Supplementary Materials: Pixel Adaptive Deep Unfolding Transformer for Hyperspectral Image Reconstruction

In this supplementary material, we present more analysis and results of our proposed Pixel Adaptive Deep Unfolding Transformer (PADUT).

- We show the detailed derivation of Half Quadratic Splitting algorithm in Section A.
- We provide the comparison and visualization of weight maps in Section B to further illustrate our proposed pixel adaptive data module.
- We give a detailed analysis of our Non-local Spectral Transformer and show ablation study in Section C.
- We show more visual comparisons between our method and the state-of-the-art methods in Section D.

A. Detailed Derivation of HQS Algorithm

In this section, we give a detailed derivation of the Half Quadratic Splitting (HQS) algorithm for the optimization of x, as only a brief version was presented in the main paper. As mentioned in the main paper, the HQS algorithm approximates the optimization of the reconstruction problem for x through the following iterative subproblem:

$$\mathbf{x}^{k+1} = (\mathbf{\Phi}^T\mathbf{\Phi} + \mu\mathbf{I})^{-1}(\mathbf{\Phi}^T\mathbf{y} + \mu\mathbf{z}^k). \tag{1}$$

For the HSI reconstruction problem, sensing matrix $\mathbf{\Phi}$ is a shifting binary mask and $\mathbf{\Phi}\mathbf{\Phi}^T$ is a diagonal matrix:

$$\mathbf{\Phi}\mathbf{\Phi}^T = \mathrm{diag}\left\{\phi_1, \phi_2, ..., \phi_M\right\}, \tag{2}$$

where M is the number of row in $\mathbf{\Phi}$. The inverse matrix in Eq (1) can be calculated as follows:

$$(\mathbf{\Phi}^T\mathbf{\Phi}+\mu\mathbf{I})^{-1} = \mu^{-1}\mathbf{I}-\mu^{-1}\mathbf{\Phi}^T(\mathbf{I}+\mathbf{\Phi}\mu^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mu^{-1}, \tag{3}$$

where $(\mathbf{I} + \mathbf{\Phi}\mu^{-1}\mathbf{\Phi}^T)^{-1}$ is formulated with Eq. (2) as:

$$(\mathbf{I}+\mathbf{\Phi}\mu^{-1}\mathbf{\Phi}^T)^{-1} = \mathrm{diag}\left\{\frac{\mu}{\mu+\phi_1}, \frac{\mu}{\mu+\phi_2}, ..., \frac{\mu}{\mu+\phi_M}\right\}. \tag{4}$$

According to Eq. (4), Eq.(1) is given by:

$$\begin{aligned}\mathbf{x}^{k+1} &= [\mu^{-1}\mathbf{I} - \mu^{-1}\mathbf{\Phi}^T(\mathbf{I} + \mathbf{\Phi}\mu^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mu^{-1}] \\ &\quad (\mathbf{\Phi}^T\mathbf{y} + \mu\mathbf{z}^k) \qquad (5)\\ &= z^k + \mu^{-2}\mathbf{\Phi}^T[\mu\mathbf{I} - (\mathbf{I} + \mu^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{\Phi}^T]\mathbf{y} \\ &\quad - \mu^{-1}\mathbf{\Phi}^T(\mathbf{I} + \mathbf{\Phi}\mu^{-1}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{z}^k \qquad (6)\\ &= \mathbf{z}^k + \mathbf{\Phi}^T\frac{\mathbf{y} - \mathbf{\Phi}\mathbf{z}^k}{\mu\mathbf{I} + \mathbf{\Phi}\mathbf{\Phi}^T} \qquad (7)\\ &= \mathbf{z}^k + \frac{\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}}{\mu(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1} + \mathbf{I}}(\mathbf{y} - \mathbf{\Phi}\mathbf{z}^k). \qquad (8)\end{aligned}$$

Since $\mathbf{\Phi}\mathbf{\Phi}^T$ is related to the sensing matrix and is pre-calculated, we simply the learning procedure as:

$$\mathbf{x}^{k+1} = \mathbf{z}^k + \sigma\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}(\mathbf{y} - \mathbf{\Phi}\mathbf{z}^k). \tag{9}$$

In our Pixel Adaptive module, we learn 3D parameters σ for each HSI considering the difference of each pixel in the degradation process.

B. Visualization of Pixel Adaptive Module

In this section, we demonstrate the effectiveness of the proposed Pixel Adaptive (PA) module. First, we replace our PA module with other learning strategies for parameter estimation. We compare our pixel adaptive module with single parameter learning and degradation-aware (DA) proposed in DAUHST [1]. Specifically, the details of three ways of parameter learning are illustrated in Figure 1. Both Figure 1 (a) and Figure 1 (b) learn a single parameter, while our PA method learns 3D parameters for pixel-level adaptive reconstruction. Table 1 presents the quantitative results of the three methods. As DA and our PA learn the parameters based on the input HSI and degradation matrix, they obtain better results than the one that only uses a learnable parameter. In addition, because our PA learns a weight for each pixel, our result is more promising.

Furthermore, we conduct a visual analysis of the intermediate weight map obtained by our PA module. The visualization of our PA module is shown in Figure 2, where we

	(a) Single Parameter	(b) DAU [1]	(c) PA (Ours)
Params	1.32M	1.36M	1.35M
GFLOPs	20.82	21.75	22.91
PSNR	36.21	36.45	36.95
SSIM	0.959	0.959	0.962

Table 1: Ablation study of different learning strategies in the data module to learn parameter.

only show the features of one channel. Our PA module gets different weights in different regions across the image. In particular, it focuses more on the textured areas which are more difficult to recover than flat areas.

C. Analysis of Non-local Spectral Transformer

In this section, we present a further analysis of our proposed Non-local Spectral Transformer (NST). Firstly, we provide a detailed introduction to our prior module, which consists of several layers of NST and other components. Secondly, we present additional experimental ablation results to demonstrate the effectiveness of our proposed NST.

The details of our prior module are depicted in Figure 4. In particular, a shift operation in the spatial dimension is performed at intervals of one layer, facilitating interactions between adjacent 3D cubes.

We compare our proposed NST with three Transformer-based networks: Restormer [3], Swin [2], and Half Shuffle Transformer (HST). HST is derived from DAUHST [1]. Additionally, we compare our NST with a variant that omits the shift operation. We only replace the Transformer layer and other settings keep the same. The experimental results are presented in Table 2. Our proposed NST leverages the features of 3D data and achieves higher PSNR with less computation cost.

	Restormer [3]	Swin [2]	HST [1]	No-Shift	NST (Ours)
Params	1.35M	1.06M	1.43N	1.35M	1.35M
GFLOPs	22.91	25.05	26.23	22.91	22.91
PSNR	36.84	35.32	36.42	36.67	36.95
SSIM	0.962	0.954	0.960	0.961	0.962

Table 2: Ablation study on the NST

D. Additional Visual Comparisons

We present additional visual results in Figure 6, along with their corresponding spectral curves displayed in Figure 3. Traditional model-based methods (Twist and GAP-TV) demonstrate poor performance in the magnified area. Although DAUHST-9stg shows relatively effective overall recovery, it fails to recover sufficient detailed information. In contrast, our method achieves the most satisfactory reconstruction results. The corresponding spectral curves prove that our method has superior reconstruction performance in terms of spectral fidelity.

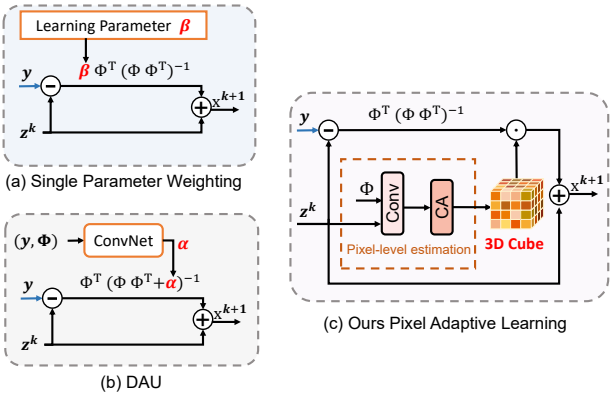
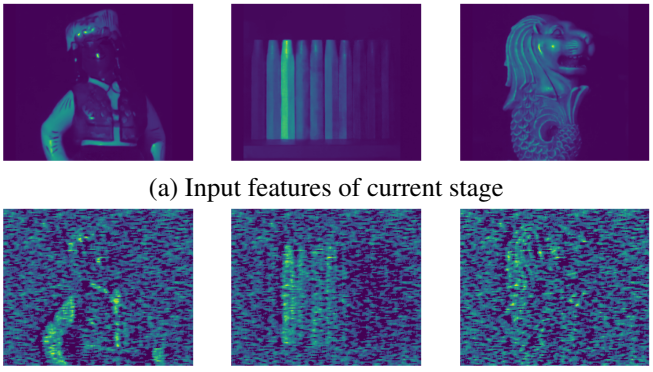


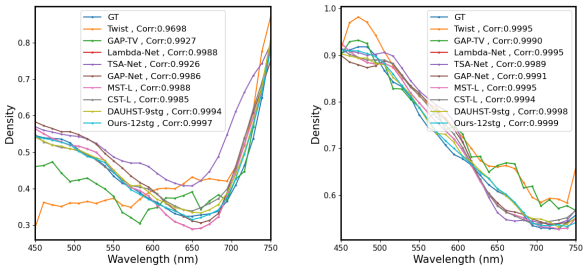
Figure 1: Different learning strategies of parameter learning. (a) Learnable parameter (b) Parameter estimated from convolutional neural network (c) Ours. Pixel-level adaptive parameter estimation



(a) Input features of current stage

(b) Visualization of weights from corresponding PA module

Figure 2: Visualization of the features and pixel-level weights in our PA module.



(a) Simulated scene 2 (b)Simulated scene 5

Figure 3: Spectral curves of reconstruction results on simulated HSIs.

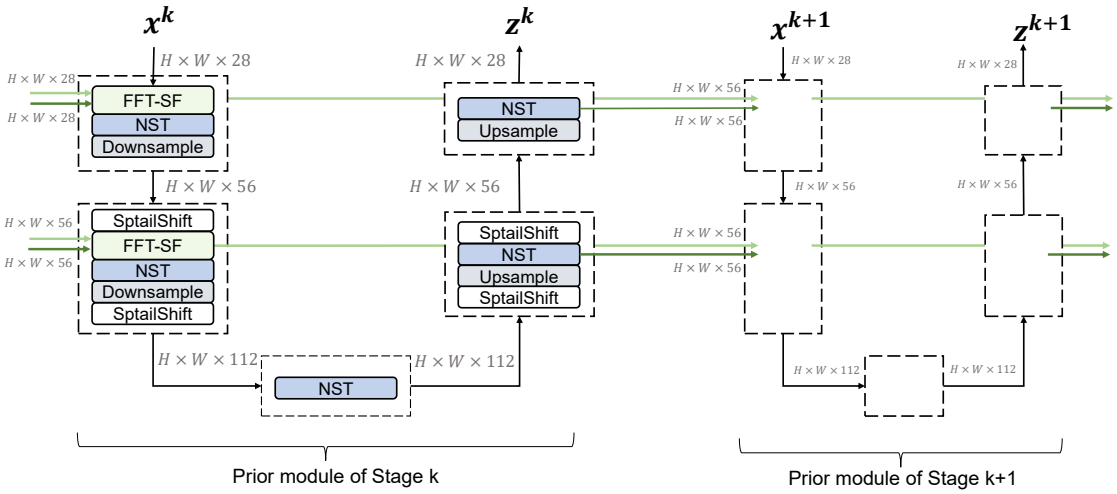


Figure 4: Details of prior module.

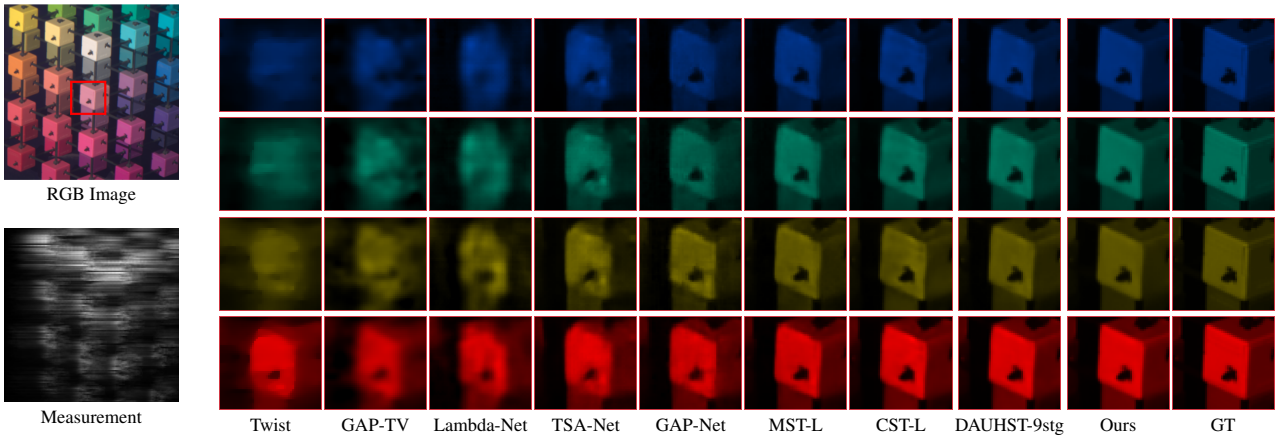


Figure 5: Simulated HSI reconstruction results with 4 channels (channel 4, 9, 21 and 27) on scene 2.

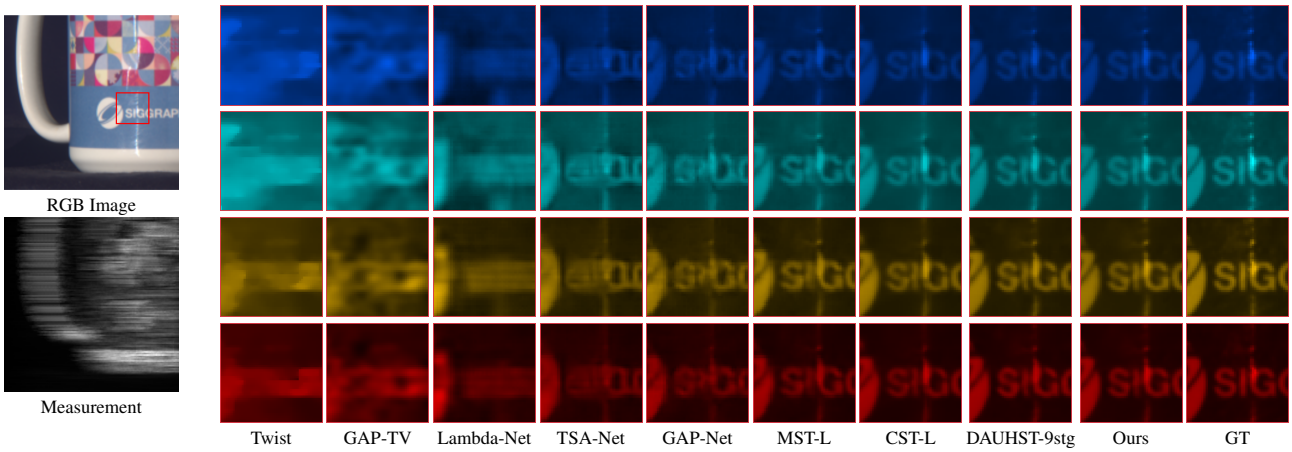


Figure 6: Simulated HSI reconstruction results with 4 channels (channel 4, 18, 22 and 26) on scene 5.

References

[1] Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. 2022. 1, 2

[2] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. pages 10012–10022, 2021. 2

[3] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5728–5739, 2022. 2