haijin Zeng, Jiezhang Cao, Kai Zhang, Yongyong Chen, Hiep Luong, and Wilfried Philips. Unmixing diffusion for self-supervised hyperspectral image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27820–27830, 2024.