Supplementary Material: Joint RGB-Spectral Decomposition Model Guided Image Enhancement in Mobile Photography

Kailai Zhou[1], Lijing Cai[1], Yibo Wang[1], Mengya Zhang[1], Bihan Wen[3], Qiu Shen[†1,2], and Xun Cao[1,2]

[1] School of Electronic Sciene and Engineering, Nanjing University, Nanjing, China
[2] Key Laboratory of Optoelectronic Devices and Systems with Extreme Performances of MOE, Nanjing University, Nanjing, China
[3] Nanyang Technological University, Singapore
{calayzhou, cailijing, ybwang, 522023230161}@smail.nju.edu.cn, bihan.wen@ntu.edu.sg, {caoxun, shenqiu}@nju.edu.cn

1 More Details about Mobile-Spec Dataset

1.1 Mobile-Spec Dataset Overview

In Figure 1, we present a diverse collection of representative samples from the Mobile-Spec dataset. It can be found that the Mobile-Spec dataset comprises two parts. The first encompasses images captured by the smartphone and hyperspectral camera, while the second comprises images related to shading, reflectance, and material segmentation. The 16-bit input paired with corresponding 8-bit target images are constructed for the task of tone enhancement, predominantly featuring outdoor scenes with high dynamic range. It should be noted that both 16-bit input and 8-bit target images are in the sRGB color space, and we focus on the tone enhancement task rather than learning the whole pipeline from raw data to final output. The hyperspectral images, obtained using the GaiaSky-mini2 [1], are intended to overcome the limitations in the spectral imaging capabilities of mobile devices. On the right side of Figure 1, we showcase shading, reflectance, and material segmentation images for each sample. The shading and material segmentation images serve as targets for the prediction of the joint RGB-Spectral decomposition model. Notably, the shading images are averages of the 850-1000 nm bands in hyperspectral images. It reveals that the near-infrared bands can approximately delineate the distribution of shadows and illumination in outdoor environments. The material segmentation images, on the other hand, are meticulously labeled by human annotators. Our Mobile-Spec dataset aims to establish a high-quality foundation, offering insights and laying the groundwork for further exploration of spectral information in mobile photography.

1.2 Alignment of Dual-Camera System

Dual Camera System. Figure 2(a) showcases our dual-camera setup, featuring the high-end commercial smartphone and the GaiaSky-mini2 hyperspec-

tral camera [1], which operates on a scanning-based imaging principle to guarantee the quality of the captured hyperspectral images. The two cameras are positioned in close proximity to each other to minimize differences in viewing angles as much as possible. During the capture of each scene, both cameras are set to their default shooting settings and operate synchronously.

Image Matching. As observed in Figure 2(b), the RGB images captured by the smartphone own a larger field of view and higher spatial resolution ($4096 \times 3072 \times 3$), compared to the hyperspectral images, which feature a smaller field of view and resolution ($1057 \times 960 \times 176$). To facilitate the study of tone enhancement tasks, we align the RGB images to the hyperspectral images. Specifically, we first convert the hyperspectral images into the pseudo-RGB format. Then, using the SIFT algorithm [14], we calculate feature points between the smartphone-captured RGB images and the pseudo-RGB images from the hyperspectral data. The feature points are matched to compute the transformation homography matrix between the two images. Finally, using the homography matrix, we perform an affine transformation on the smartphone-captured RGB images, resulting in aligned pairs of RGB and hyperspectral images, as illustrated in Figure 2(c).

Filtering of Mismatched Samples. Due to factors such as depth of field and errors in feature point matching, there can be pixel alignment errors between RGB and hyperspectral image pairs. We filter out samples with significant registration errors. To visually assess the alignment of each sample, we substitute the R channel in the smartphone-captured RGB image with the shading image from the hyperspectral data. This method allows for a visual evaluation of the image registration quality. As shown in Figure 3(b), samples that exhibit artifacts and misalignments are excluded. Conversely, samples from our Mobile-Spec dataset demonstrate high alignment accuracy, as displayed in Figure 3(a).

Compared to the RGB images ($1057 \times 960 \times 3$), the Lr-MSI possesses a markedly limited spatial resolution ($16 \times 16 \times 10$). Given that the spatial resolution setting for Lr-MSI is very small, the alignment errors can be negligible in the JDM-HDRNet. We believe that high-precision RGB-hyperspectral image pairs are meaningful for many other tasks, such as joint RGB-spectral pansharpening [13], reconstruction [6], segmentation [10], and illumination estimation [15]. Therefore, we ensure that the Mobile-Spec dataset maintains minimal alignment errors.

1.3 Material Segmentation

Figure 4 showcases material segmentation images from our dataset, wherein the segmentation categories have been designated as plant, trunk, building, road, sky, and others, reflecting the most commonly encountered subjects in outdoor scenes. It is evident that our semantic segmentation annotations are of exceptionally high granularity, particularly notable in the detailing of individual leaves within the plant area. Our Mobile-Spec dataset is also expected to advance research in the joint analysis of RGB and spectral images for material segmentation.

1.4 The Approximated Shading Prior

Since shading GT of complex outdoor scenes is difficult to obtain, the near-infrared images serve as the guide map to approximate shading instead of real shading GT. We consider four methods for approximating shading: (a) Intrinsic decomposition of hyperspectral images [5]: This method fails to handle complex outdoor scenes. (b) Implicit learning (e.g., Retinexformer [3]): This method produces results more like the grey image instead of shading. (c) Intrinsic decomposition of RGB images: PIE-Net [9] relies heavily on training data and lacks interpretability, which is unstable and may produce artifacts and over smoothness. (d) Near-infrared images [7]: Figure 5 shows near-infrared images stands out as reliable guide map for approximating shading, since they are less sensitive to reflectance variations [7]. So we adopt near-infrared images for shading estimation as a simple yet effective way.

Figure 6 delineates the spectral reflectance values sampled across 24 colorants on a color checker. The spectral curves exhibit a trend towards flattening and convergence with increasing wavelengths, suggesting the near-infrared spectrum's viability as a proxy for shading [7]. It is pertinent to note that the precise estimation of shading is not the focal point of this paper; rather, our objective is to harness the shading prior approximated by near-infrared bands to enhance the image quality in tone enhancement tasks.

As illustrated in Figure 8, the spectral curves of sunlight and LED light source are captured by our hyperspectral camera with the white board. The sunlight encompasses a broad and continuous spectrum across the visible range, extending into the ultraviolet (UV) and infrared (IR) regions, with its intensity being relatively uniform across the visible spectrum, albeit with minor variations. Conversely, the spectrum of the LED light source is characterized by a pronounced peak in the blue region, with its intensity swiftly diminishing at near-infrared wavelengths.

Furthermore, we have drawn spectral curves of 24 distinct colors on a color checker using our hyperspectral camera under two light sources. As evident from Figure 9(a), for the illumination of sunlight within the visible light spectrum (400-750nm), different colors display unique spectral curve characteristics. However, in the near-infrared spectrum (850-1000nm), the spectral curves of different colors tend to converge, aligning with the conclusion drawn in Figure 6. However, under LED illumination, various colors demonstrate no significant response within the near-infrared spectrum. Given that most indoor lighting sources are LED lights, the assumption of using the near-infrared spectrum as a proxy for shading does not hold indoors. Consequently, the samples in our Mobile-Spec dataset are exclusively captured in outdoor settings. As shown in Figure 7, under the single outdoor illumination of sunlight, the spectral curve variations of different materials in the near-infrared spectrum (850-1000nm) tend to be consistent. Notably, the divergence in spectral response, both among various materials and across distinct pixels of the same material, is more pronounced in terms of intensity variation. This suggests the near-infrared band may serve as an approximation for the shading prior in the outdoor environment.

Table 1: Objective quality assessment of tone-mapped images across four representative datasets. Our Mobile-Spec dataset achieves comparable structural fidelity with PPR10K [12] and MIT FiveK [2].

	HDR+ [11]	FiveK [2]	PPR10K [12]	Mobile-Spec
Structural Fidelity↑ [17]	0.130	0.150	**0.193**	0.176

Table 2: Comparisons with previous hyperspectral datasets.

	Harvard [4]	CAVE [8]	KAIST [16]	PaviaU	Mobile-Spec
scenes	75	32	30	1	200
spatial (x-y)	1392×1040	512×512	2704×3376	610×340	1057×960
spectral (λ,nm)	420-720	400-700	400-700	430-860	400-1000
channels	30	31	31	103	176
segmentation	×	×	×	✓	✓
aligned RGB	×	×	×	×	✓

1.5 8-bit Target Image

Since tone-mapped targets are highly subjective, as different individuals have varying aesthetic preferences, we employ a commercial privacy model to generate targets of Mobile-Spec dataset that integrate the aesthetics of multiple experts. This model is trained by a large-scale and high-quality commercial dataset, which is elaborately adjusted by professional photographers and artists. Then we meticulously filter the target images through subjective assessments, considering factors such as chromatic aberration, sharpness, noise and artifacts. Samples that do not meet the criteria are excluded. We adopt the structural fidelity term in TMQI [17] to evaluate the objective quality of the Mobile-Spec dataset. It should be noted that we only reserve the structural fidelity term and the statistical naturalness term is removed, since the tone enhancement images in Mobile-Spec are high dynamic range scenes, which do not comply with statistical naturalness of common pictures. Table 1 shows the Mobile-Spec dataset achieves comparable structural fidelity with PPR10K [12] and MIT FiveK [2], highlighting its potential as a solid foundation for exploring the role of Lr-MSI in the tone enhancement task.

1.6 Comparisons with Previous Hyperspectral Datasets

From Table 2, it can be seen that our Mobile-Spec dataset encompasses more diverse scenes and a broader spectral range compared to previous hyperspectral datasets. Additionally, the Mobile-Spec dataset owns high resolutions in both spatial and spectral dimensions. We provide aligned RGB images captured by smartphones and meticulously labeled segmentation maps. Consequently, the Mobile-Spec dataset is anticipated to contribute to related fields such as hyperspectral reconstruction, segmentation, and pansharpening.

2 Qualitative Comparisons

2.1 Test Set of Mobile-Spec

Figure 10 presents additional comparative visualization of the results between JDM-HDRNet and other methods. From these samples, we can draw the following conclusions: (1) In areas with drastic local brightness variations, such as sunlight-dappled foliage exhibiting diverse light and shadow distributions, JDM-HDRNet adaptively accommodates high-dynamic-range scene tone adjustments by isolating the shading component. (2) Incorporating the reflectance prior of Lr-MSI with an expanded color channel, and designing a specialized grid expert for each material enable JDM-HDRNet to produce tone-enhanced outcomes with minimal color deviation. For instance, the exterior walls of the building in the fifth sample and the red wall in the third sample. Given that the Mobile-Spec dataset predominantly features sky and plant, cooler tones like blue and green dominate most colors, leading to inadequate learning for mapping warmer tones such as red. Introducing additional Lr-MSI can mitigate the imbalance in learning across different colors to some extent. JDM-HDRNet achieves more accurate and aesthetically pleasing colors compared to the competitive methods. The qualitative comparison underscores the efficacy of decomposing Lr-MSI into three components: shading, reflectance, and material semantic priors. This framework offers explicit guidance for tone enhancement and overcomes the intrinsic complexity of spectral images. Through the exploration of Lr-MSI in the tone enhancement task, we aim to lay the foundation for the broader application of spectral information in mobile photography.

2.2 Unseen Samples

To validate the generalization, we capture extra unseen test samples from entirely new locations, which contain scenes outside of Mobile-Spec (e.g., pool, yellow leaves). Qualitative results in Figure 11 show JDM-HDRNet generates more vivid images with a natural appearance than HDRNet. The benefits of Lr-MSI lie in the following aspects: (a) Enhanced dynamic range with S: Shading prior enhances adaptability in dealing with localized brightness variations. (b) More accurate color with R: JDM-HDRNet generates more realistic color than HDRNet on unseen images. (c) Context consistency with M: Introducing M prior reduces color inconsistency within the same context region. Explicit priors help constrain JDM-HDRNet outputs, enhancing generalization and robustness on unseen images.

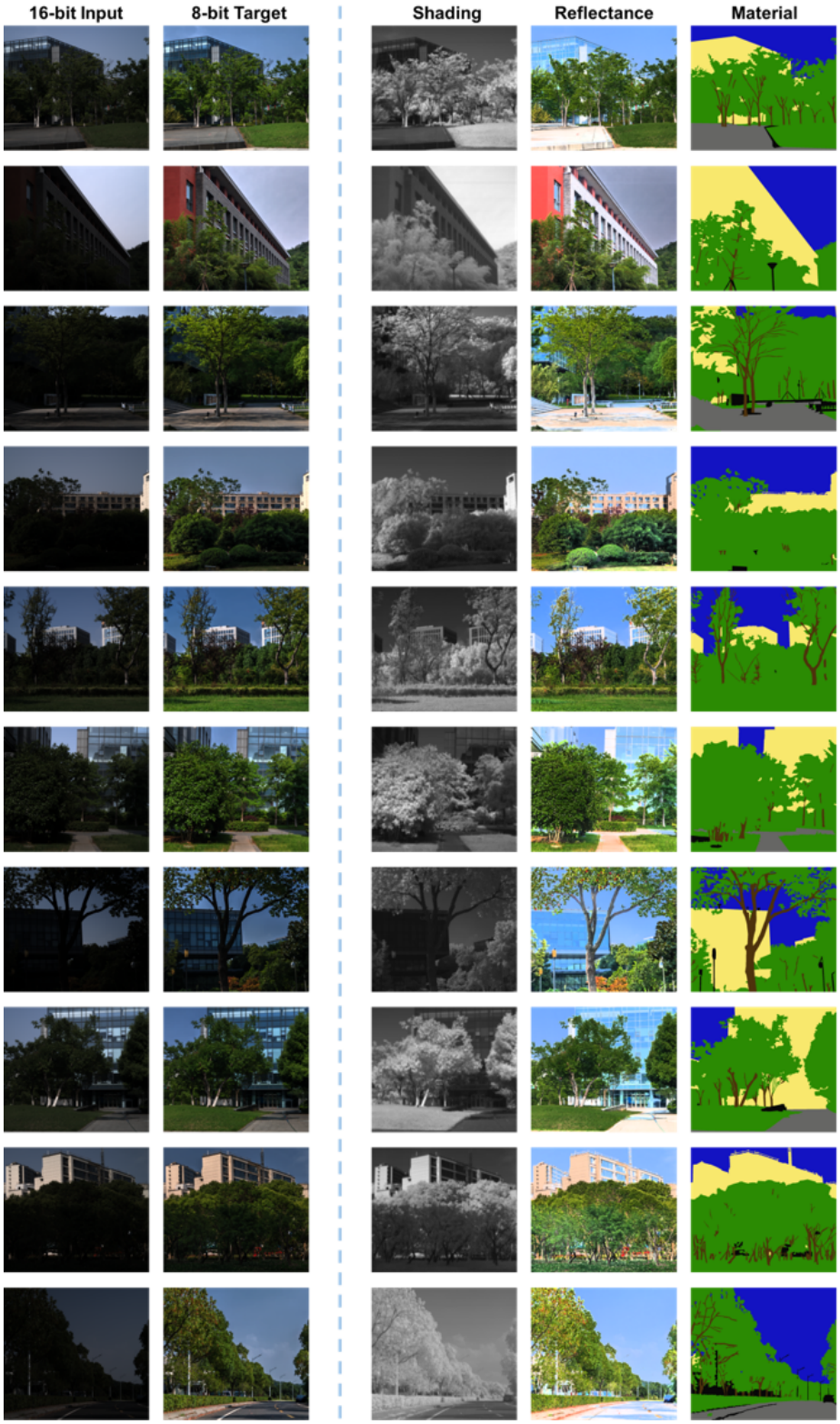

Fig. 1: More visualized samples of the Mobile-Spec dataset. The 16-bit RGB images are linear tone-mapped for visualization.

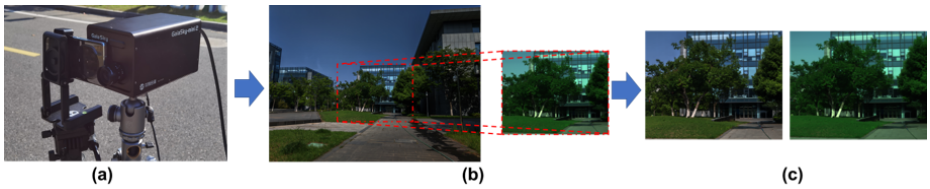

Fig. 2: (a) The dual camera system which consists of the high-end commercia smartphone and the GaiaSky-mini2 hyperspectral camera [1]. (b) Image Matching: the overlapping region is detected by the SIFT descriptor [14], then the affine transformation is performed on the smartphone-captured RGB image to align with the pseudo-RGB image from the hyperspectral data. (c) The aligned pair of RGB and hyperspectral images, the hyperspectral image is transformed into the pseudo-RGB image.

Fig. 3: We substitute the R channel in the smartphone-captured RGB image with the shading image, which allows for a visual evaluation of the image registration quality. In each sample, the first column is the smartphone-captured RGB image, the second column is the pseudo-RGB image from the hyperspectral data, the third column is the fused image which replaces the R channel with the shading image. (a) Samples in our Mobile-Spec dataset demonstrate high alignment accuracy. (b) Samples that exhibit artifacts and misalignments are discarded in the dataset filtering procedure.

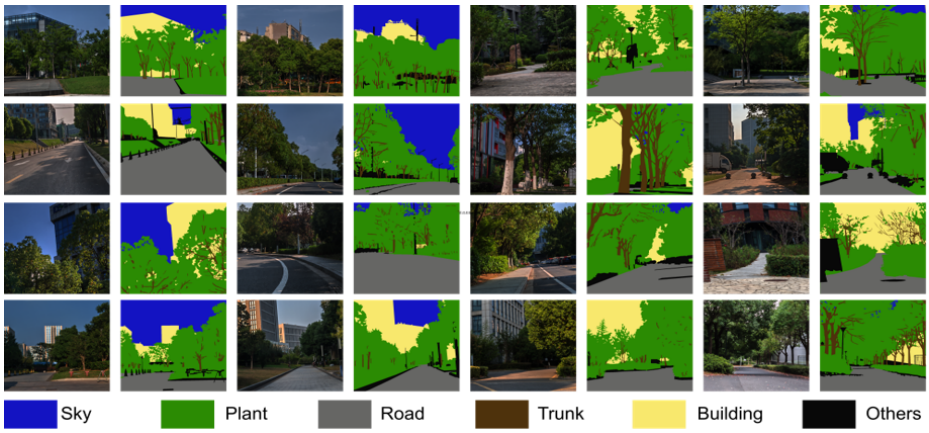

Fig. 4: The Mobile-Spec dataset is meticulously labeled by human annotators. The segmentation categories are the sky, plant, road, trunk, building, and others, which reflect the most commonly encountered subjects in outdoor scenes.

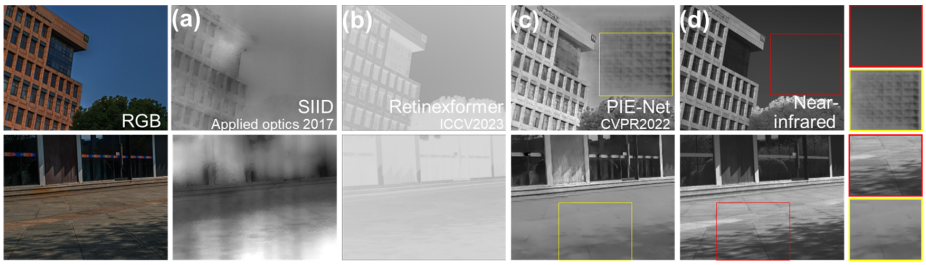

Fig. 5: Since it is difficult to obtain the shading GT for outdoor scenes, we consider four ways to approximate the shading term: (a) Intrinsic decomposition of hyperspectral images [5]. (b) Implicit learning [3]. (c) Intrinsic decomposition of RGB images [9]. (d) Near-infrared images [7]. The near-infrared images serves as a reliable guide map to estimate shading.

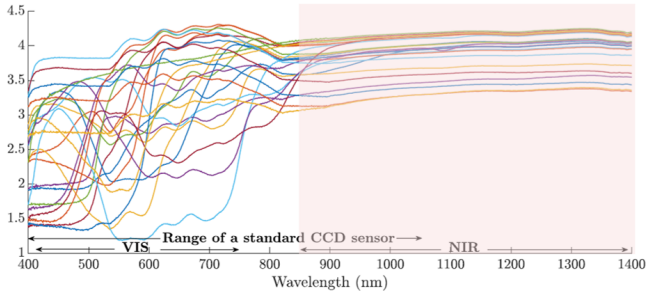

Fig. 6: The reflectance versus wavelength curves for 24 colourants on a colour checker. Different colors exhibit a trend towards flattening and convergence with increasing wavelengths in the near-infrared band. This figure is from [7].

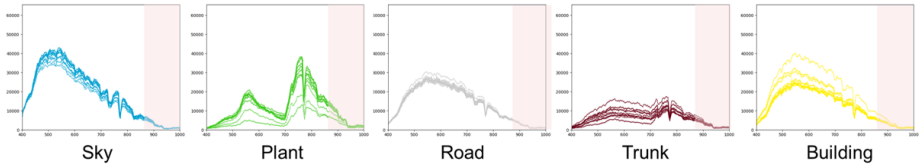

Fig. 7: Under the single outdoor illumination of sunlight, the spectral curve trends of different materials in the near-infrared spectrum (850-1000nm) tend to be consistent.

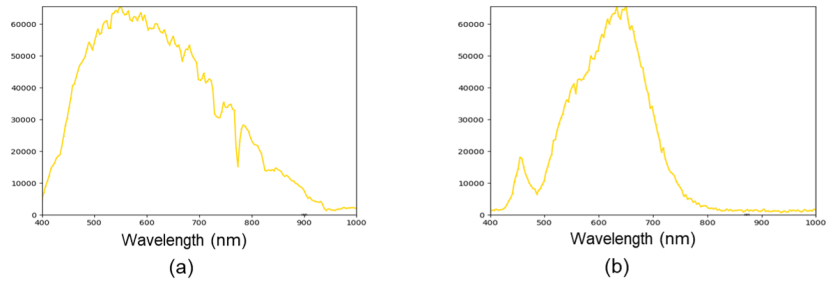

Fig. 8: (a) Sunlight. (b) LED light source. The spectral curves of sunlight and LED light source are captured by our hyperspectral camera with the white board. The spectrum intensity of LED light source swiftly diminishing at near-infrared wavelengths. While the sunlight exhibits a broad and continuous spectrum.

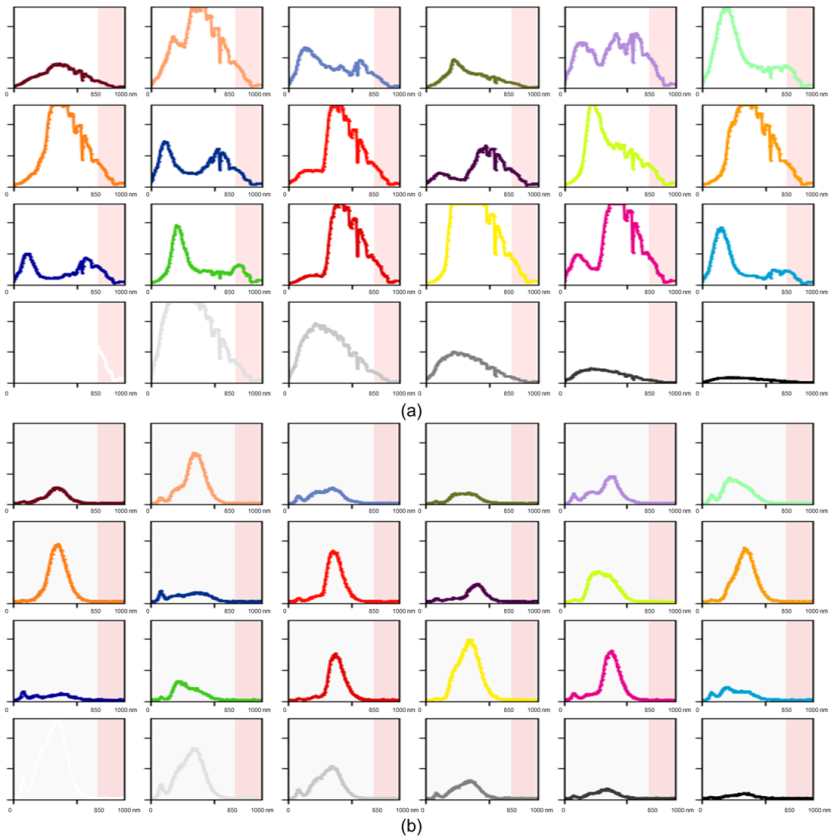

Fig. 9: (a) Sunlight. (b) LED light source. The spectral curves of 24 distinct colors on a standard color checker under illuminations of sunlight and LED light source are captured by our hyperspectral camera.

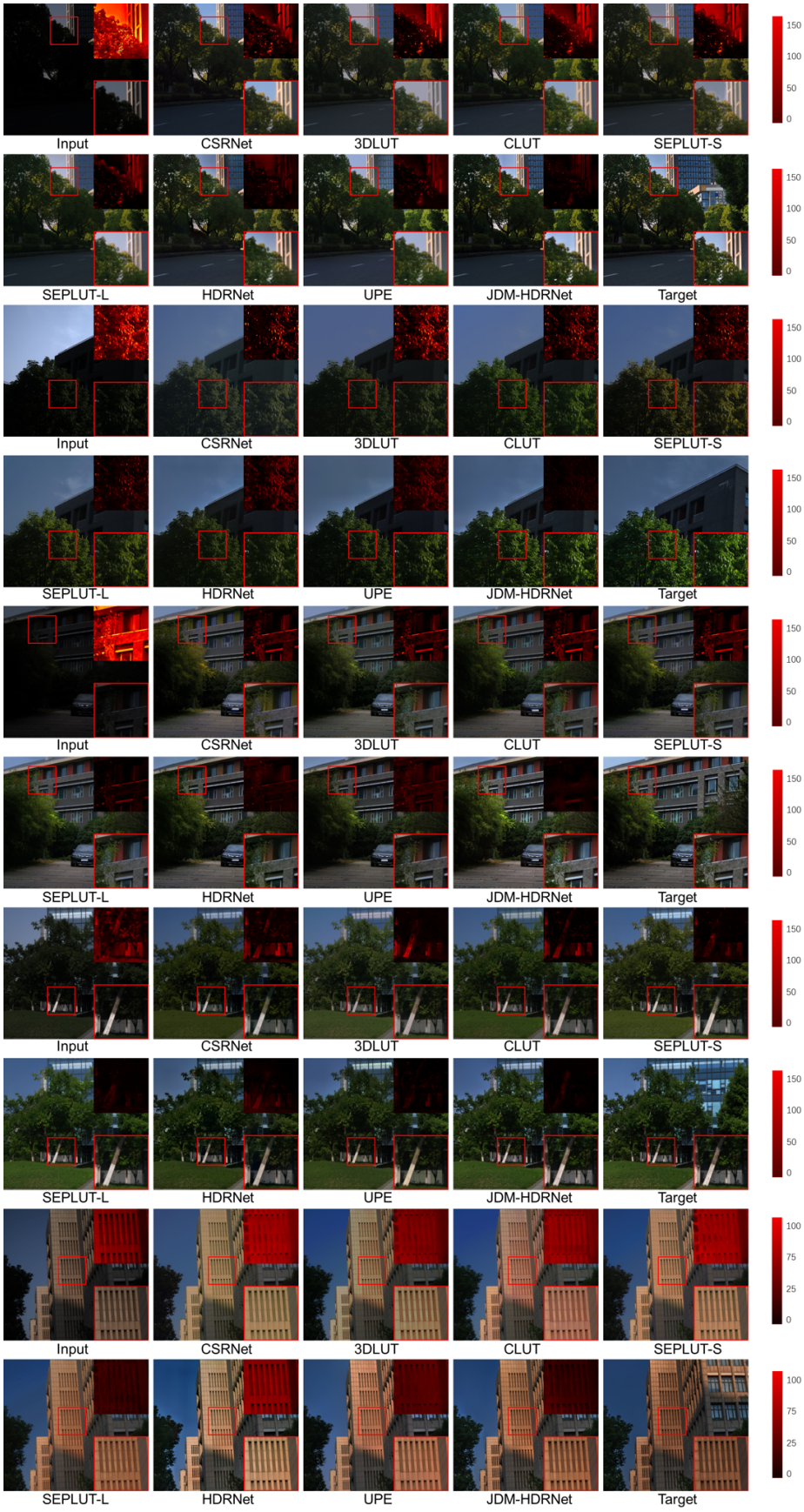

Fig. 10: More qualitative comparisons on the Mobile-Spec dataset.

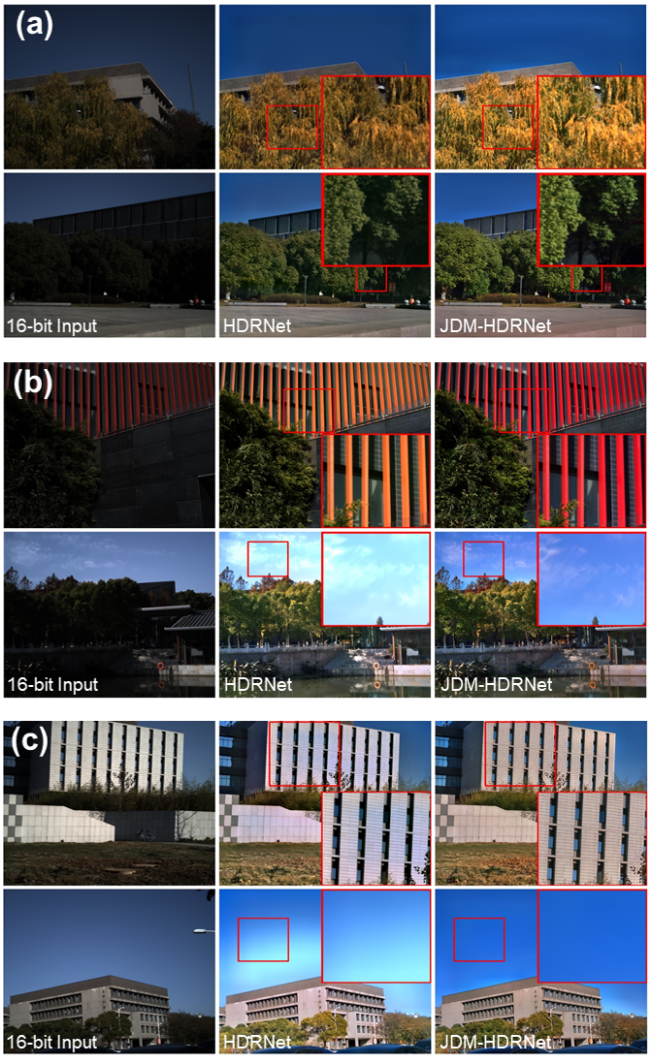

Fig. 11: Evaluations on unseen samples captured from new locations, which contain scenes out of the Mobile-Spec dataset (e.g., pool, yellow leaves). Benefits of Lr-MSI: (a) S: enhanced dynamic range; (b) R: more accurate color; (c) M: context consistency. Explicit priors help constrain JDM-HDRNet outputs, enhancing generalization and robustness on unseen images.

References

1. Gaiasky-mini hyperspectral imaging camera. https://www.dualix.com.cn/en/Goods/desc/id/123/aid/954.html
2. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: CVPR 2011. pp. 97–104. IEEE (2011)
3. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12504–12513 (2023)
4. Chakrabarti, A., Zickler, T.: Statistics of real-world hyperspectral images. In: CVPR 2011. pp. 193–200. IEEE (2011)
5. Chen, X., Zhu, W., Zhao, Y., Yu, Y., Zhou, Y., Yue, T., Du, S., Cao, X.: Intrinsic decomposition from a single spectral image. Applied optics **56**(20), 5676–5684 (2017)
6. Chen, Y., Wang, Y., Zhang, H.: Prior image guided snapshot compressive spectral imaging. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
7. Cheng, Z., Zheng, Y., You, S., Sato, I.: Non-local intrinsic decomposition with near-infrared priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2521–2530 (2019)
8. Choi, I., Kim, M., Gutierrez, D., Jeon, D., Nam, G.: High-quality hyperspectral reconstruction using a spectral prior. Tech. rep. (2017)
9. Das, P., Karaoglu, S., Gevers, T.: Pie-net: Photometric invariant edge guided network for intrinsic image decomposition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19790–19799 (2022)
10. Habili, N., Kwan, E., Li, W., Webers, C., Oorloff, J., Armin, M.A., Petersson, L.: A hyperspectral and rgb dataset for building façade segmentation. In: European Conference on Computer Vision. pp. 258–267. Springer (2022)
11. Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics (ToG) **35**(6), 1–12 (2016)
12. Liang, J., Zeng, H., Cui, M., Xie, X., Zhang, L.: Ppr10k: A large-scale portrait photo retouching dataset with human-region mask and group-level consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 653–661 (2021)
13. Loncan, L., De Almeida, L.B., Bioucas-Dias, J.M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G.A., Simoes, M., et al.: Hyperspectral pansharpening: A review. IEEE Geoscience and remote sensing magazine **3**(3), 27–46 (2015)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60**, 91–110 (2004)
15. Thomas, J.B.: Illuminant estimation from uncalibrated multispectral images. In: 2015 Colour and Visual Computing Symposium (CVCS). pp. 1–6. IEEE (2015)
16. Yasuma, F., Mitsunaga, T., Iso, D., Nayar, S.K.: Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. IEEE transactions on image processing **19**(9), 2241–2253 (2010)
17. Yeganeh, H., Wang, Z.: Objective quality assessment of tone-mapped images. IEEE Transactions on Image processing **22**(2), 657–667 (2012)