DiffSCI: Zero-Shot Snapshot Compressive Imaging via Iterative Spectral Diffusion Model

Supplementary Material

Summary

In the supplementary material, we extensively present the closed-form solution derivation for the data subproblem and discuss the theoretical foundation of diffusion as a generative denoiser prior. Moreover, we address the enhancements in convergence speed and reconstruction quality resulting from the implementation of the acceleration algorithm. Simultaneously, this supplementary material provides an extensive visual comparison among three methods for diffusion adaptation in MSI. Additionally, we offer a comparative analysis of our DiffSCI method against the current state-of-the-art (SOTA) method using both simulation and real datasets.

Appendix Section

1. HQS-based Sampling

The inversion problem of SCI can be modeled as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi x}\|^2 + \lambda \mathcal{P}(\mathbf{x}), \tag{1}$$

where $\lambda\mathcal{P}(\mathbf{x})$ denotes the prior term with regularization parameter λ. By adopting HQS algorithm [3] and introducing an auxiliary variable $\mathbf{z}$ to decouple the inversion problem, Eq. (1) can be solved by minimizing the following problem:

$$\mathcal{L}_\mu(\mathbf{x}, \mathbf{z}) = \frac{1}{2}||\mathbf{y} - \mathbf{\Phi x}||^2 + \lambda\mathcal{P}(\mathbf{z}) + \frac{\mu}{2}||\mathbf{z} - \mathbf{x}||^2, \tag{2}$$

where μ is a penalty parameter. Thus, the problem can be solved by iteratively solving following two subproblems:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|^2 + \mu\|\mathbf{x} - \mathbf{z}_k\|^2, \tag{3}$$

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} \frac{\mu}{2}\|\mathbf{z} - \mathbf{x}_{k+1}\|^2 + \lambda\mathcal{P}(\mathbf{z}). \tag{4}$$

1.1. Closed-form Solution to Data Subproblem

The closed-form solution of Eq. (3) can be easily found as [1, 4]:

$$\mathbf{x_{k+1}} = (\mathbf{\Phi}^T\mathbf{\Phi} + \mu\mathbf{I})^{-1}(\mathbf{\Phi}^T\mathbf{y} + \mu\mathbf{z_k}), \tag{5}$$

where $\mathbf{I}$ is an identity matrix. Due to the diagonal structure of $\mathbf{\Phi}\mathbf{\Phi}^T$ in image-plane coding, the calculation of $(\mathbf{\Phi}^T\mathbf{\Phi} + \mu\mathbf{I})^{-1}$ can be done efficiently using the matrix inversion lemma (Woodbury matrix identity) [2], i.e.,

$$(\mathbf{\Phi}^T\mathbf{\Phi} + \mu\mathbf{I})^{-1} = \mu^{-1}\mathbf{I} - \mu^{-1}\mathbf{\Phi}^T(\mathbf{I} + \mu\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mu^{-1}. \tag{6}$$

In SCI system, specifically in the coded aperture snapshot spectral compressive imaging (CASSI) systems, the $\mathbf{\Phi}\mathbf{\Phi}^T$ is a diagonal matrix, so that Eq. (5) can be simplified and the final solution is [1, 4]:

$$\mathbf{x}_{k+1} = \mathbf{z}_k + \mathbf{\Phi}^T[\mathbf{y} - \mathbf{\Phi z}_k] \oslash [Diag(\mathbf{\Phi}\mathbf{\Phi}^T) + \mu]. \tag{7}$$

where $Diag(\cdot)$ extracts the diagonal elements of the ensured matrix, $\oslash$ denotes the element-wise division of Hadamard division. Let $Diag(\mathbf{\Phi}\mathbf{\Phi}^T) \overset{\text{def}}{=} [\psi_1, \psi_2, ..., \psi_n]$, $\mathbf{y} = [y_1, ...y_n]^T$, and $[\mathbf{\Phi z}_k]_i$ denote the i_{th} element of vector $\mathbf{\Phi z}_k$ [1]. Then Eq. (7) can be rewritten as:

$$\mathbf{x}_{k+1} = \mathbf{z}_k + \mathbf{\Phi}^T[\frac{y_1 - [\mathbf{\Phi z}_k]_1}{\mu + \psi_1}, ..., \frac{y_n - [\mathbf{\Phi z}_k]_n}{\mu + \psi_n}]^T. \tag{8}$$

In this way, x_{k+1} can be updated efficiently.

1.2. One Step Reverse Diffusion Step

Each reverse step t of conditional diffusion can be treated as solving the MAP prediction problem with β_t as the noise level:

$$\begin{aligned} \hat{\mathbf{x}}_t &= \arg\min_{\mathbf{x}_t} \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi x}_t\|^2 + \lambda\mathcal{P}(\mathbf{z}_t), \\ s.t.\ \ \mathbf{z}_t &= \mathbf{x}_t = \sqrt{1-\beta_t}\mathbf{z}_{t-1} + \sqrt{\beta_t}\epsilon. \end{aligned} \tag{9}$$

With HQS trick, Eq. (9) can be decoupled into two subproblems similar to Eq.(3) and Eq.(4) as:

$$\tilde{\mathbf{x}}_t = \arg\min_{\mathbf{x}_t} \|\mathbf{y} - \mathbf{\Phi x}_t\|^2 + \mu\|\mathbf{x}_t - \tilde{\mathbf{z}}_t\|^2, \tag{10}$$

$$\tilde{\mathbf{z}}_t = \arg\min_{\mathbf{z}_t} \frac{1}{2(\sqrt{\lambda/\mu})^2}\|\mathbf{z}_t - \tilde{\mathbf{x}}_t\|^2 + \mathcal{P}(\mathbf{z}_t), \tag{11}$$

where $\sigma_t = \sqrt{\lambda/\mu}$, and it denotes as relative noise level between $\mathbf{x}_t$ and $\mathbf{z}_t$ with $\sigma_t = \sqrt{\frac{\beta_t}{1-\beta_t}}$.

We assume that $\tilde{\mathbf{x}}_t^{(b)}$ is the three-channel image obtained by WM(B_b) for spectral band b of the MSI at timestep t. To establish connection between Eq. (11) and one step reverse diffusion, Eq. (11) can be rewritten as:

$$\begin{aligned} \tilde{\mathbf{z}}_{t-1}^{(b)} = \arg\min_{\mathbf{z}_{t-1}^{(b)}} &\frac{1}{2(\frac{\beta_t}{1-\beta_t})}\|\sqrt{1-\beta_t}\mathbf{z}_{t-1}^{(b)} + \sqrt{\beta_t}\epsilon - \tilde{\mathbf{x}}_t^{(b)}\|^2 \\ &+ \mathcal{P}(\sqrt{1-\beta_t}\mathbf{z}_{t-1}^{(b)} + \sqrt{\beta_t}\epsilon). \end{aligned} \tag{12}$$

Note that $\nabla_{\mathbf{x}}\mathcal{P}(\mathbf{x}) = -\nabla_{\mathbf{x}} \log p(\mathbf{x}) = -\mathbf{s}_\theta(\mathbf{x})$, we have

$$\sqrt{1-\beta_t}\tilde{\mathbf{z}}_{t-1}^{(b)} + \sqrt{\beta_t}\epsilon \approx \tilde{\mathbf{x}}_t^{(b)} + \frac{\beta_t}{1-\beta_t}\mathbf{s}_\theta(\tilde{\mathbf{x}}_t^{(b)}, t), \quad (13)$$

which can be obtained by minimising the first order approximation on the Right-hand side of Eq. (12). Meanwhile, it can be treated as a gradient step with step length $\frac{\beta_t}{1-\beta_t}$. So that we can solve $\tilde{\mathbf{z}}_{t-1}^{(b)}$ as:

$$\begin{aligned}
\tilde{\mathbf{z}}_{t-1}^{(b)} &= \frac{1}{\sqrt{1-\beta_t}}(\tilde{\mathbf{x}}_t^{(b)} + \frac{\beta_t}{1-\beta_t}\mathbf{s}_\theta(\tilde{\mathbf{x}}_t^{(b)}, t)) + \frac{\sqrt{\beta_t}}{\sqrt{1-\beta_t}}\epsilon_0 \\
&= \frac{1}{\sqrt{1-\beta_t}}(\tilde{\mathbf{x}}_t^{(b)} + [\beta_t + o(\beta_t)]\mathbf{s}_\theta(\tilde{\mathbf{x}}_t^{(b)}, t)) + [(\sqrt{\beta_t} + o(\sqrt{\beta_t})]\epsilon_0 \\
&\approx \frac{1}{\sqrt{1-\beta_t}}[(\tilde{\mathbf{x}}_t^{(b)} + \beta_t\mathbf{s}_\theta(\tilde{\mathbf{x}}_t^{(b)}, t)] + \sqrt{\beta_t}\epsilon_0,
\end{aligned} \quad (14)$$

where $\epsilon_0 = -\epsilon$ is also obeying $N(0, I)$ and β_t is set as a linear schedule from $(0.1/1000)$ to $(20/1000)$. Eq. (14) is same as one step reverse of DDPM. Then we can get $\tilde{\mathbf{z}}_{t-1}$ through combining each $\tilde{\mathbf{z}}_{t-1}^{(b)}$ which $b \in [1, B]$.

1.3. Acceleration algorithm

The idea of the acceleration algorithm we use is residual accumulation. By introducing an additional momentum term $\mathbf{y}_1$, an inertia is introduced in the parameter update, so that the parameter update is not only affected by the current gradient, but also by the gradient of the previous steps. As Momentum's accelerated update rules:

$$\mathbf{v}_{n+1} = \beta\mathbf{v}_n + \nabla f(\mathbf{x}_n) \quad (15)$$

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \alpha\mathbf{v}_{n+1}, \quad (16)$$

where $\mathbf{v}_n$ is the momentum term i.e. $\mathbf{y}_1$ and α is the learning rate i.e. sc in our acceleration algorithm. The performance improvement resulting from the acceleration algorithm is shown in Table 1.

Table 1. Avg PSNR, SSIM and time of sampling with and without acceleration on simulation dataset

	sampling w/o acc	sampling w/ acc
Avg PSNR (db)	29.92	35.28 ↑ 5.36
Avg SSIM	0.826	0.916 ↑ 0.09
Avg Time (s)	925.3	730.5 ↑ 194.8

2. Additional Visual Results

In this section, we provide more visualisation examples for the simulation and real datasets to demonstrate the capabilities of our approach.

2.1. Results on Simulation Dataset

2.1.1 Comparison among Three Methods

In this section, we provide more visual examples to show that our proposed Wavelength Matching (WM) approach has a significant improvement over the Non-overlapping Bands Method and Spectral Correlation Modelling Method. The Spectral Correlation Modelling Method improves the reconstruction quality compared to selecting three Non-overlapping Bands Method, however, the wavelength mismatch leads to poor reconstruction in the forward bands, which is shown in Fig. 1 and Fig. 2. Our proposed WM method provides better recovery of the forward bands compared to the previous two methods.

2.1.2 Comparison with Other SOTA Methods

We conduct additional comparative experiments with current SOTA methods. As shown in Fig. 3 and Fig. 4, we find that our DiffSCI allows for better reconstruction in low-light and sharp-edged scenes, e.g., in $Scene$ 7, we can observe that the clarity of the boundary lines and the fidelity of the colours are better than in other methods. Meanwhile, in $Scene$ 8, our DiffSCI method even reconstructs the gaps between the blocks that other methods can't.

However, our DiffSCI method still has some limitations. As shown in Fig. 4, the reconstruction of $Scene$ 10 is unsatisfactory. The reason behind this is due to the fact that diffusion is an unconditional generative model, whereas our approach turns it into a conditional generative model by means of using data proximity. Nevertheless this conditional constraint is not enough to accomplish a better reconstruction of structurally complex regions.

2.2. Results on Real Dataset

2.2.1 Comparison with Other SOTA Methods

In this section, we provide more examples of visualisations of real datasets to illustrate the superior performance of our DiffSCI approach. As shown in from Fig. 5 to Fig. 8, our DiffSCI method is able to reconstruct the structural outline better compared to current methods. However, artifacts still appear in the reconstruction results due to the interference of noise, which will be a concern for subsequent research.

References

[1] Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc V Gool. Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. In *NeurIPS*, 2022. 1

[2] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004. 1

[3] Donald Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE TIP*, 4(7):932–946, 1995. 1

[4] Siming Zheng, Yang Liu, Ziyi Meng, Mu Qiao, Zhishen Tong, Xiaoyu Yang, Shensheng Han, and Xin Yuan. Deep plug-and-

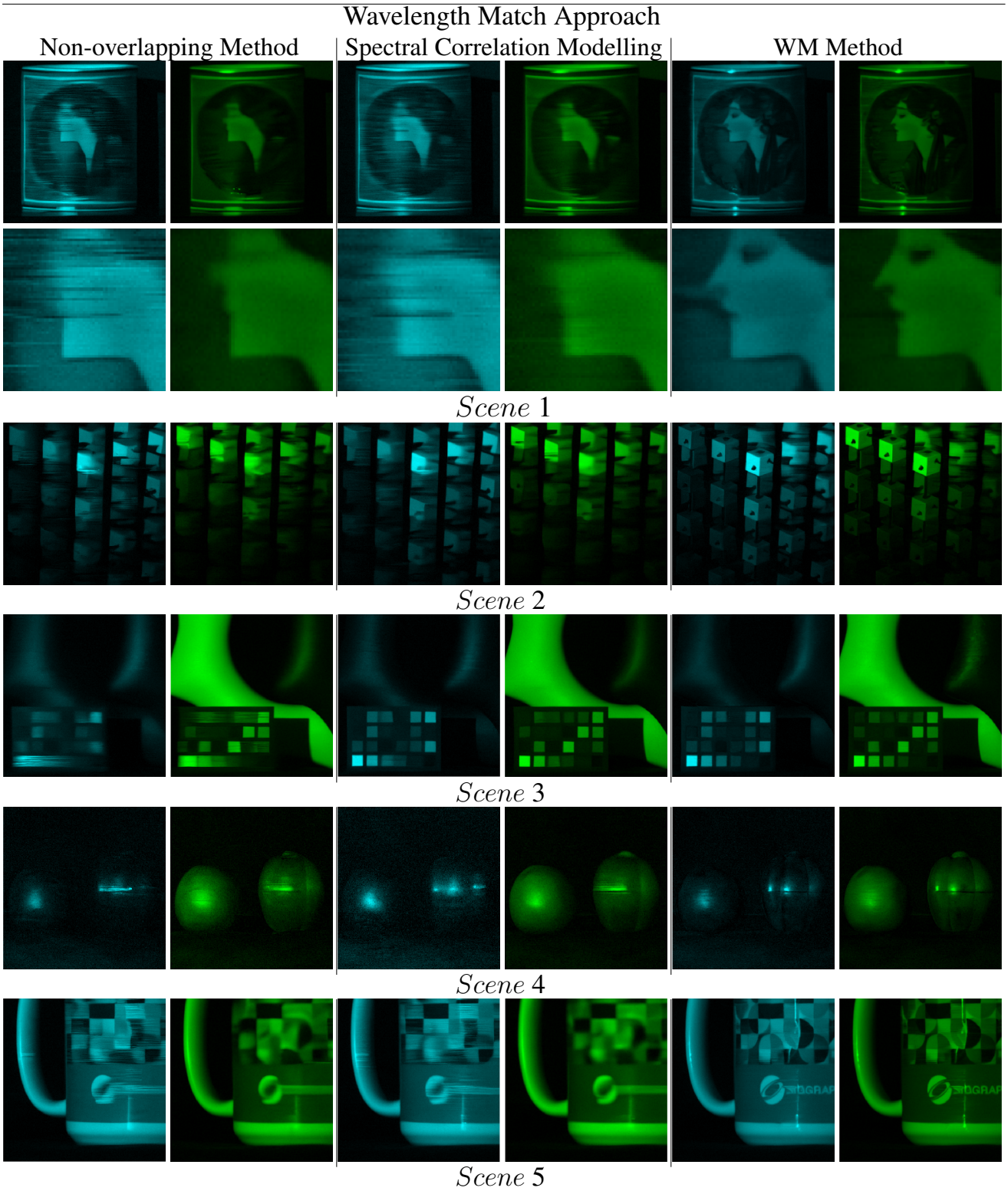

Figure 1. Visual comparison among three methods at 487nm and 536.5nm wavelengths on *Scene* 1 to *Scene* 5 in KAIST dataset.**Two columns on the left:** selecting Non-overlapping Bands Method, **Middle two columns:** Spectral Correlation Modelling Method, **Two columns on the right:** WM Method.

play priors for spectral snapshot compressive imaging. *PR*, 9(2):B18–B29, 2021. 1

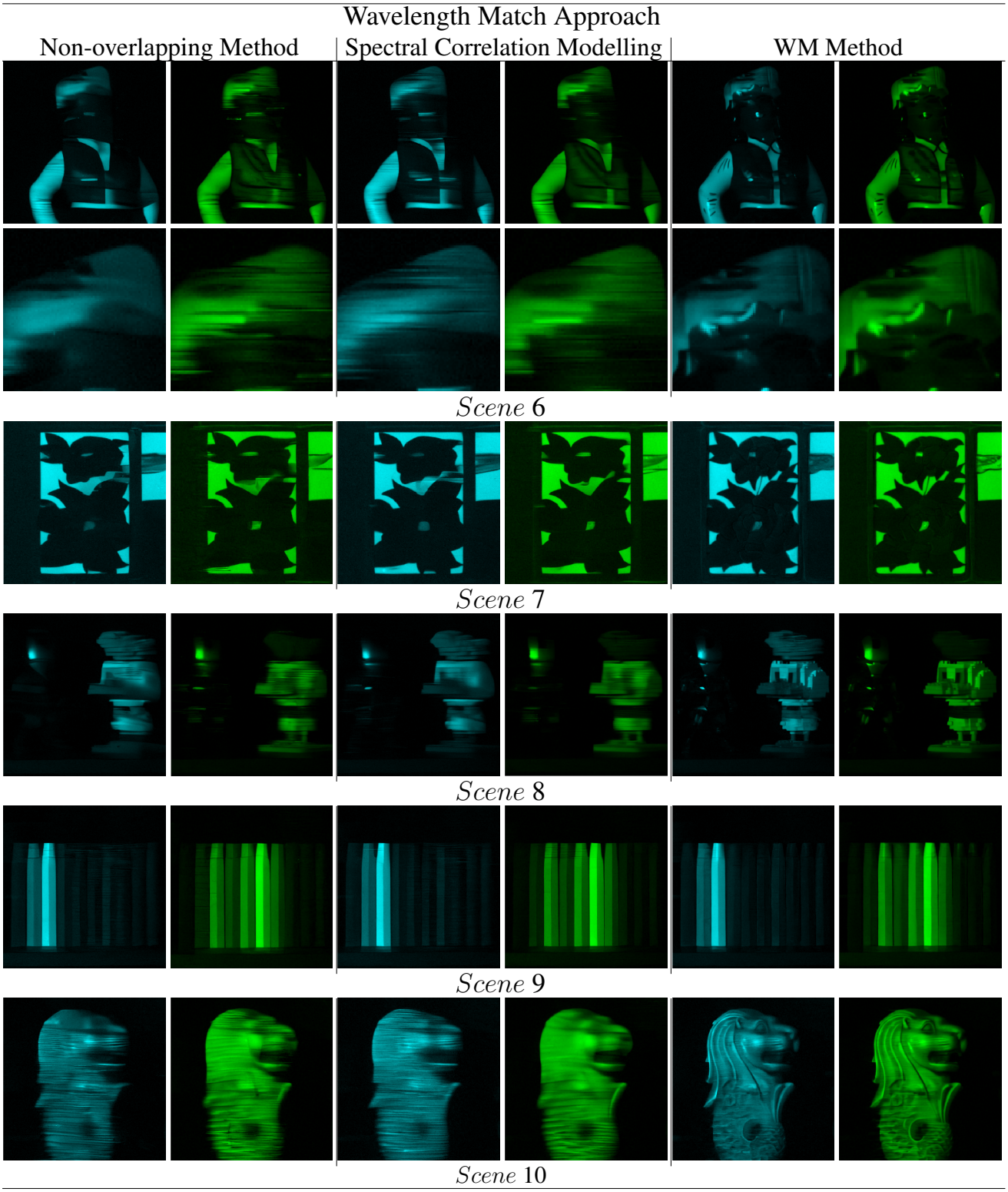

Figure 2. Visual comparison among three methods at 487nm and 536.5nm wavelengths on *Scene* 5 to *Scene* 10 in KAIST dataset.**Two columns on the left:** selecting Non-overlapping Bands Method, **Middle two columns:** Spectral Correlation Modelling Method, **Two columns on the right:** WM Method.

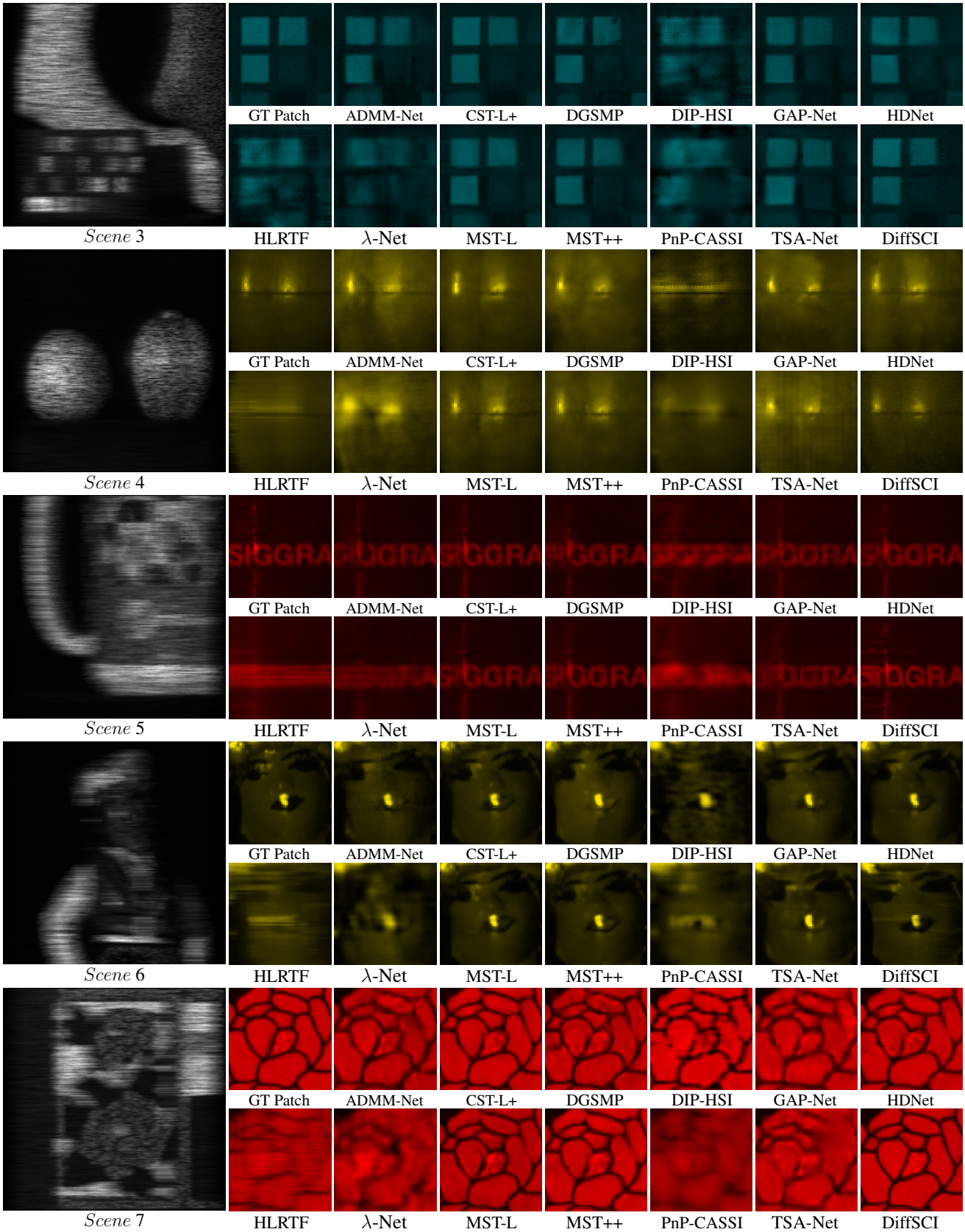

Figure 3. Visual comparison from *Scene* 3 to *Scene* 7 of KAIST dataset.

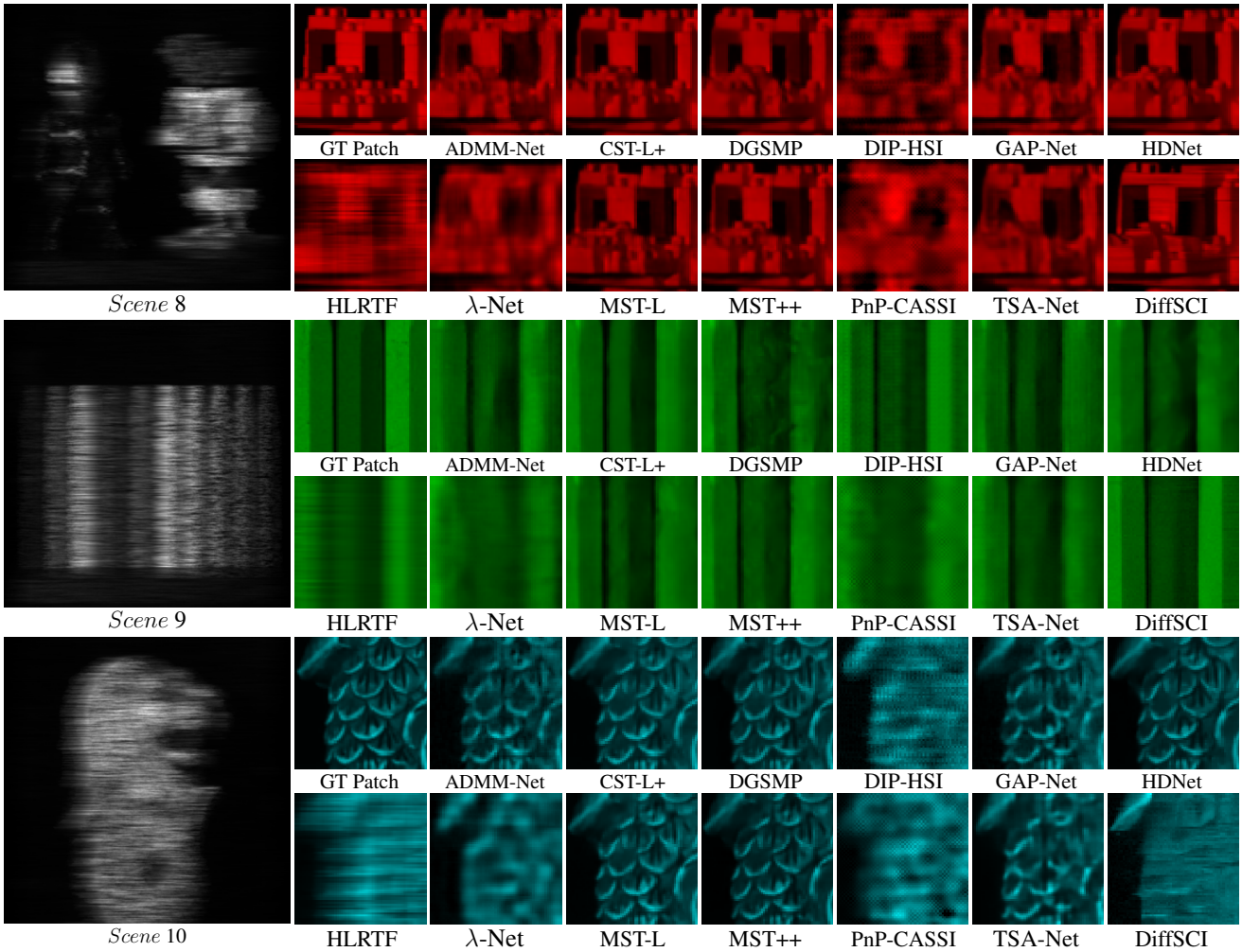

Figure 4. Visual comparison from *Scene* 8 to *Scene* 10 of KAIST dataset.

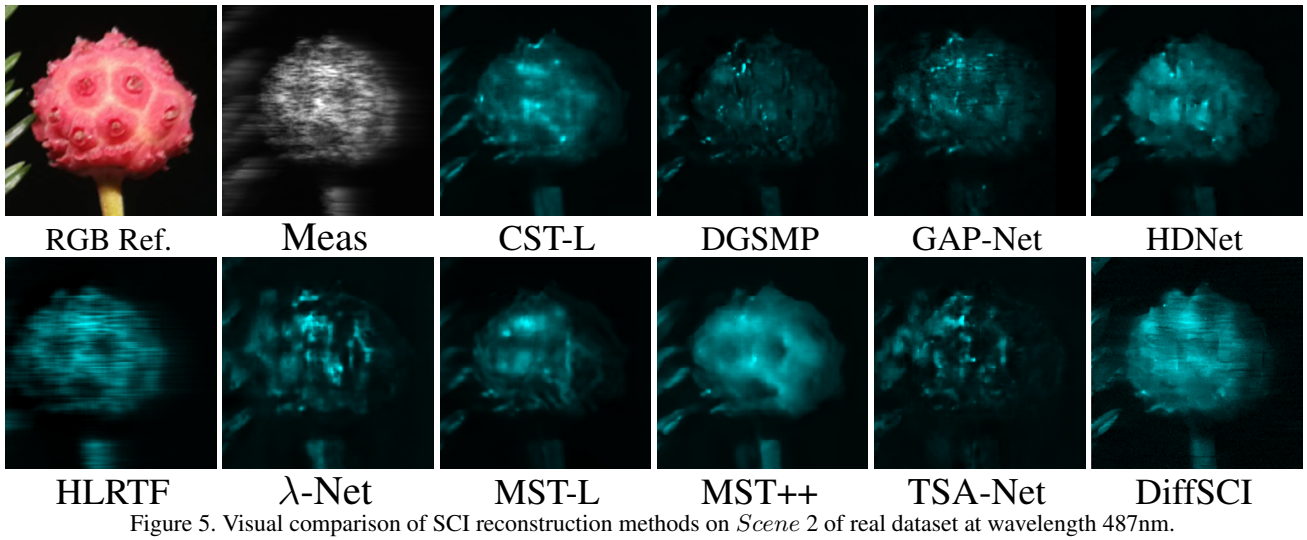

Figure 5. Visual comparison of SCI reconstruction methods on *Scene* 2 of real dataset at wavelength 487nm.

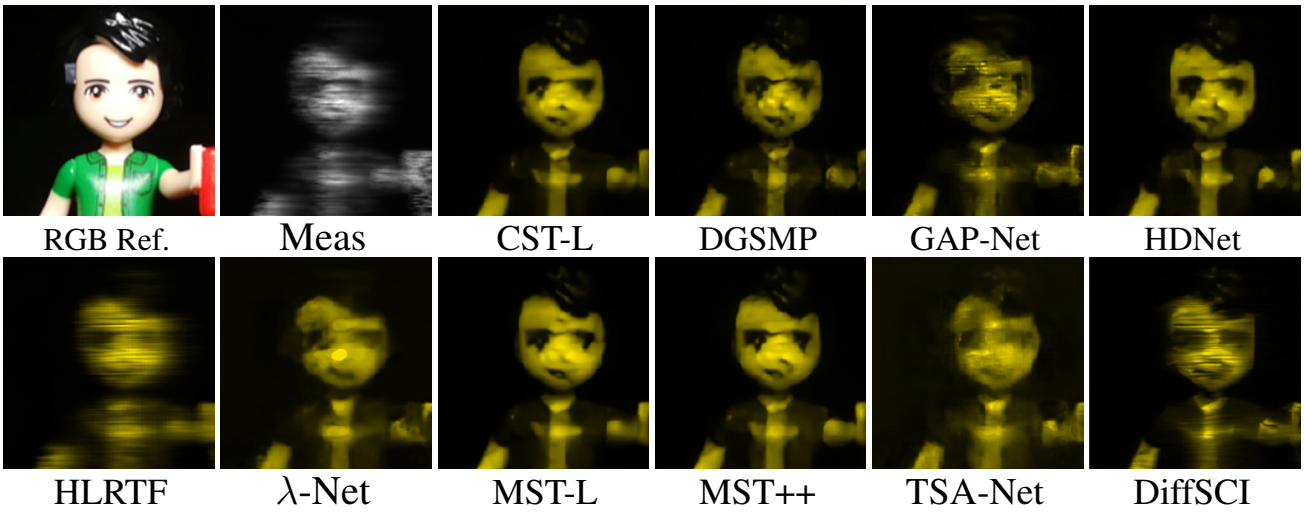

Figure 6. Visual comparison of SCI reconstruction methods on *Scene* 3 of real dataset at wavelength 575.5nm.

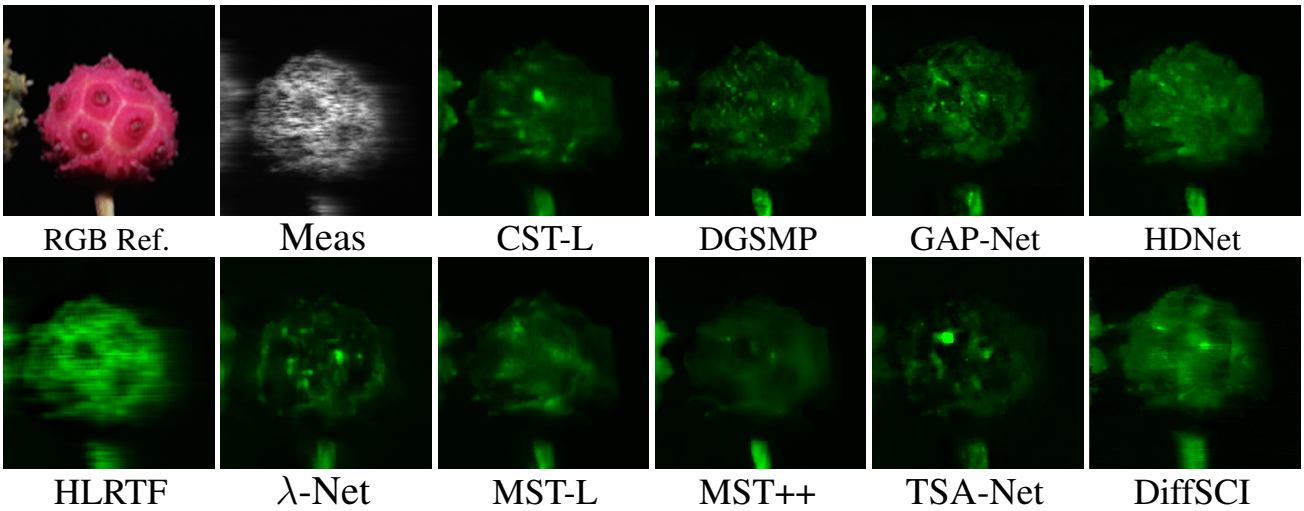

Figure 7. Visual comparison of SCI reconstruction methods on *Scene* 4 of real dataset at wavelength 536.5nm.

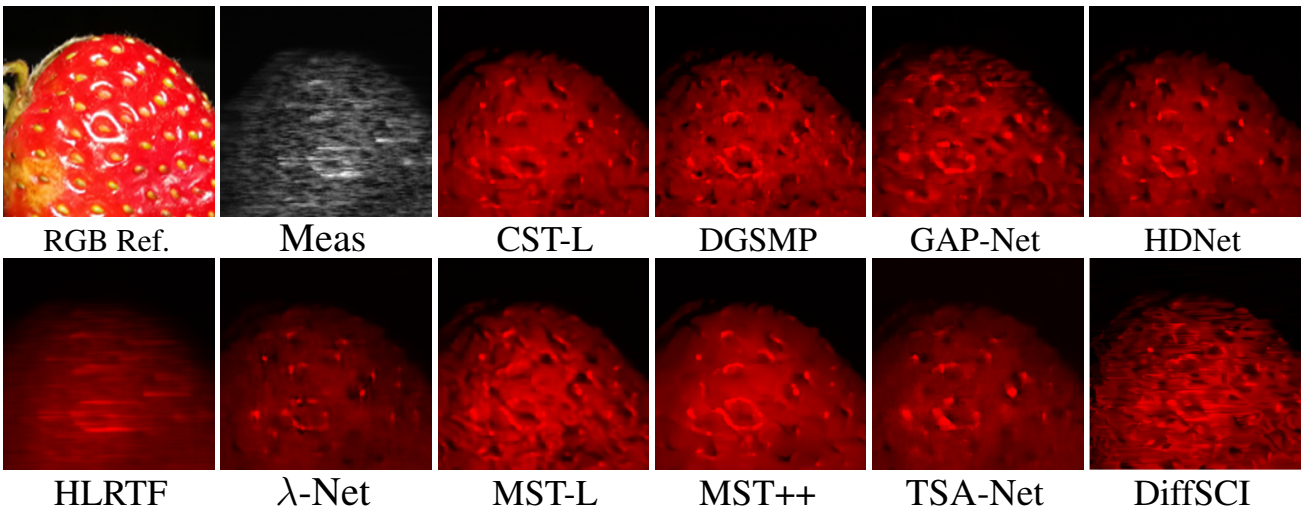

Figure 8. Visual comparison of SCI reconstruction methods on *Scene* 5 of real dataset at wavelength 648nm.