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

Zhenghao Pan [1,*] Haijin Zeng [2,*] Jiezhang Cao [3] Kai Zhang [4] Yongyong Chen [1,†]
[1] Harbin Institute of Technology (Shenzhen) [2] IMEC-UGent [3] ETH Zurich [4] Nanjing University

Abstract

This paper endeavors to advance the precision of snapshot compressive imaging (SCI) reconstruction for multispectral image (MSI). To achieve this, we integrate the advantageous attributes of established SCI techniques and an image generative model, propose a novel structured zero-shot diffusion model, dubbed DiffSCI. DiffSCI leverages the structural insights from the deep prior and optimization-based methodologies, complemented by the generative capabilities offered by the contemporary denoising diffusion model. Specifically, firstly, we employ a pre-trained diffusion model, which has been trained on a substantial corpus of RGB images, as the generative denoiser within the Plug-and-Play framework for the first time. This integration allows for the successful completion of SCI reconstruction, especially in the case that current methods struggle to address effectively. Secondly, we systematically account for spectral band correlations and introduce a robust methodology to mitigate wavelength mismatch, thus enabling seamless adaptation of the RGB diffusion model to MSIs. Thirdly, an accelerated algorithm is implemented to expedite the resolution of the data subproblem. This augmentation not only accelerates the convergence rate but also elevates the quality of the reconstruction process. We present extensive testing to show that DiffSCI exhibits discernible performance enhancements over prevailing self-supervised and zero-shot approaches, surpassing even supervised transformer counterparts across both simulated and real datasets. Code is at https://github.com/PAN083/DiffSCI.

1. Introduction

Contrary to conventional RGB images, multispectral images (MSIs) incorporate an expanded array of spectral bands, enabling the retention of more comprehensive and detailed information. Therefore, MSIs are widely applied in remote sensing [4, 21, 34, 57], medical imaging [2, 30, 36], environmental monitoring [47], etc. Owing to the advancement of snapshot compressive imaging (SCI) systems [9, 18, 29, 33, 50, 51, 54], it has become feasible to acquire two-dimensional measurements of MSIs. The decoding stage of the SCI system aims to reconstruct the three-dimensional MSIs from its degraded two-dimensional measurement.

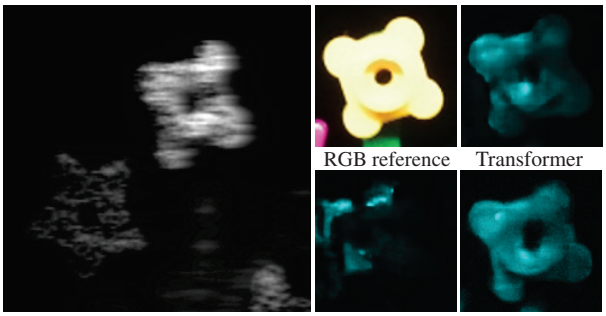


Figure 1. Comparison of Transformer (MST [5]), Deep Unfolding (TSA-Net [35]), and the proposed DiffSCI for real SCI reconstruction. The RGB image from the same scene serves as the reference. DiffSCI can reconstruct some unsampled and compressed scene contents by rethinking SCI through the generative diffusion model.

Given the ill-posed nature of SCI reconstruction as an inverse problem, existing methods still face several key challenges in accurately reconstructing certain aspects. For instance, inadequately illuminated regions or areas with sharp edges remain problematic as shown in Fig. 1. The underlying reason may be that insufficient sampling occurred in the above areas, then the reconstruction algorithm may not be able to accurately recover the detail information. Moreover, contemporary end-to-end (E2E) models [23, 35, 36, 39], while processing both two-dimensional measurements and three-dimensional MSIs maps, may inadvertently lose crucial high-dimensional information due to necessary dimensionality reduction. And current unsupervised methods also fail to achieve satisfactory results. Furthermore, the performance of the reconstruction on real-world datasets frequently deviates from the ideal, primarily attributable to discrepancies between the training dataset and the novel, unseen testing images, as evidenced in Fig. 1.

Diffusion model [12, 16, 27, 42] has demonstrated notable proficiency in generating content from RGB images [59]. Leveraging its generative capacity to address

*Equal contribution. † Corresponding Author: Yongyong Chen (YongyongChen.cn@gmail.com)

challenging-to-reconstruct segments holds promise for enhancing MSIs SCI results [1, 12, 13, 22, 44]. Nonetheless, two significant challenges must be confronted: **(i)** MSIs lack extensive training data compared to RGB images due to their broad band spectrum, amplifying temporal and GPU resource requirements for training. Consequently, directly training a diffusion model on MSIs is a formidable task. **(ii)** Utilizing pre-trained diffusion models, primarily trained on large RGB datasets with three channels, poses challenges for MSI reconstruction due to their extensive band spectrum. The task involves decoding a spatial-spectral MSI from a single measurement, presenting significant differences in input and output dimensions. Consequently, directly applying diffusion models to MSI reconstruction is non-trivial.

Plug-and-Play (PnP) [10, 37, 43, 55, 56, 58] framework incorporates pre-trained denoising networks into traditional model-based methods, due to its interpretability of the principles underlying SCI and its flexibility across different SCI systems, has emerged as one of the most predominant reconstruction techniques in the current scenario. Also because the noise type in the iterative steps is additive Gaussian noise with variance σ_n^2 [49], we thought of using PnP framework to apply the pre-trained diffusion model based on massive RGB images as denoiser to the reconstruction of MSIs. However, there are four key challenges to embedding the diffusion model into MSIs at present. **(i)**It is impossible to input MSIs involving dozens of spectral bands directly into existing diffusion models pre-trained on RGB images. **(ii)** There exists a spectral connection among the bands of MSIs, and many existing denoisers trained on RGB do not have a good grasp of this connection. **(iii)** The wavelength range of RGB images is much smaller than that of MSIs, making wavelength mismatch issues inevitable. This discrepancy could significantly impact the performance of the diffusion model. **(iv)** The sampling time required by the diffusion model in RGB images is already substantial. For our MSIs problem, the time required will be even greater. In order to address these challenges, this paper makes the following contributions:

- Initially, the proposed DiffSCI leverages a diffusion model trained on a substantial corpus of RGB images for multispectral SCI reconstruction through the PnP framework, harnessing its generative potential to enhance SCI restoration outcomes. This is the first attempt to fill the research gap to fuse the diffusion model into the PnP framework for multispectral SCI.
- Acknowledging the inherent spectral band correlations in MSIs that are not present in RGB images, we embark on a comprehensive modeling of spectral correlation.
- We introduce a method to address the inevitable issue of wavelength mismatch, given the broader spectral range of MSIs compared to RGB images.
- We implement an accelerated strategy to get the analytic solution of the data subproblem within DiffSCI, which improves the convergence rate and reconstruction quality.

We validate DiffSCI through experiments on simulated and real datasets. Comparative assessments with state-of-the-art methods confirm DiffSCI's superior efficiency in restoring MSIs, as demonstrated by visual examples in Fig. 1.

2. Background

2.1. Degradation Model of CASSI

In Coded Aperture Snapshot Spectral Compressive Imaging (CASSI) systems [19, 35, 50], two-dimensional measurements $\mathcal{Y} \in \mathbb{R}^{H\times(W+d\times(B-1))}$ can be modulated from three-dimensional MSI $\mathcal{X} \in \mathbb{R}^{H\times W\times B}$ as shown in Fig. 2, where H, W, d and B denote the MSI's height, width, shifting step and total number of wavelengths. As [8, 32], we denote the vectorized measurement $\mathbf{y} \in \mathbb{R}^n$ with $n = H(W + d(B-1))$. Then, given vectorized shifted MSI $\mathbf{x} \in \mathbb{R}^{nB}$ and mask $\mathbf{\Phi} \in \mathbb{R}^{n\times nB}$, the degradation model can be formulated as:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathbb{R}^n$ represents the noise on measurement. SCI reconstruction is to obtain $\mathbf{x}$ from the captured $\mathbf{y}$ and the pre-set $\mathbf{\Phi}$ using a reconstruction algorithm [17, 26, 48].

2.2. Denoising Diffusion Probabilistic Models

Diffusion model includes two processes: forward process and reverse process. The forward process is to continuously add Gaussian noise to the clean image ($\mathbf{x}_0$) and eventually turn the initial image into pure Gaussian noise. Thus sampling $\mathbf{x}_t$ at any given timestep t can be formulated as [22]:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{2}$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{k=1}^{t} \alpha_k$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ and β_t is a gradually increasing arithmetic sequence. The reverse process is to gradually restore a clean image from Gaussian noise. One reverse step of Denoising Diffusion Probabilistic Models (DDPM) is [22]:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t)) + \sqrt{\beta_t}\epsilon_t, \tag{3}$$

where $\epsilon_\theta(\mathbf{x}_t, t)$ is the noise predicted by the network at t_{th} step and ϵ_t is standard Gaussian noise. Briefly, DDPM can be interpreted as a process of gradually subtracting the predicted noise from $\mathbf{x}_t$ to restore a clean image $\mathbf{x}_0$.

2.3. Score-based Diffusion Model

Compared to DDPM, the score-based model can use methods like Langevin dynamics for more efficient sampling [46], and at the same time learn the data distribution (i.e., score function) under various noise levels, thus acquiring more training signals. This could help to improve the performance

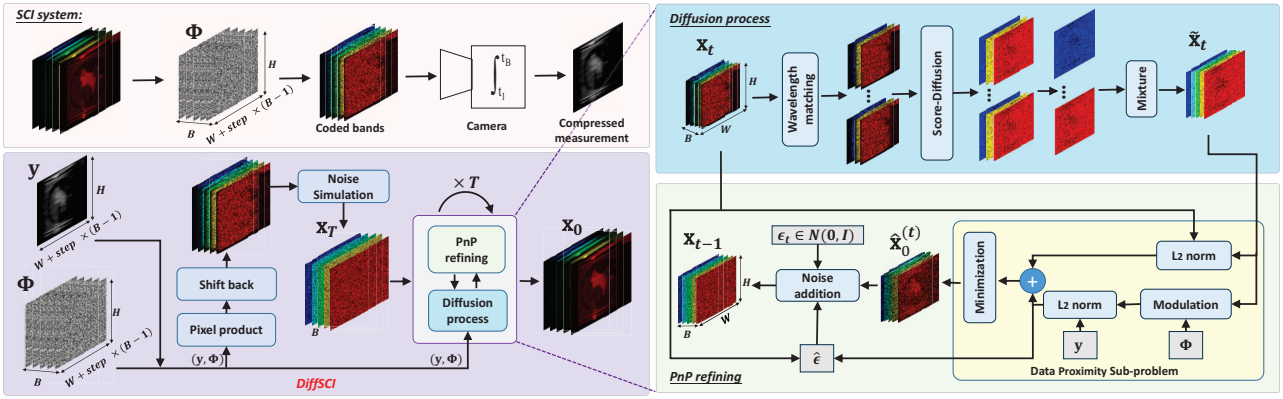


Figure 2. **Top Left:** Obtaining 2D measurements $\mathbf{y}$ of 3D MSI through the SCI system with mask $\mathbf{\Phi}$. **Bottom Left:** DiffSCI generates desired reconstructed MSI ($\mathbf{x}_0$) with $\mathbf{y}$ and $\mathbf{\Phi}$ through reverse diffusion and PnP framework. **Right:** Integrating diffusion model with PnP method with wavelength matching (WM) method as a module of our DiffSCI method.

of the model. The forward process can also be described in the form of a Stochastic Differential Equation (SDE):

$$\mathbf{dx} = f(\mathbf{x}, t)\mathbf{d}t + g(t)\mathbf{d}w, \tag{4}$$

where $\mathbf{d}w$ is infinitesimal white noise, $f(\cdot, t)$ is a vector function called the drift coefficient, and $g(\cdot, t)$ is a real-valued function called the diffusion coefficient. The reverse process can be written as:

$$\mathbf{dx} = [f(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})]\mathbf{d}t + g(t)\mathbf{d}w, \tag{5}$$

where $p_t(\mathbf{x})$ is terminal distribution density [1], and the only unknown part $\nabla_{\mathbf{x}} log p_t(\mathbf{x})$ can be predicted through a score-based model $s_\theta(\mathbf{x}, t)$ [25, 45].

2.4. Denoising Diffusion Implicit Models

In order to accelerate the reverse diffusion process, Denoising Diffusion Implicit Models (DDIM) generates new samples with a non-Markovian process. At each step, the model computes a denoised version of the image and then mixes this denoised version with some noise to generate the image for the next step. This process allows for more efficient estimation and sampling of multiple future states within the same time step, thus improving sampling efficiency and saving time. Therefore, Eq. (3) can be rewritten as:

$$\begin{aligned}\mathbf{x}_{t-1} =& \sqrt{\bar{\alpha}_{t-1}} \left(\frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{\bar{\alpha}_t}} \right) \\ &+ \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_{\eta_t}^2} \cdot \epsilon_\theta(\mathbf{x}_t, t) + \sigma_{\eta_t}\epsilon_t,\end{aligned} \tag{6}$$

the term inside the first bracket can be treated as denoised image $\tilde{\mathbf{x}}_t$ predicted via current $\mathbf{x}_t$, σ_{η_t} controls randomness.

2.5. Conditional Diffusion Model

In the context of conditional generation tasks, we are presented with a condition $\mathbf{y}$, and our objective is to optimize the probability of $p(\mathbf{x}|\mathbf{y})$. Applying Bayes' theorem, we can rewrite Eq. (6) as [46]:

$$\mathbf{dx} = [f(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}}(\log p_t(\mathbf{x}) + \log p_t(\mathbf{y}|\mathbf{x}))]\mathbf{d}t + g(t)\mathbf{d}w, \tag{7}$$

where the unconditionally pre-trained diffusion model achieves conditional generation by adding a classifier. So that, given Eq. (7), one step of reverse sampling under conditional circumstances can be accomplished by first taking one reverse sampling step in the unconditional diffusion model, and then merging it with the conditional constraint.

3. Proposed Method

3.1. Problem Definition and Solution

Diffusion-based methods could theoretically recover the details of dark areas better through their powerful generative ability [41, 52]. Unfortunately, the existing diffusion-based methods are mostly designed for RGB images in which the input and output are with three channels, while the task of SCI reconstruction involves decoding a complete multi-band MSI from a single-band measurement. Meanwhile, limited by the inadequate datasets of MSI and high dimension of the data, resource consumption required for retraining diffusion model on MSIs is high. To leverage the generative power of diffusion models and thus compensate for the shortcomings of current methods, our idea is to insert the pre-trained diffusion model on RGB images as a denoiser into the PnP framework to accomplish SCI reconstruction.

There are now four key problems: **(i)** How can diffusion models, trained on RGB images, be effectively applied to MSIs? **(ii)** How does one capture spectral correlation in MSIs that do not exist in RGB images? **(iii)** What strategies mitigate wavelength mismatching arising from inconsistencies between MSI and RGB wavelengths? **(iv)** How can fast and efficient sampling be achieved for MSIs with numerous bands? To address the above issues, we proposed the DiffSCI method with three modules: Denoising Diffusion PnP-SCI Model, Diffusion Adaptation for MSI, and Acceleration Algorithm. See Fig. 2 for an overall view.

3.2. Denoising Diffusion PnP-SCI Model

The inversion problem of SCI can be modeled as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi x}\|^2 + \lambda \mathcal{P}(\mathbf{x}), \tag{8}$$

where $\mathcal{P}(\mathbf{x})$ denotes diffusion MSI prior, λ is a trade-off parameter. By adopting the half-quadratic splitting (HQS) [20] algorithm and introducing an auxiliary variable $\mathbf{z}$, Eq. (8) can be solved by iteratively solving following two subproblems:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|^2 + \mu\|\mathbf{x} - \mathbf{z}_k\|^2, \tag{9}$$

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} \frac{\mu}{2}\|\mathbf{z} - \mathbf{x}_{k+1}\|^2 + \lambda \mathcal{P}(\mathbf{z}). \tag{10}$$

Closed-form Solution to Data Subproblem. In CASSI system, $\mathbf{\Phi}^T\mathbf{\Phi}$ is a diagonal matrix [8, 58], so that by using matrix inversion theorem (Woodbury matrix identities), the closed-form solution of Eq. (9) can be easily found with fast operation guarantee [14]:

$$\mathbf{x}_{k+1} = \mathbf{z}_k + \mathbf{\Phi}^T[\mathbf{y} - \mathbf{\Phi z}_k] \oslash [Diag(\mathbf{\Phi}\mathbf{\Phi}^T) + \mu], \tag{11}$$

where $Diag(\cdot)$ extracts the diagonal elements of the ensured matrix, $\oslash$ is the element-wise division of Hadamard division.

Diffusion Models as Generative Denoiser Prior. Unlike conventional denoisers, diffusion models possess powerful generative capabilities [15]. To utilize this generative capability, our DiffSCI model explores diffusion as the generative denoiser prior as shown in Fig. 2 to address hard-to-recover parts of SCI reconstruction, such as low-light and sharp edges. We firstly establish the correlation between Eq. (10) and diffusion model. Let $\mathbf{x}_k^{(b)}$ be a three-channel image corresponding to b_{th} band of MSI $\mathbf{x}_k$, from Eq. (10) we have:

$$\mathbf{z}_{k+1}^{(b)} = \arg\min_{\mathbf{z}^{(b)}} \frac{1}{2(\sqrt{\lambda/\mu})^2}\|\mathbf{z}^{(b)} - \mathbf{x}_{k+1}^{(b)}\|^2 + \mathcal{P}(\mathbf{z}^{(b)}), \tag{12}$$

where $\mathbf{z}_{k+1}^{(b)}$ can be treated as clean image from noisy image $\mathbf{x}_{k+1}^{(b)}$ with noise level $\bar{\sigma}_t = \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}$. Letting $\bar{\sigma}_t = \sqrt{\lambda/\mu}$, with $\nabla_{\mathbf{x}}\mathcal{P}(\mathbf{x}) = -\nabla_{\mathbf{x}} \log p(\mathbf{x}) = -\mathbf{s}_\theta(\mathbf{x})$ [59], Eq. (12) can be rewritten as:

$$\mathbf{z}_{k+1}^{(b)} \approx \mathbf{x}_{k+1}^{(b)} + \frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}\mathbf{s}_\theta(\mathbf{x}_{k+1}^{(b)}, t). \tag{13}$$

Hence, we can perceive $\mathbf{z}_{k+1}^{(b)}$ as the clean three-channel image $\tilde{\mathbf{x}}_{k+1}^{(b)}$ reversed from $\mathbf{x}_{k+1}^{(b)}$.

3.3. Diffusion Adaptation for MSI

Applying an RGB pre-trained denoising diffusion model directly to MSI would cause issues such as band number mismatching, insufficient spectral correlation, and wavelength mismatching. This section will investigate these problems.

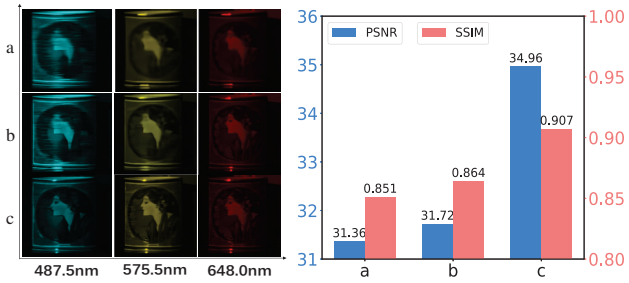


Figure 3. Visual effects and PSNR/SSIM presentation of (a) independently selecting non-overlapping bands method, (b) spectral correlation modeling, and (c) wavelength matching method of *Scene* 1 of 3 (out of 28) spectral channels.

Spectral Correlation Modeling. MSIs exhibit spectral correlation between neighboring bands, denoted as $[B_{i-1}, B_i, B_{i+1}]$. One approach to address this correlation is to partition the MSIs into distinct, non-overlapping bands,

$$C_k = [B_{i-1}, B_i, B_{i+1}], C_{k+1} = [B_{i+2}, B_{i+3}, B_{i+4}], \tag{14}$$

but it just models the part spectral correlation which may cause pixel jump between B_{i+1} and B_{i+2}. Here, to model the spectral correlation, for each band reconstruction, we extract adjacent bands for combination,

$$C_k = [B_{i-1}, B_i, B_{i+1}], C_{k+1} = [B_i, B_{i+1}, B_{i+2}], \tag{15}$$

the combined representation serves as the input for the diffusion model. Subsequently, the corresponding band from the output is selected as the recovered band R_i for the MSIs,

$$R_i = D(C). \tag{16}$$

Quality Comparison: The quality (Q) of the reconstructed MSIs obtained through the spectral correlation modeling method is significantly superior compared to individually selecting non-overlapping bands as shown in Fig. 3, i.e.,

$$Q(D(C)) > Q(D([B_{i-1}, B_i, B_{i+1}])). \tag{17}$$

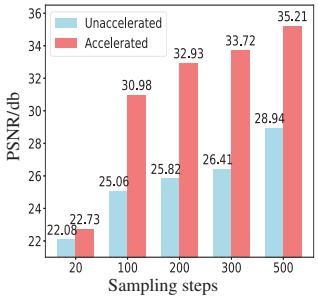
Wavelength Matching. Based on previous experiments illustrated in Fig. 3, it was observed that the reconstruction performance of forward bands was significantly inferior compared to later bands. Analyzing the *Spectral Bands and Range* within the simulated dataset revealed the division of MSIs into 28 spectral bands spanning from 450nm to 720nm,

$$\text{Bands} = \{B_i\}_{i=1}^{28}, \quad \lambda(B_i) \in [450, 720]. \tag{18}$$

While the spectral bands of the RGB image are only a subset of these, i.e.,

$$\lambda(\text{RGB}) = \{660, 520, 450\} \subset [453, 720], \text{RGB} \subset \text{MSIs}. \tag{19}$$

Hence, establishing wavelength matching (WM) between MSIs and RGB images is imperative. In the context of recovering bands with wavelengths significantly distant from



steps \ time(s)	unaccelerated	accelerated
20	31.04	30.36 0.68↑
100	128.03	130.15 2.12↓
200	250.83	251.98 1.12↓
300	395.69	372.77 22.92↑
500	899.20	721.50 177.7↑

Time of each step

Figure 4. Effect of sampling steps and acceleration algorithm on scene5 of simulation dataset on PSNR and time.

RGB images, our DiffSCI method integrates them with two bands featuring matched wavelengths, thereby mitigating interference arising from wavelength mismatching,

$$\mathrm{WM}(B_i) = \mathrm{Merge}(B_i, B_{i+n}, B_{i+m}). \tag{20}$$

Enhanced Metrics: Experimental findings demonstrate significant improvement in both PSNR and SSIM when employing this approach in conjunction with previous spectral correlation modeling methods, as illustrated in Fig. 3.

3.4. Acceleration Algorithm

Motivated by the fact that the sampling process of diffusion model is time-consuming and unconditional, we employ an acceleration algorithm to achieve faster and more efficient sampling. As mentioned in Eq. (11), current methods usually calculate residuals by $(\mathbf{y} - \Phi\mathbf{z}_k)$, which only uses information about the current $\mathbf{z}_k$ for iterative updates. As a result, this approach leads to slow convergence speed and fails to effectively address the issue of data proximity.

On this basis, we introduce a variable $\mathbf{y}_1$, which can be defined as $\mathbf{y}_1 = \mathbf{y}_1 + (\mathbf{y} - \mathbf{\Phi}\mathbf{z}_k)$, can be treated as the accumulation of residuals and calculate residuals by calculating $(\mathbf{y}_1 - \mathbf{\Phi}\mathbf{z}_k)$ iteratively. On the one hand, $\mathbf{y}_1$ can be used to incorporate more residual information for updating $\mathbf{z}$, thereby improving reconstruction quality. On the other hand, a form similar to Nesterov acceleration [40] is employed to expedite the convergence speed.

Accumulation of Residuals. Since $\mathbf{y}_1$ is updated at each iteration, it contains all the residual information from previous iterations. This means that when we update $\mathbf{z}$ using $\mathbf{y}_1$, we are effectively utilizing information from all previous iterations, not just the most recent one.

Methods Pertaining to Nesterov-Type Acceleration. The closed-form solution Eq. (11) can be rewritten as:

$$\mathbf{x}_{k+1} = \mathbf{z}_k + \mathbf{\Phi}^T[\mathbf{y}_1 - \mathbf{\Phi}\mathbf{z}_k] \oslash [Diag(\mathbf{\Phi}\mathbf{\Phi}^T) + \mu] \tag{21}$$

$$= \mathbf{z}_k + \mathbf{\Phi}^T[\sum_{i=1}^{k}(\mathbf{y} - \mathbf{\Phi}\mathbf{z}_i) - \mathbf{\Phi}\mathbf{z}_k] \oslash [Diag(\mathbf{\Phi}\mathbf{\Phi}^T) + \mu].$$

Thus, we can approximate that $\mathbf{x}_{k+1}$ is derived from $\sum_{i=1}^{k}(\mathbf{y} - \mathbf{\Phi}\mathbf{z}_i)$ and $\mathbf{z}_k$, resembling Nesterov's acceleration concept. This enhances the efficacy of the data fidelity term and accelerates the overall convergence rate of the algorithm, as evidenced by experimental comparisons in Fig. 4.

Meanwhile, we define *guidance_scale* (sc) as the iterative step size as the data subproblem and test the effect of different sc on the results, which are shown in Fig. 10.

Algorithm 1 DiffSCI sampling

Require: $\mathbf{s}_\theta$, T,B, $\mathbf{y}$, $\mathbf{\Phi}$, σ_n, $\{\bar{\sigma}_t\}_{t=1}^{T}$, ζ, λ

Initialize $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{y}_1 = 0$, pre-calculate $\rho_t \triangleq \lambda\sigma_n^2/\bar{\sigma}_t^2$.
for $t = T$ **to** 1 **do**
  for $b = 1$ **to** B **do**
    $\mathbf{x}_t^{(b)} = \mathrm{WM}(\mathbf{B}_b)$ *// wavelength mathcing method*
    $\tilde{\mathbf{x}}_t^{(b)} = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t^{(b)} + (1-\bar{\alpha}_t)\mathbf{s}_\theta(\mathbf{x}_t^{(b)}, t))$ *//predict clean image from $\mathbf{x}_t^{(b)}$ with score based model*
  end for
  Get $\tilde{\mathbf{x}}_t$ *// combination*
  $\mathbf{y}_1 = \mathbf{y}_1 + (\mathbf{y} - \mathbf{\Phi}\tilde{\mathbf{x}}_t)$ *// calculate and accumulate residuals*
  $\hat{\mathbf{x}}_0^{(t)} = \tilde{\mathbf{x}}_t + sc \cdot \mathbf{\Phi}^T(\mathbf{y}_1 - \mathbf{\Phi}\tilde{\mathbf{x}}_t) \oslash [Diag(\mathbf{\Phi}\mathbf{\Phi}^T) + \rho_t]$ *// acceleration for data subproblem*
  $\hat{\epsilon} = \frac{1}{\sqrt{1-\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\hat{\mathbf{x}}_0^{(t)})$
  $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\mathbf{x}}_0^{(t)} + \sqrt{1-\bar{\alpha}_{t-1}}(\sqrt{1-\zeta}\hat{\epsilon} + \sqrt{\zeta}\epsilon_t)$ *// diffusion to $\mathbf{x}_{t-1}$ to finish one step sampling*
end for
return $\mathbf{x}_0$

3.5. DiffSCI Method

In DiffSCI, we embed diffusion model into SCI via PnP framework. To elaborate, we can rewrite it as:

$$\mathbf{x}_t^{(b)} \xleftarrow{\mathrm{WM}(B_b)} \mathbf{x}_t, \tag{22}$$

$$\tilde{\mathbf{x}}_t^{(b)} = \arg\min_{\mathbf{z}^{(b)}} \frac{1}{2\bar{\sigma}_t^2}\|\mathbf{z}^{(b)} - \mathbf{x}_t^{(b)}\|^2 + \mathcal{P}(\mathbf{z}^{(b)}), \tag{23}$$

$$\tilde{\mathbf{x}}_t \xleftarrow{\text{combination}} \tilde{\mathbf{x}}_t^{(b)}, \tag{24}$$

$$\hat{\mathbf{x}}_0^{(t)} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}(\mathbf{x})\|^2 + \rho_t\|\mathbf{x} - \tilde{\mathbf{x}}_t\|^2, \tag{25}$$

$$\mathbf{x}_{t-1} \leftarrow \hat{\mathbf{x}}_0^{(t)}, \tag{26}$$

where $\rho_t = \lambda(\sigma_n/\bar{\sigma}_t)^2$, $\mathbf{x}_t$ is noisy MSI at timestep t, $\mathbf{x}_t^{(b)}$ denotes the three-channel image at timestep t obtained by $\mathrm{WM}(B_b)$ from $\mathbf{x}_t$, $\tilde{\mathbf{x}}_t^{(b)}$ is noiseless three-channel image of $\mathbf{x}_t^{(b)}$ and $\tilde{\mathbf{x}}_t$ denotes noiseless MSI through combination.

DiffSCI Sampling. According to previous discussion, the clean estimated MSI $\hat{\mathbf{x}}_0^{(t)}$ can be obtained from $\mathbf{x}_t$ with the condition $\mathbf{y}$. However, this estimation is not accurate, we can add noise and diffusion to timestep $t-1$ as Eq. (26). $\hat{\mathbf{x}}_0^{(t)}$ with condition $\mathbf{y}$ can be firstly gotten, whose conditional

Table 1. Comparisons between DiffSCI and SOTA methods on 10 simulation scenes (S1∼S10). Category, PSNR (upper entry in each cell), and SSIM (lower entry in each cell) are reported. The best and second best results are highlighted in bold and underlined, respectively.

Algorithms	Category	Reference	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg
TwIST [3]	Model	TIP 2007	25.16 0.700	23.02 0.604	21.40 0.711	30.19 0.851	21.41 0.635	20.95 0.644	22.20 0.643	21.82 0.650	22.42 0.690	22.67 0.569	23.12 0.669
GAP-TV [53]	Model	ICIP 2016	26.04 0.817	21.66 0.724	24.86 0.732	30.51 0.875	24.33 0.778	25.11 0.790	18.28 0.730	23.94 0.780	21.77 0.732	23.08 0.721	23.96 0.768
DeSCI [28]	Model	TPAMI 2019	28.38 0.803	26.00 0.701	23.11 0.730	28.26 0.855	25.41 0.778	24.66 0.764	24.96 0.725	24.15 0.747	23.56 0.701	24.17 0.677	25.27 0.748
λ-Net [39]	CNN (Supervised)	ICCV 2019	30.10 0.849	28.49 0.805	27.73 0.870	37.01 0.934	26.19 0.817	28.64 0.853	26.47 0.806	26.09 0.831	27.50 0.826	27.13 0.816	28.53 0.841
TSA-Net [35]	CNN (Supervised)	ECCV 2020	32.31 0.894	31.03 0.863	32.15 0.916	37.95 0.958	29.47 0.884	31.06 0.902	30.02 0.880	29.22 0.886	31.14 0.909	29.18 0.861	31.35 0.895
HDNet [23]	Transformer (Supervised)	CVPR 2022	34.96 0.937	35.64 0.943	35.55 0.946	41.64 0.976	32.56 0.948	34.33 0.954	33.27 0.928	32.26 0.945	34.17 0.944	32.22 0.940	34.66 0.946
MST-L [5]	Transformer (Supervised)	CVPR 2022	35.30 0.944	36.13 0.948	35.66 0.954	40.05 0.976	32.84 0.949	34.56 0.955	33.80 0.930	32.74 0.950	34.37 0.944	32.63 0.943	34.81 0.949
MST++ [7]	Transformer (Supervised)	CVPR 2022	35.57 0.945	36.22 0.949	37.00 0.959	**42.86** 0.980	33.27 0.954	35.27 0.960	34.05 0.936	33.50 0.956	36.17 0.956	**33.26** 0.949	**35.72** 0.955
CST-L+ [6]	Transformer (Supervised)	ECCV 2022	**35.64** **0.951**	**36.79** **0.957**	37.71 **0.965**	41.38 **0.981**	32.95 **0.957**	**35.58** **0.966**	34.54 **0.947**	**34.07** **0.964**	35.62 **0.959**	32.82 **0.949**	35.71 **0.960**
DGSMP [24]	Deep Unfolding (Supervised)	CVPR 2021	33.26 0.915	32.09 0.898	33.06 0.925	40.54 0.964	28.86 0.882	33.08 0.937	30.74 0.886	31.55 0.923	31.66 0.911	31.44 0.925	32.63 0.917
ADMM-Net [32]	Deep Unfolding (Supervised)	ICCV 2019	34.03 0.919	33.57 0.904	34.82 0.933	39.46 0.971	31.83 0.924	32.47 0.926	32.01 0.898	30.49 0.907	33.38 0.917	30.55 0.899	33.26 0.920
GAP-Net [38]	Deep Unfolding (Supervised)	IJCV 2023	33.63 0.913	33.19 0.902	33.96 0.931	39.14 0.971	31.44 0.921	32.29 0.927	31.79 0.903	30.25 0.907	33.06 0.916	30.14 0.898	32.89 0.919
PnP-CASSI [58]	PnP (Zero-Shot)	PR 2021	29.09 0.799	28.05 0.708	30.15 0.850	39.17 0.939	27.45 0.798	26.16 0.752	26.92 0.736	24.92 0.710	27.99 0.752	25.58 0.664	28.55 0.771
DIP-HSI [37]	PnP (Zero-Shot)	ICCV 2021	31.32 0.855	25.89 0.699	29.91 0.839	38.69 0.926	27.45 0.796	29.53 0.824	27.46 0.700	27.69 0.802	33.46 0.863	26.10 0.733	29.75 0.803
HLRTF [31]	Tensor Network (Self-Supervised)	CVPR 2022	34.82 0.909	33.83 0.887	34.16 0.958	38.67 0.979	32.27 0.924	32.52 0.906	32.69 0.913	31.28 0.871	36.62 0.958	30.06 0.855	33.69 0.916
DiffSCI	PnP-Diffusion (Zero-Shot)	Ours	34.96 0.907	34.60 0.905	**39.83** 0.949	42.65 0.951	**35.21** 0.946	33.12 0.917	**36.29** 0.944	30.42 0.887	**37.27** 0.931	28.49 0.821	35.28 0.916

distribution is $p(x|y)$, and estimated clean image can be used to calculate the noise with condition $\mathbf{y}$, which is $\hat{\epsilon} = \frac{1}{\sqrt{1-\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\hat{\mathbf{x}}_0(t))$. Then, the diffusion expression like Eq. (6) is:

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\mathbf{x}}_0^{(t)} + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_{\eta_t}^2}\,\hat{\epsilon} + \sigma_{\eta_t}\epsilon_t. \quad (27)$$

Based on previous experience [59], the noise term σ_{η_t} could be set to 0, and hyperparameter ζ can be used to introduce noise to balance ϵ_t and $\hat{\epsilon}$, and Eq. (27) can be rewritten as:

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{\mathbf{x}}_0^{(t)} + \sqrt{1-\bar{\alpha}_{t-1}}(\sqrt{1-\zeta}\hat{\epsilon} + \sqrt{\zeta}\epsilon_t), \quad (28)$$

where ζ controls the variance of the noise added at each step, when $\zeta = 0$, our method becomes a deterministic process.

Finally, we summarize the algorithm for DiffSCI-based MSI reconstruction in Algorithm 1. Further details regarding the model are presented in the supplementary materials.

4. Experiments

4.1. Experiment Setup

Similar to most existing methods [8, 23, 24, 35], we select 10 scenes with spatial size 256×256 and 28 bands from KAIST [11] as simulation dataset. Meanwhile, we select 5 MSIs with spatial size 660×660 and 28 bands, captured by the CASSI system for real dataset [35], then we crop data blocks of size 256×256 for testing. The pre-trained diffusion model uses a model trained by [59].

Parameter Setting. Through all our experiments, we use the same linear noise schedule $\{\beta_t\}$, and DDIM sampling. The shift step is set to 2. And in the wavelength matching method, we choose 21_{th} and 28_{th} bands to form a three-channel image. Meanwhile, we set the reverse initial time step to 600 and set the sampling steps to 20, 100, 200, 300 and 500 respectively for testing. After experiments, we find setting $\lambda = 15$, $\eta = 1$, $\zeta = 1$ in DDIM process and $sc = 1$ in data proximal subproblem can achieve the best results.

Comparisons with SOTA Methods. We test the performance of our proposed DiffSCI method on the simulation dataset. We compare the results of our DiffSCI method with 15 SOTA methods set up strictly according to the paper code including three model-based methods (TwIST [3], GAP-TV [53], DESCI [28]), six E2E methods (λ-Net [39], TSA-NET [35], HDNET [23], MST-L [5], MST++ [7], CST-L-PLUS [6]), three deep unfolding methods (DGSMP [24], GAP-NET [38], ADMM-NET [32]), two PnP methods (PnP-CASSI [58], DIP-MSI [37]) and one tensor network method (HLRTF [31]) on 10 simulation scenes. From Table 1, it can be observed that our unsupervised method has a significant improvement compared to other unsupervised methods. The gap between its performance on PSNR and current super-

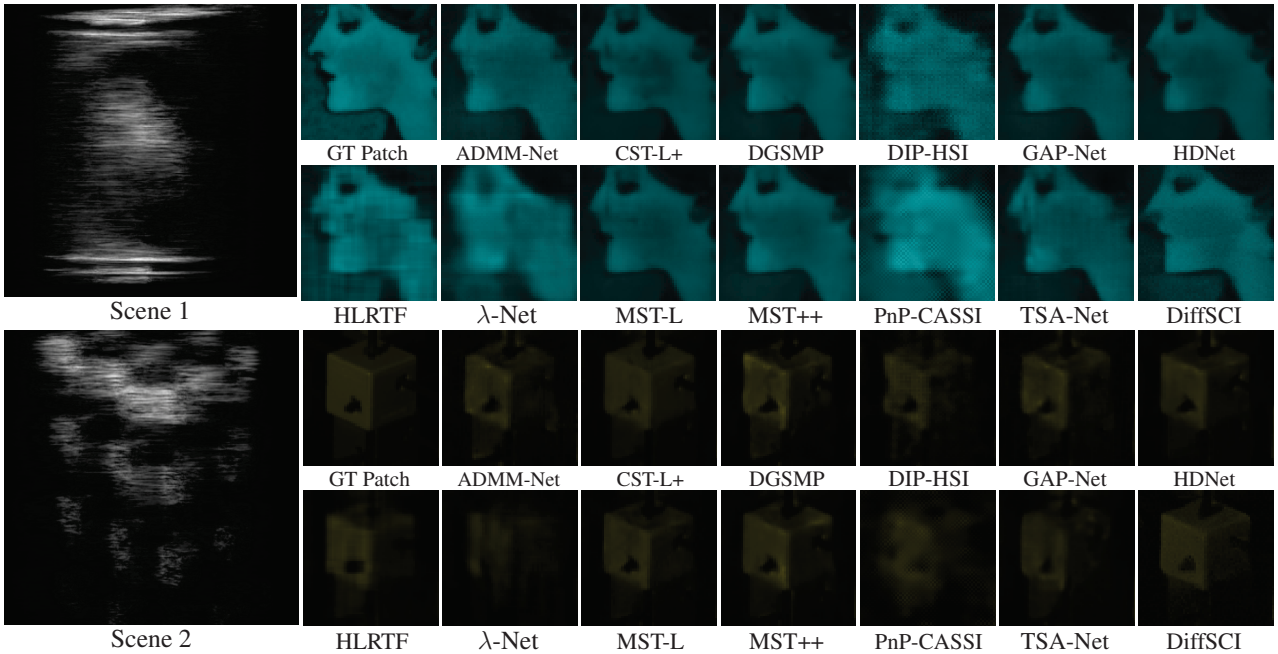


Figure 5. Visual comparison on KAIST dataset. **Top** is $Scene$ 1 at wavelength 487.0nm. **Bottom** is $Scene$ 2 at wavelength 575.5nm

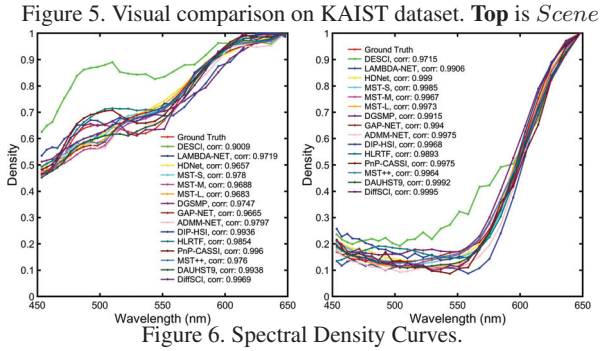


Figure 6. Spectral Density Curves.

vised SOTA methods such as MST-L [5] and MST++ [7] is also narrowing. Moreover, we do not need to retrain a model on MSIs. Therefore, the proposed DiffSCI achieves a balance between flexibility and performance.

4.2. Qualitative Experiments

Results on Simulation Dataset. Fig. 5 shows the display effects of MSI reconstruction between our DiffSCI method and other SOTA methods on the 8_{th} band of $Scene$ 1 (top) and 21_{th} band of $Scene$ 2 (bottom). From the enlarged part of the $Scene$ 1 image, we can see that our DiffSCI provides superior visual effects of detailed contents, cleaner textures, and fewer artifacts compared to other SOTA methods. Furthermore, to demonstrate the powerful generative capabilities of the diffusion model, we can observe the magnified section of $Scene$ 2. Our method makes the edges of the blocks sharper, the shapes and patterns closer to the GT, whereas previous methods either generate over-smoothed results thus losing the complexity of fine-grained structures or introduce artifacts. This suggests that the generative capabilities of diffusion can be effectively applied to reconstruct darker regions, thereby filling in the gaps in the current method.

Fig. 6 presents the density-wavelength spectral curves. The spectral curves from DiffSCI achieve the highest correlation with the reference curves, even exceeding the performance of the current leading method, DAUHST-9 [8]. This demonstrates the superiority of our proposed DiffSCI in terms of spectral-dimension consistency.

Results on Real Dataset. We also test the reconstruction capability of DiffSCI on real dataset. Fig. 7 and Fig. 11 show the visual comparison between DiffSCI and other SOTA methods. It is evident that our reconstruction results are more detailed and have fewer artifacts. Compared to the blurred results reconstructed by other methods, our method DiffSCI demonstrates that the generative ability of the diffusion model can provide good robustness against noise, leading to enhanced results in MSI reconstruction. More experimental results are shown in the supplementary materials.

5. Ablation Study

Effects of Acceleration Algorithm. We propose a residual accumulation method aimed at achieving acceleration. Through experimentation, employing this accelerated algorithm showcases enhancements not only in convergence speed but also in performance, maintaining consistent parameters. Figure 4 demonstrates the impact of the acceleration algorithm on both PSNR and time, utilizing an identical number of sampling steps. Evidently, the accelerated algorithm yields an improvement of 5-6dB in average performance while expediting convergence.

Effects of t_{start}. Our DiffSCI can perform the reverse process from partially noisy images instead of starting the recovery from pure Gaussian noise. To demonstrate the impact of t_{start} on performance briefly, we show how PSNR changes

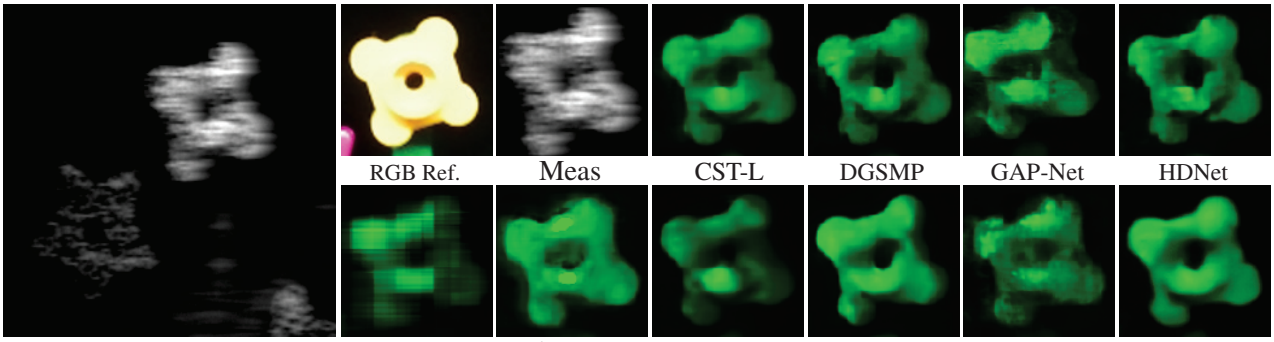
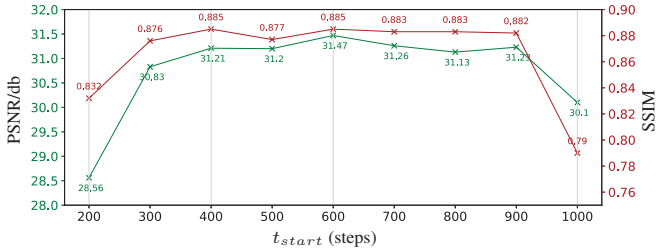
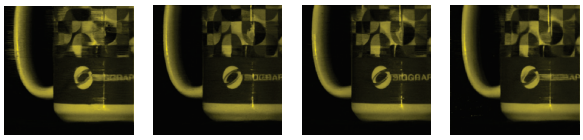


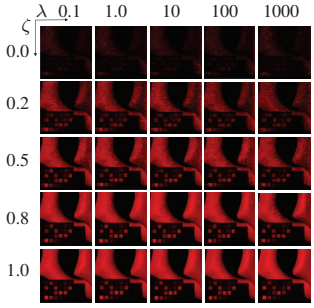
Figure 7. Visual comparison of SCI reconstruction methods on $Scene$ 1 of real dataset at wavelength 536.6nm.





(a) 200 (b) 400 (c) 600 (d) 1000

Figure 8. Effect of t_{start} on $Scene$ 5 of KAIST.



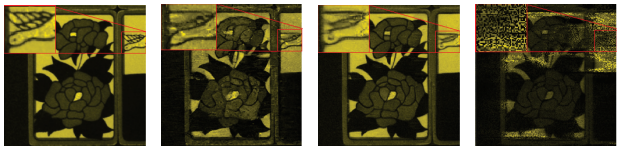
PSNR SSIM	0.1	1.0	10	100	1000
0	0.09 0.023	1.83 0.024	9.47 0.028	15.58 0.144	15.63 0.323
0.2	10.07 0.121	11.54 0.124	16.46 0.153	18.71 0.255	18.04 0.368
0.5	18.57 0.345	19.18 0.349	20.96 0.386	21.22 0.489	20.76 0.539
0.8	26.80 0.668	26.62 0.671	27.04 0.698	26.94 0.730	25.69 0.725
1.0	33.75 0.856	33.68 0.856	33.77 0.874	33.77 0.907	31.27 0.901

Figure 9. Visual comparison (left) and PSNR comparison (right) of the effect of hyperparameters ζ and λ on $Scene$ 3 of KAIST.

in Fig. 8. We select sampling steps with 100 for all experiments and find that our method achieves the best results in terms of PSNR and SSIM at $t_{start} = 600$.

Effects of Sampling Steps. To study the impact of the number of sampling steps on the reconstruction quality assessment parameters PSNR and SSIM, and thus balance the sampling speed with the recovery quality, we conduct experiments with different numbers of sampling steps. As shown in Fig. 4, we set sampling steps $T \in [20, 100, 200, 300, 500]$.

Effects of λ, ζ and sc. DiffSCI has three hyperparameters λ, ζ and sc, which manage the strength of the condition guidance, the level of noise added at each timestep and the update step size in data proximity subproblems. As shown in the left figure of Fig. 9, when ζ approaches 1, we get



(a) GT (b) 0.5 (c) 1.0 (d) 2.0

Figure 10. Effect of sc (0.5, 1.0, 2.0) on $Scene$ 7 of KAIST.

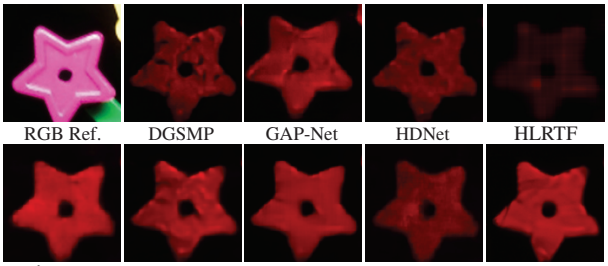


Figure 11. Visual comparison on $Scene$ 1 of real dataset.

the best reconstruction quality. Meanwhile, Fig. 10 demonstrates a close relationship between sc and the quality of the reconstruction.

6. Conclusion

In this paper, we are the first to integrate diffusion model with Plug-and-Play algorithm, applying the generative capabilities of the diffusion model to MSI reconstruction, which compensated for the shortcomings of current methods. Specifically, by utilizing the wavelength matching method and HQS method, we successfully applied the HQS-based diffusion model, which was pre-trained on RGB images, as a denoising prior in MSI reconstruction. Meanwhile, we introduced acceleration algorithms when solving the data subproblem. Experimental results on both simulated and real datasets highlighted the superior adaptability, efficiency, and applicability of DiffSCI compared to SOTA methods.

7. Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62106063, by the Guangdong Natural Science Foundation under Grants 2022A1515010819 and 2022A1515010800, and Guangdong Major Project of Basic and Applied Basic Research under Grant 2023B0303000010.

References

[1] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. 2, 3

[2] V Backman, Michael B Wallace, LT Perelman, JT Arendt, R Gurjar, MG Müller, Q Zhang, G Zonios, E Kline, T McGillican, et al. Detection of preinvasive cancer cells. *Nature*, 2000. 1

[3] José M Bioucas-Dias and Mário AT Figueiredo. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE TIP*, 2007. 6

[4] Marcus Borengasser, William S Hungate, and Russell Watkins. Hyperspectral remote sensing: principles and applications. *CRC press*, 2007. 1

[5] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *CVPR*, 2022. 1, 6, 7

[6] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In *ECCV*, pages 686–704. Springer, 2022. 6

[7] 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 *CVPRW*, 2022. 6, 7

[8] 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. 2, 4, 6, 7

[9] Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan, Qionghai Dai, Lawrence Carin, and David J. Brady. Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE SPM*, 2016. 1

[10] Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE TCI*, 3(1):84–98, 2016. 2

[11] Inchang Choi, MH Kim, D Gutierrez, DS Jeon, and G Nam. High-quality hyperspectral reconstruction using a spectral prior. In *Technical report*, 2017. 6

[12] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021. 1, 2

[13] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 2

[14] 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. 4

[15] Kamil Deja, Anna Kuzina, Tomasz Trzcinski, and Jakub Tomczak. On analyzing generative and denoising capabilities of diffusion-based deep generative models. In *NeurIPS*, pages 26218–26229, 2022. 4

[16] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, pages 8780–8794, 2021. 1

[17] David L Donoho. Compressed sensing. *IEEE TIT*, 2006. 2

[18] Hao Du, Xin Tong, Xun Cao, and Stephen Lin. A prism-based system for multispectral video acquisition. In *ICCV*, 2009. 1

[19] Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express*, 2007. 2

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

[21] Alexander FH Goetz, Gregg Vane, Jerry E Solomon, and Barrett N Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153, 1985. 1

[22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, pages 6840–6851, 2020. 2

[23] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *CVPR*, 2022. 1, 6

[24] Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale mixture prior for spectral compressive imaging. In *CVPR*, 2021. 6

[25] Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *JMLR*, 6(4), 2005. 3

[26] Shirin Jalali and Xin Yuan. Snapshot compressed sensing: Performance bounds and algorithms. *IEEE TIT*, 2019. 2

[27] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS*, pages 23593–23606, 2022. 1

[28] Yang Liu, Xin Yuan, Jinli Suo, David Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE TPAMI*, 2019. 6

[29] Patrick Llull, Xuejun Liao, Xin Yuan, Jianbo Yang, David Kittle, Lawrence Carin, Guillermo Sapiro, and David J Brady. Coded aperture compressive temporal imaging. *Optics Express*, 2013. 1

[30] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, 2014. 1

[31] Yisi Luo, Xi-Le Zhao, Deyu Meng, and Tai-Xiang Jiang. Hlrtf: Hierarchical low-rank tensor factorization for inverse problems in multi-dimensional imaging. In *CVPR*, pages 19303–19312, 2022. 6

[32] Jiawei Ma, Xiao-Yang Liu, Zheng Shou, and Xin Yuan. Deep tensor admm-net for snapshot compressive imaging. In *ICCV*, 2019. 2, 6

[33] Xiao Ma, Xin Yuan, Chen Fu, and Gonzalo R Arce. Led-based compressive spectral-temporal imaging. *Optics Express*, 2021. 1

[34] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE TGRS*, 2004. 1

[35] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *ECCV*, 2020. 1, 2, 6

[36] Ziyi Meng, Mu Qiao, Jiawei Ma, Zhenming Yu, Kun Xu, and Xin Yuan. Snapshot multispectral endomicroscopy. *Optics Letters*, 2020. 1

[37] Ziyi Meng, Zhenming Yu, Kun Xu, and Xin Yuan. Self-supervised neural networks for spectral snapshot compressive imaging. In *ICCV*, 2021. 2, 6

[38] Ziyi Meng, Xin Yuan, and Shirin Jalali. Deep unfolding for snapshot compressive imaging. *IJCV*, 131(11):2933–2958, 2023. 6

[39] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos. l-net: Reconstruct hyperspectral images from a snapshot measurement. In *ICCV*, 2019. 1, 6

[40] Yurii Nesterov. A method for unconstrained convex minimization problem with the rate of convergence o (1/k2). In *Dokl. Akad. Nauk. SSSR*, page 543, 1983. 5

[41] Cindy M Nguyen, Eric R Chan, Alexander W Bergman, and Gordon Wetzstein. Diffusion in the dark: A diffusion model for low-light text recognition. *arXiv preprint arXiv:2303.04291*, 2023. 3

[42] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICML*, pages 8162–8171. PMLR, 2021. 1

[43] Mu Qiao, Xuan Liu, and Xin Yuan. Snapshot spatial–temporal compressive imaging. *Optics letters*, 2020. 2

[44] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 2

[45] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *NeurIPS*, 2019. 3

[46] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 2, 3

[47] Prasad S Thenkabail, Murali Krishna Gumma, Pardhasaradhi Teluguntla, and AM Irshad. Hyperspectral remote sensing of vegetation and agricultural crops. *Photogrammetric Engineering & Remote Sensing (TSI)*, 80(8):695–723, 2014. 1

[48] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE TIT*, 2007. 2

[49] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pages 945–948. IEEE, 2013. 2

[50] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 2008. 1, 2

[51] Ashwin A Wagadarikar, Nikos P Pitsianis, Xiaobai Sun, and David J Brady. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics Express*, 2009. 1

[52] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *ICCV*, pages 12302–12311, 2023. 3

[53] Xin Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *ICIP*, 2016. 6

[54] Xin Yuan, Tsung-Han Tsai, Ruoyu Zhu, Patrick Llull, David Brady, and Lawrence Carin. Compressive hyperspectral imaging with side information. *IEEE JSTSP*, 2015. 1

[55] Xin Yuan, Yang Liu, Jinli Suo, and Qionghai Dai. Plug-and-play algorithms for large-scale snapshot compressive imaging. In *CVPR*, 2020. 2

[56] Xin Yuan, Yang Liu, Jinli Suo, Fredo Durand, and Qionghai Dai. Plug-and-play algorithms for video snapshot compressive imaging. *IEEE TPAMI*, 2021. 2

[57] Yuan Yuan, Xiangtao Zheng, and Xiaoqiang Lu. Hyperspectral image superresolution by transfer learning. *IEEE JSTAEORS*, 2017. 1

[58] Siming Zheng, Yang Liu, Ziyi Meng, Mu Qiao, Zhishen Tong, Xiaoyu Yang, Shensheng Han, and Xin Yuan. Deep plug-and-play priors for spectral snapshot compressive imaging. *PR*, 9 (2):B18–B29, 2021. 2, 4, 6

[59] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *CVPR*, pages 1219–1229, 2023. 1, 4, 6