LRDUN: A Low-Rank Deep Unfolding Network for Efficient Spectral Compressive Imaging

Supplementary Material

1. QR Retraction in $\mathrm{ProxyNet}_{\mathrm{E}}$

The low-rank decomposition of a tensor $\mathcal{X}$ into $\mathcal{X} = \mathcal{A} \times_3 \mathbf{E}$, where $\mathcal{A} \in \mathbb{R}^{H \times W \times k}$ and $\mathbf{E} \in \mathbb{R}^{B \times k}$, inherently suffers from scale ambiguity (or indeterminacy). Specifically, for any non-zero scalar $a \in \mathbb{R}$, the pair $\mathbf{E}' = a\mathbf{E}$ and $\mathcal{A}' = \mathcal{A}/a$ constitutes an equally valid decomposition. To resolve this indeterminacy and stabilize the optimization process, we enforce an orthogonality constraint on the spectral basis $\mathbf{E}$, thereby requiring $\mathbf{E}$ to reside on the Stiefel manifold ($\mathbf{E}^T\mathbf{E} = \mathbf{I}$). For the resulting constrained optimization problem, we adopt the widely-used QR retraction method [8] to project the updated basis $\mathbf{E}$ back onto the Stiefel manifold after each iteration. Mathematically, this involves first performing a QR decomposition on the updated $\mathbf{E}$:

$$\mathbf{E} = \mathbf{Q}\mathbf{R},$$

where $\mathbf{Q} \in \mathbb{R}^{B \times k}$ is a column-orthogonal matrix and $\mathbf{R} \in \mathbb{R}^{k \times k}$ is an upper triangular matrix. The retracted spectral basis is then set as $\mathbf{E} = \mathbf{Q}$.

To validate the efficacy of QR Retraction, we compare it against a constraint-free Baseline-3 and a soft orthogonal loss ($\mathcal{L} = \|\mathbf{E}^T\mathbf{E} - \mathbf{I}\|_F^2$). As shown in Table S1, the baseline fails to converge due to numerical instability. The orthogonal loss, acting merely as a soft constraint, yields suboptimal performance as it cannot enforce strict orthogonality. In contrast, QR Retraction guarantees strict orthogonality via manifold optimization, significantly improving both stability and reconstruction quality.

Table S1. Ablation Study on QR Retraction.

Metric	Baseline-3	Orth. Loss	QR Retraction
PSNR (dB)	not converged	38.18	39.44
SSIM	not converged	0.963	0.972

2. SCAB in $\mathrm{ProxyNet}_{\mathrm{A}}$

Convolutional attention mechanisms have proven highly successful in various computer vision tasks, such as image classification [4, 6], semantic segmentation [1, 5], and image resolution [7], owing to their efficient and effective encoding of contextual information. In our LRDUN framework, the $\mathrm{ProxyNet}_{\mathrm{A}}$ module employs a SCAB to efficiently model long-range spatial dependencies within the subspace features. As illustrated in Fig. 3 of the main paper, the SCAB architecture processes input features through two parallel pathways. The first is the **Attention Path**, where the input feature is processed by a `Conv2D` 1×1 layer, followed by a large-kernel `DWConv2D` 11×11 (11×11 Depth-wise Convolution) to capture features with long-range spatial dependencies. The resulting output is passed through a `GELU` activation to generate the spatial attention map. Concurrently, the **Value Path** processes the input feature with a separate `Conv2D` 1×1 layer to produce the value representation. The two representations are then fused via element-wise multiplication ($\odot$), allowing the long-range contextual information from the attention path to selectively modulate the features in the value path. This fused feature is then passed through a final `Conv2D` 1×1 layer, aggregated with the initial input via a residual connection, and finally refined by a ConvFFN block. This design provides an efficient mechanism for modeling long-range dependencies and enhancing feature representations, which is validated by its superior performance (39.44 dB) against other attention mechanisms in our ablation study (Table 3 of the main paper).

3. Comparison with lightweight methods.

To further contextualize the efficiency of our proposed LRDUN method, we compare its parameter count and reconstruction performance against two lightweight methods: BiSRNet [3] and MST++ [2], as listed in Table S2. Remarkably, LRDUN-3stg outperforms MST++ by 3.45 dB while requiring approximately 48% fewer parameters (0.69M vs. 1.33M). Compared to the ultra-lightweight BiSRNet, LRDUN provides a substantial gain of 9.68 dB with a still-modest parameter count. This comparison underscores the superior balance of efficiency and effectiveness offered by LRDUN, delivering high-fidelity reconstruction with a highly optimized model size.

Table S2. Comparison with related lightweight methods.

Method	Params (M)	FLOPS (G)	PSNR (dB)	SSIM
BiSRNet	0.036	1.18	29.76	0.837
MST++	1.33	19.42	35.99	0.951
LRDUN-3stg	0.69	10.26	39.44	0.972

4. Runtime Comparison with SOTA Methods

To further demonstrate the efficiency of our proposed LRDUN method, we compare its inference runtime against five SOTA methods on a single scene with dimensions $256 \times 256 \times 28$. As shown in Table S3, LRDUN achieves the lowest inference latency among all compared methods, fur-

ther validating its superior efficiency in reconstructing high-quality hyperspectral images while significantly reducing computational costs.

Table S3. Avg. Runtime.

Runtime (ms)	85	131	103	218	146	**78**

5. More Visual Comparison Results

To further validate the effectiveness of our proposed LRDUN method, we provide additional visual comparison results on both real and simulated datasets. Figures S1 to S4 present reconstructed results of real-world scenes, showcasing 4 out of 28 spectral channels for each scene. Figures S5 to S8 display reconstructed results of simulated scenes obtained by various DUNs, along with spectral analysis of representative regions. These visualizations further demonstrate the superior reconstruction quality and spectral fidelity achieved by our LRDUN approach.

6. Limitations and Future Work

While our proposed LRDUN method achieves SOTA reconstruction quality with significantly reduced computational costs, several limitations warrant further investigation. First, the current framework is primarily validated on simulated datasets where the sensing model is assumed to be perfectly linear and known, yet real-world optical systems introduce non-linearities, signal-dependent noise, and optical aberrations that the current linear low-rank model may not fully encapsulate. This model mismatch explains the performance gap observed in real CASSI measurements, where reconstruction quality is less competitive due to calibration errors and hardware artifacts, which remains an inherent challenge in the HSI field that we aim to address through physics-informed models or adaptive correction strategies. Second, the implementation relies on a fixed physical rank k that may not be optimal for all scenes, suggesting a need for adaptive rank selection based on spectral complexity. Furthermore, although LRDUN is more efficient than existing DUNs, it still requires more resources than some lightweight methods, necessitating further architectural optimization. Finally, evaluating LRDUN across a wider range of real-world applications, such as remote sensing and medical imaging, remains essential to fully understand its generalizability and practical utility in complex, non-ideal sensing environments.

References

[1] Han Cai, Junyan Li, Muyan Hu, Chuang Gan, and Song Han. Efficientvit: Lightweight multi-scale attention for high-resolution dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17302–17313, 2023. 1

[2] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. MST++: Multi-stage Spectral-wise Transformer for Efficient Spectral Reconstruction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 744–754, New Orleans, LA, USA, 2022. IEEE. 1

[3] Yuanhao Cai, Yuxin Zheng, Jing Lin, Xin Yuan, Yulun Zhang, and Haoqian Wang. Binarized spectral compressive imaging. *Advances in Neural Information Processing Systems*, 36: 38335–38346, 2023. 1

[4] Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, and Zicheng Liu. Dynamic convolution: Attention over convolution kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11030–11039, 2020. 1

[5] Meng-Hao Guo, Cheng-Ze Lu, Qibin Hou, Zhengning Liu, Ming-Ming Cheng, and Shi-Min Hu. Segnext: Rethinking convolutional attention design for semantic segmentation. *Advances in neural information processing systems*, 35:1140–1156, 2022. 1

[6] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *Computational visual media*, 9(4):733–752, 2023. 1

[7] Dongheon Lee, Seokju Yun, and Youngmin Ro. Emulating Self-attention with Convolution for Efficient Image Super-Resolution, 2025. 1

[8] Hiroyuki Sato and Kensuke Aihara. Cholesky qr-based retraction on the generalized stiefel manifold. *Computational Optimization and Applications*, 72(2):293–308, 2019. 1

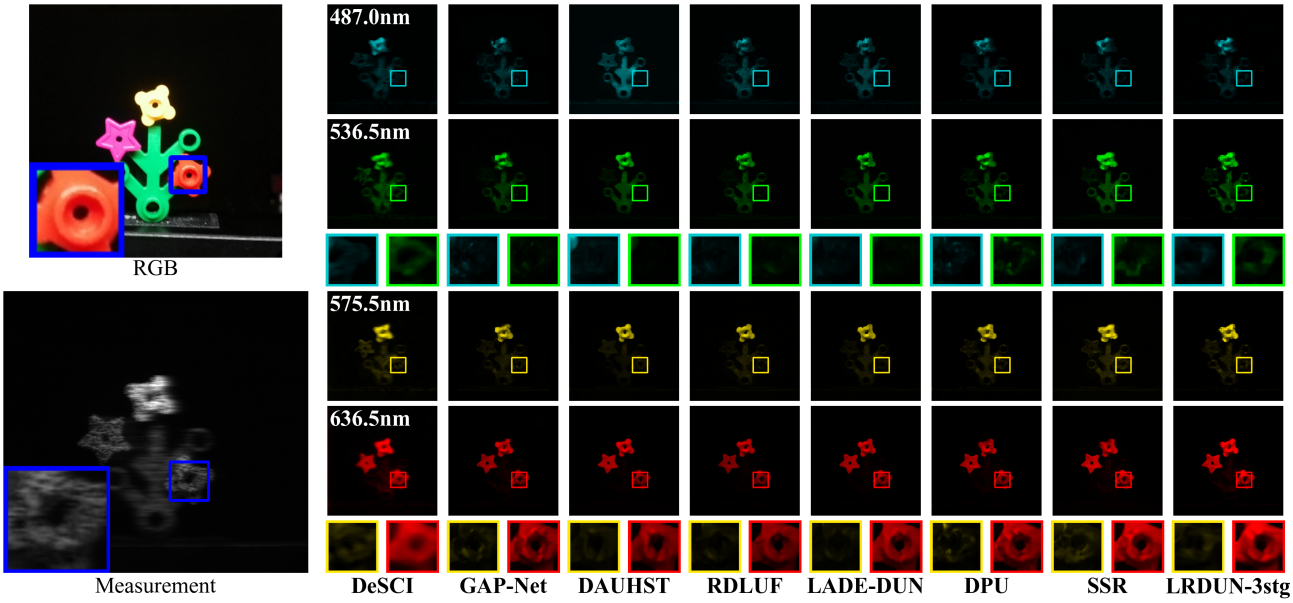

Figure S1. Reconstructed results of real-world Scene 1, displaying 4 out of 28 spectral channels.

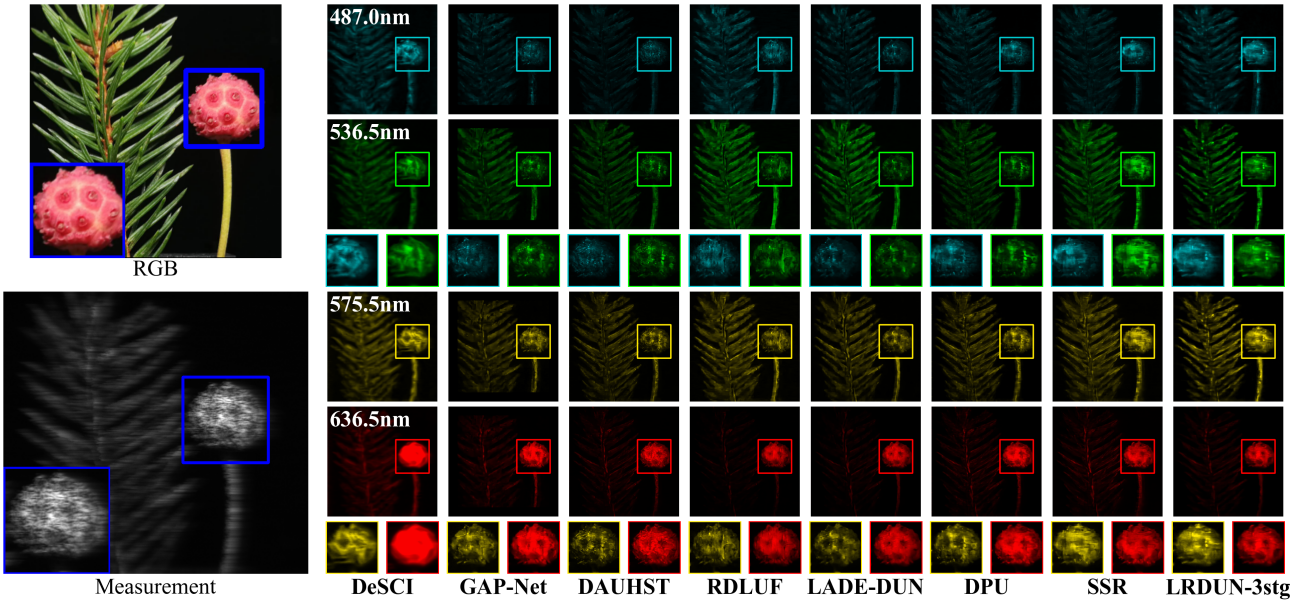

Figure S2. Reconstructed results of real-world Scene 2, displaying 4 out of 28 spectral channels.

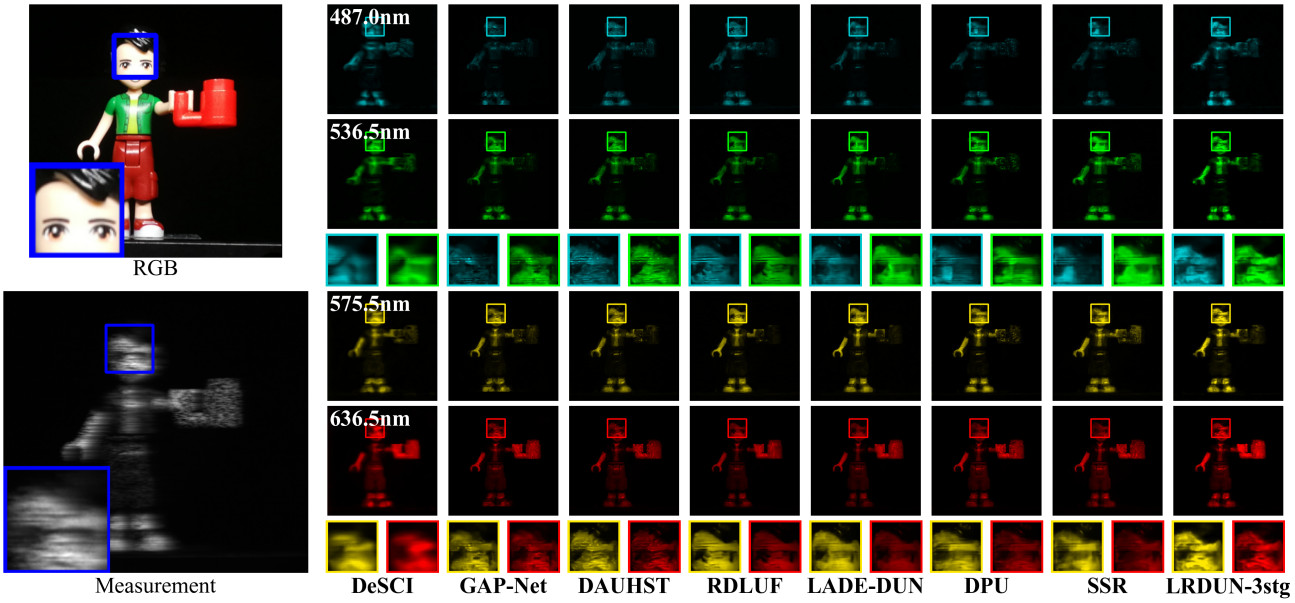

Figure S3. Reconstructed results of real-world Scene 3, displaying 4 out of 28 spectral channels.

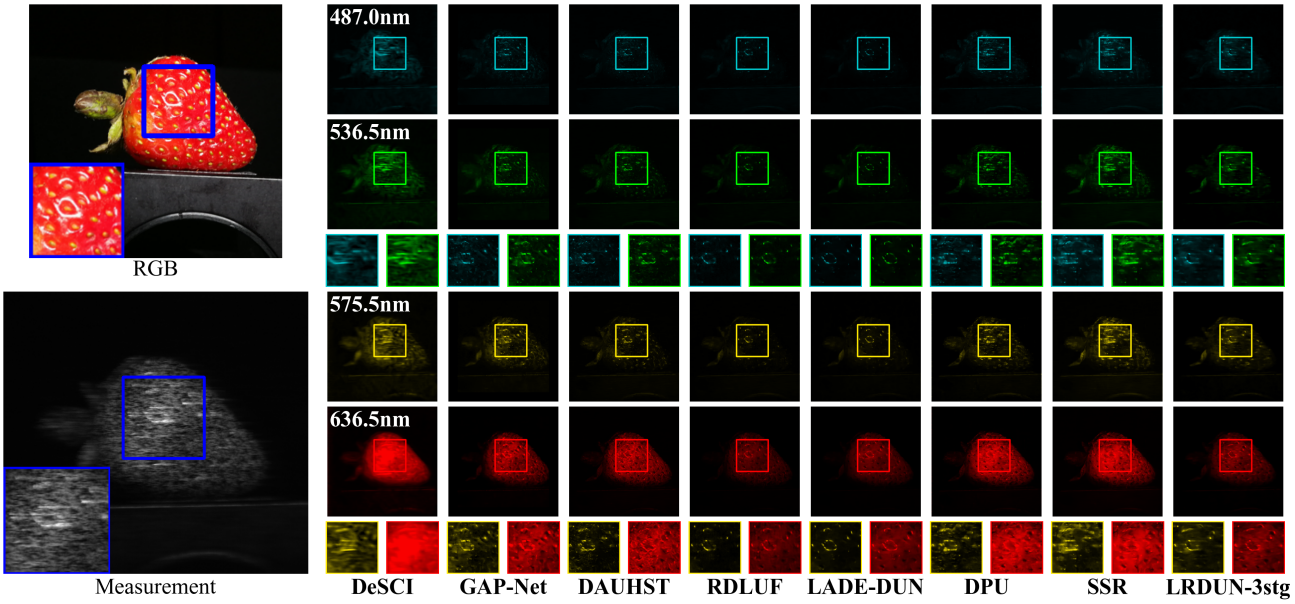

Figure S4. Reconstructed results of real-world Scene 5, displaying 4 out of 28 spectral channels.

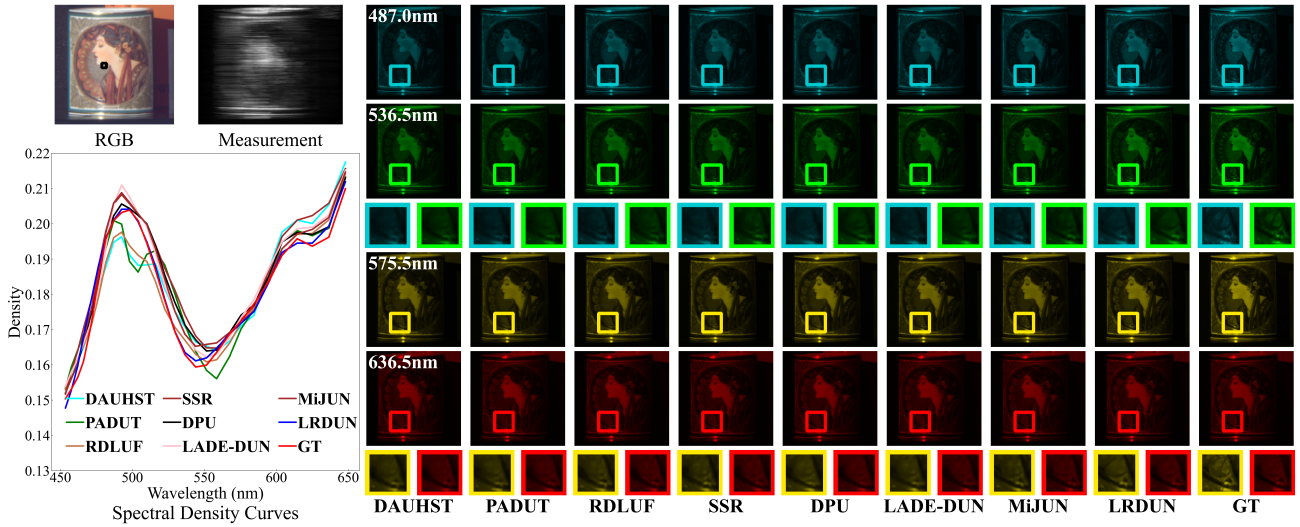

Figure S5. Reconstructed results of the simulated Scene 1, showing 4 out of 28 spectral channels obtained by state-of-the-art methods. One representative region is selected for spectral analysis. The proposed LRDUN achieves the best spatial details (see the zoomed figures of 575.5nm and 636.5nm) and spectral fidelity.

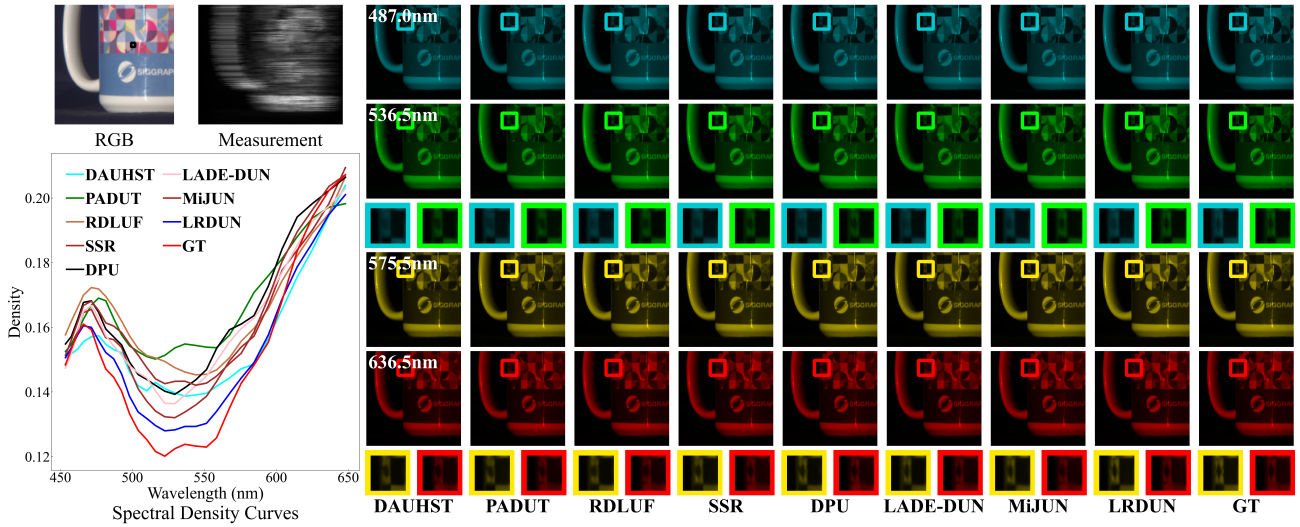

Figure S6. Reconstructed results of the simulated Scene 5, showing 4 out of 28 spectral channels obtained by state-of-the-art methods. One representative region is selected for spectral analysis. The proposed LRDUN achieves the best spectral fidelity.

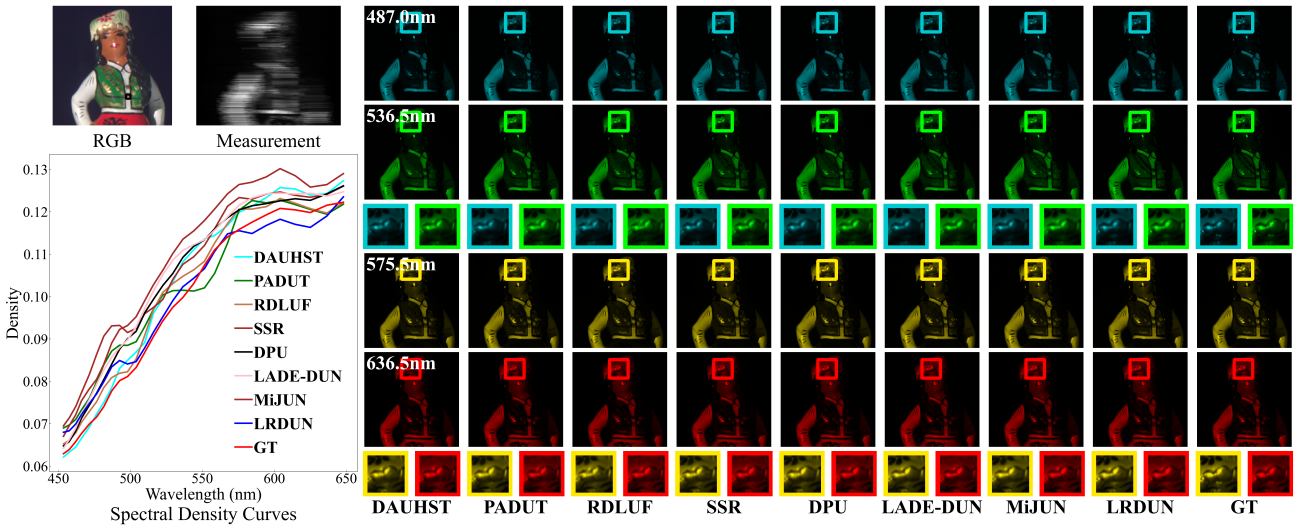

Figure S7. Reconstructed results of the simulated Scene 6, showing 4 out of 28 spectral channels obtained by state-of-the-art methods. One representative region is selected for spectral analysis. The proposed LRDUN achieves the best spectral fidelity.

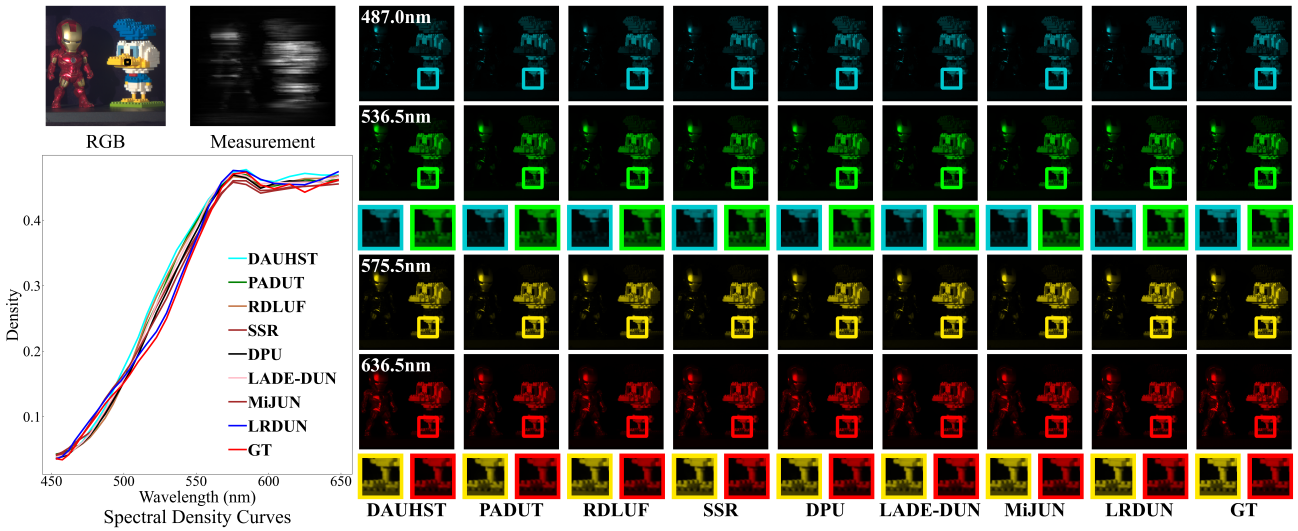

Figure S8. Reconstructed results of the simulated Scene 8, showing 4 out of 28 spectral channels obtained by state-of-the-art methods. One representative region is selected for spectral analysis. The proposed LRDUN achieves the best spatial details (see the zoomed figures of 487.0nm).

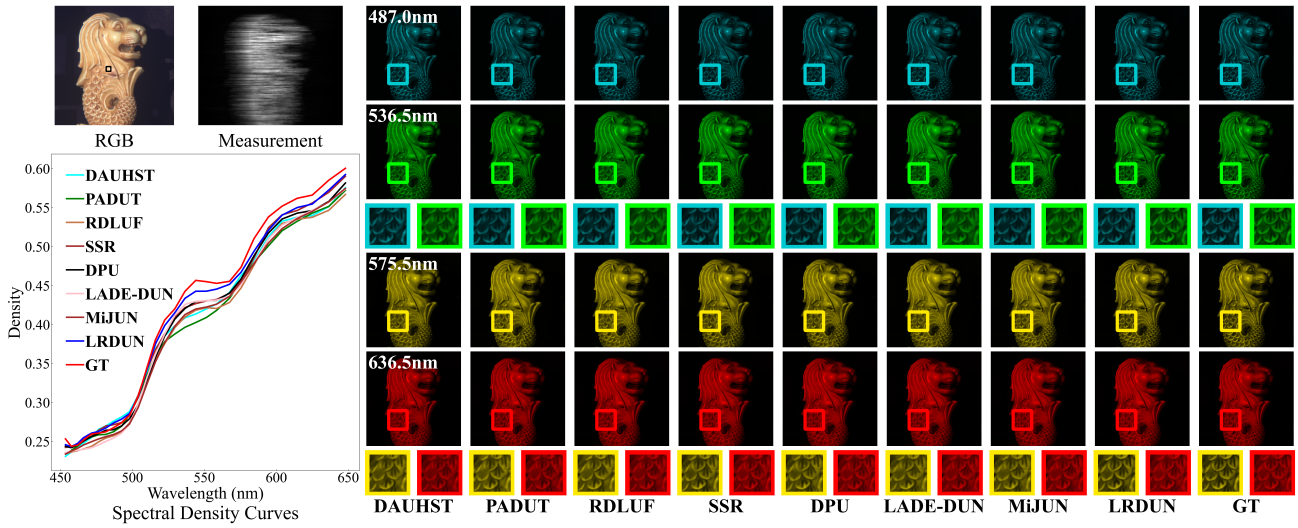

Figure S9. Reconstructed results of the simulated Scene 10, showing 4 out of 28 spectral channels obtained by state-of-the-art methods. One representative region is selected for spectral analysis. The proposed LRDUN achieves the best spectral fidelity.