International Journal on Semantic Web and Information Systems
Volume 20 • Issue 1

Reversible-Prior-Based Spectral-Spatial Transformer for Efficient Hyperspectral Image Reconstruction

Zeyu Cai, Southeast University, China
Zheng Liu, Nanjing Institute of Argicultural Mechanization, Ministry of Agriculture and Rural Affairs, China*
Jian Yu, Southeast University, China
Ziyu Zhang, Nanjing University, China
Feipeng Da, Southeast University, China
Chengqian Jin, Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, China

ABSTRACT

The task of reconstructing a 3D cube from a 2D measurement is not well-defined in spectral imaging. Unfortunately, existing Deep Unfolding Network (DU) and End-to-End (E2E) approaches can't strike an optimal balance between computational complexity and reconstruction quality. The goal of this study is to think about ways to merge the E2E's violent mapping with DU's iterative method. Our proposed deep learning framework, the Reversible-prior-based Spectral-Spatial Transformer, combines the high-quality reconstruction capabilities of DU with the advantages of having fewer parameters and lower computing cost, similar to the E2E approach. SST-ReversibleNet uses a reversible prior to project the end-to-end mapping reconstruction results back into the measurement space, construct the residuals between the reprojection and the actual measurement, and improve reconstruction accuracy. Extensive trials show that our SST-ReversibleNet outperforms cutting-edge approaches by at least 0.8 dB and only use 34.3% Params and 44.1% giga floating-point operations per second (GFLOP).

KEYWORDS

CASSI, Hyperspectral-Imaging, Reconstruction, Reversible, Spectral-Spatial

INTRODUCTION

With rich and unique features (Cao et al., 2016), hyperspectral images (HSIs) have been widely used for analysis and scene applications such as remote sensing (Deng et al., 2023), precision agriculture (Ishida et al., 2018), national security (Udin et al., 2019), environmental protection (Wright et al., 2019), and astronomical observations (De Angelis et al., 2015). In computer vision, HSIs can be extensively used for object tracking (Li et al., 2022; Kim et al., 2012),

DOI: 10.4018/IJSWIS.344457 *Corresponding Author

material classification (Yu et al., 2022; Hong et al., 2022), feature extraction (Li et al., 2020), and medical image analysis (Liu et al., 2019).

To obtain spectral images, traditional methods typically scan scenes along the one-dimensional (1D) or two-dimensional (2D) spatial dimension, or along spectral channels, sacrificing time through multiple exposures to reconstruct the spectral data of the scene. Although traditional methods perform well in terms of spectral detection range and accuracy (Wang et al., 2021), they are unsuitable for dynamic detection and therefore consumer applications. Recently, researchers have used developments in compressed sensing (CS) theory to collect HSIs using snapshot compressed imaging (SCI) systems (Du et al., 2009), which combine information from snapshots throughout the process of reducing the spectral dimensionality to a single 2D observation. Coded aperture snapshot spectrum imaging (CASSI) (Wagadarikar et al., 2008) is considered a very promising field among the current SCI systems.

Although the CASSI technology modulates the spectral three-dimensional (3D) cube with a coded mask before dispersing it, the redundant picture information can be used to recreate the entire 3D cube. Spectral reconstruction methods are categorized into four types: conventional, deep unfolding network (DU), end-to-end (E2E), and plug-and-play (PnP).

Traditional methods perform reconstructions based on over-complete dictionaries or sparse spectral features that rely on hand-crafted priors and assumptions (Zhang et al., 2019; Wang et al., 2016). The primary limitation of these conventional techniques is the requirement for human parameter adjustment, leading to inadequate resilience and sluggish repair. Deep learning approaches have shown significant prowess in recent years in image production (Qian et al., 2022; Yu et al., 2018; Li et al., 2019; Chopra et al., 2022), image retrieval (Nhi et al., 2022; Chu et al., 2022; Wang et al., 2020), image-semantic analysis (Hu et al., 2022), image classification (Ghoneim et al., 2018; Mandle et al., 2022) and reconstruction (Arnab et al., 2021), such as image denoising, image super-resolution, and rain and fog removal (Jia et al., 2023; Liu et al., 2022; Liang et al., 2022), and have also been applied to spectral image reconstruction. PnP introduces a denoising module based on the traditional method, but with limited improvement in reconstruction speed and accuracy. The current state-of-the-art (SOTA) methods all come from E2E and DU. The E2E directly establishes the mapping between the measurement and truth data, and the DU uses a depth module to simulate the iterations in a convex optimization algorithm. Although both E2E and DU have achieved good performance, there are still limitations to the current methods.

1. The E2E method is similar to an open-loop control system where the measurements no longer guide the reconstruction process during the reconstruction, in addition to lacking a DU-like iterative framework and interpretability. As a result, E2E is inefficient at improving network performance by increasing network depth, and limits the ability to further improve accuracy.
2. The DU networks are based on convex optimization algorithms, but require transposition and invertible operations on the operation matrix during iteration. These conditions limit the structure of the network module and impose requirements on the design of the coding mask, as described in Sec. 2.
3. The denoising modules in the Transformer-based E2E methods and DU networks learn either the global self-similarity of the spectral dimension or the local correlation of the spatial dimension, disregarding the spectral and spatial correlations that exist on a global scale within the spectral cube.

The motivation of this paper is to find a new framework that employs end-to-end mapping and is able to guide the reconstruction process through measurements. Its structure is similar to the iterations of the DU but is not subject to the constraints of the convex optimization method, thus combining the advantages of both E2E and DU methods. Furthermore, the authors seek a denoising module that learns both the self-similarity of the transformer-based spectral dimension and the spatial global

dependence of the transformer-based spatial dimension, taking into account memory consumption and computational complexity.

To address the above issues, we, inspired by the reversible nature of the optical path, propose a framework based on the reversible optical path prior (Reversible-prior). The new framework uses a reversible prior to project the reconstruction results of the end-to-end mapping back into the measurement space after each iteration, and to construct the residuals between the reprojection and the actual measurement. By incorporating residuals and measurement errors into the learning process, as well as by utilizing raw measurement data from the sensors and network guidance, the model's reconstruction capability can be significantly enhanced. The structure of this improved model is illustrated in Figure 1. The new framework incorporates a mapping network called Spectral-Spatial Transformer, which is designed to acquire knowledge of both spectral and spatial self-similarity and global correlation. This is achieved through the utilization of efficient spectral self-attention and spatial self-attention, respectively. We plug a spectral-spatial transformer (SST) into the reversible prior-based framework to establish a novel HSI reconstruction method, a spectral-spatial transformer network based on reversible prior (SST-ReversibleNet). Finally, based on the unique design of the new framework, we propose a new reversible loss. By implementing the approaches mentioned above, we have developed a set of very efficient SST-ReversibleNet families that outperform the current leading methods by a wide margin. The core of the efforts can be concisely summarized as follows:

1. We propose a new framework that bridges the gap between E2E and DU, allowing E2E methods to have the iterative capabilities and interpretability of DU. In addition, we design a new reversible loss based on the new framework.
2. We present a SST module that can balance the parameters and reconstruction accuracy without deepening the depth of the module.
3. Our SST-ReversibleNet outperforms SOTA methods by an average of 5.54db higher peak signal-to-nose ratio (PSNR) on simulated datasets and requires less than 1/3 the number of parameters and less than 1/2 of GFLOPs for the same reconstruction quality. SST-ReversibleNet also produces more aesthetically pleasing outcomes when applied to HSI reconstruction in the real world.

This article continues as follows. In Section 2, we discuss similar works. CASSI systems are introduced in Section 3, while SST-ReversibleNet design and loss are described in Section 4. We summarize the experimental setting and present and analyze the results in Section 5 and conclude in Section 6.

RELATED WORKS

Methods of HSI Reconstruction

E2E Method

The E2E method works by directly finding strong mapping relationships between measurements and spectral cubes, so the network structure is concise and diverse. E2E can be divided into convolutional neural networks (CNN)-based and transformer-based networks (Wang et al., 2015; Yang et al., 2021). Following external learning, Zhang et al. (2019) developed a CNN model using coded image internal learning. Miao et al. (2019) propose λ-net, a network speeding up the reconstruction by adding conditional generative adversarial networks (cGANs) to the U-net. Cheng et al. (2022) trained recurrent neural networks (RNNs) combining CNN and self-attention to improve the reconstruction quality, but the RNN-based framework model is large and time consuming. Although Both CNN-based and transformer-based E2E methods have the advantage of fast inference, both ignore the way CASSI systems work and lack theoretical interpretability and flexibility.

DU Network

The DU uses multi-stage network iterations to map measurements down a gradient into the HSI cube. DUs are generated from convex optimization algorithms such as alternating direction method of multipliers (ADMM), proximal gradient descent (PGD), and half quadratic splitting (HQS). In order to solve the proximal mapping problem associated with the sparsity-inducing regularizer, Zhang et al. (2018) used the iterative shrinkage-thresholding algorithm (ISTA) as inspiration and converted it into a deep network structure. In their study, Juan et al. (2021) introduce the linearized alternating direction method of multipliers network (LADMM-Net), a novel approach that incorporates the alternating direction method of multipliers (ADMM) into a processing layer. This technique allows for the estimation of image details by linearizing each iteration of the ADMM. Typically, these methods break down the objective function into a term that measures how well the data fits and a term that promotes decoupling. This results in iterative schemes where solutions to a subproblem related to the data and a subproblem related to prior knowledge are alternated. However, the optimization-based approach has some conditional constraints in the solution process, and as in the HQS expansion framework (Cai et al., 2022), the 2-stage iterative process can be described as shown in Equations 1 and 2.

$$x_{k+1} = \left(\Phi^{\mathrm{T}}\Phi + \mu \mathrm{I}\right)^{-1}\left(\Phi^{\mathrm{T}} y + \mu z_{k}\right) \tag{1}$$

$$z_{k+1} = arg\min_{z} \frac{1}{2\left(\sqrt{\frac{\tau_{k+1}}{\mu_{k+1}}}\right)^{2}} z - x_{k+1}{}^{2} + R\left(z\right) \tag{2}$$

where I is a matrix of identities. Φ is a matrix. X_{k+1} and z_{k+1} are two subproblems of (k+1)-stage. μ, μ_{k+1} are hyperparameters. R(.) is a mapping function. It is clear from the formula that the optimization formula is valid on the premise that $\left(\Phi^{\mathrm{T}}\Phi + \mu \mathrm{I}\right)^{-1}$ is invertible. In addition, operations such as transpose multiplication $\Phi^{\mathrm{T}}\Phi$ and $\Phi^{\mathrm{T}} y$ are involved in the operation.

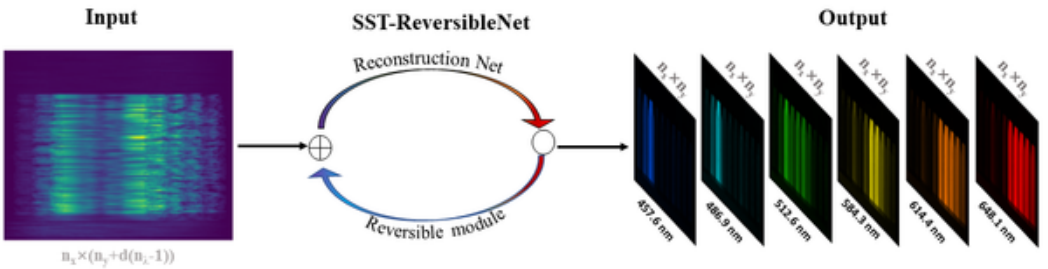
3D Cube Feature Extraction Module

Both E2E and DU require feature extraction in the measurement space. While many earlier efforts focused on utilizing CNN to extract local spatial information, these models suffer from shortcomings when it comes to representing non-local self-similarity and capturing long-range spatial interdependence. To remedy CNN's flaws, a new contender named Transformer has just emerged. To measure the reliance and similarity between different spectrums of HSI, Cai et al. (2022) suggested spectral-wise multi-head self-attention (S-MSA). However, S-MSA only extracts global features in spectral space while ignoring spatial information. The spatial-attention module in the spatial-spectral attention system suggested by Meng et al. (2020) learns 1D spatial features in the x- and y-axes before merging them with the spectral-attention feature. These methods either transform the dimensionality of the feature to reduce the parameters and operations, or ignore the interaction between spectral and spatial features. Insufficient feature learning results in less efficient network learning as the network depth increases.

MODEL OF CASSI SYSTEM

The model of our CASSI system is shown in Figure 1.

Figure 1. Schematic Diagram of Reversible Optical Path. According to the Principle of Reversible Optical Path, Our Network Also Includes Two Stages: Forward and Reverse



Imaging Process of CASSI

The 3D hyperspectral cube is first scattered by a dispersive prism in the CASSI system after being modified by a coded mask, as shown in Figure 2.

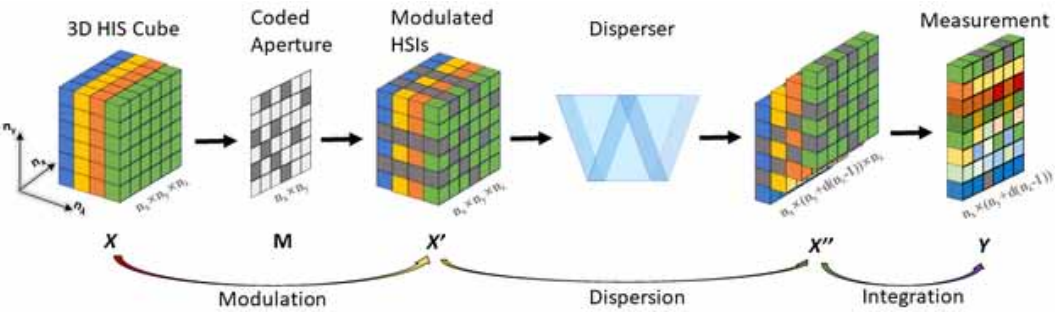
In mathematical terms, we can think of a 3D HSIs cube as $X \in \mathbb{R}^{n_x \times n_y \times c}$, where c, n_x and n_y stand for the number of wavelengths, diameter, and height of the HSIs, respectively. M, which is a pre-defined mask, is a member of the set $\mathbb{R}^{n_x \times n_y}$. The spectral image is modulated for each wavelength $m = 1,2 \cdots c$, and it can be expressed as shown in Equation 3.

$$X'(:,:,m) = X(:,:,m) \odot M \tag{3}$$

where $X' \in \mathbb{R}^{n_x \times n_y \times c}$ is spectral data cube, $\odot$ is the process of element-wise multiplying. Afterwards, the dispersive prism measures, X' is slanted and sheared along the y-axis. $X'' \in \mathbb{R}^{n_x \times \left(n_y + d(c-1) \times c\right)}$ is dispersed HSIs cube, and d indicates the shifting step. We assume λ_c is the reference wavelength, which means that $X''(:,:,m)$ is unbroken along its y-axis. Then, we have Equation 4.

$$X''(x,y,m) = X'(x, y + d_m, m) \tag{4}$$

Figure 2. Imaging Process of CASSI



where (x, y) represents coordinates of a point on the 3D HSI, d_m is the spatial shift of the channel on X''. Finally, the collected 2D compressed observation $Y \in \mathbb{R}^{n_x \times \left(n_y + d\left(c-1\right)\right)}$ can be derived as shown in Equation 5.

$$Y = \sum_{m=1}^{c} X''\left(:,:,m\right) + G \tag{5}$$

where $G \in \mathbb{R}^{n_x \times \left(n_y + d\left(c-1\right)\right)}$ is the measurement-related random noise produced by the photon sensing detector.

PROPOSED METHOD

Overall Architecture Based on Reversible Prior

Previous E2E methods look for violent mapping relations to obtain the solution to Equation 5 in a single pass, which means that they only have the upper half of the process (Reconstruction Net) of Figure 1. The single irreversible process also means that the end-to-end approach cannot fine-tune the intermediate inference results and learned features, leading to a partial degradation of the model performance.

In other fields, such as 3D reconstruction, relationships in the study of three-dimensional geometry often project the reconstruction results back into the measurement space based on the principle of reversibility of the optical path. The reconstruction results are updated using geometric distance constraints for optimization purposes. Similarly, HSI measurements also satisfy traditional physical optics. And the HSI 3D cube is easily projected back to the 2D measurement space. This process acts as an inverse process of reconstruction, and the quality of reconstruction can be improved by multiple iterations. Therefore, the primary distinction between our network and the E2E and the DU is the construction of residuals from measured and reprojected data, as well as the fine-tuning of the gap between the last learnt data and the true value based on the residuals. Figure 3 depicts the general architecture of SST-ReversibleNet, which is divided into a reversible module and a reconstruction subnet, as indicated below.

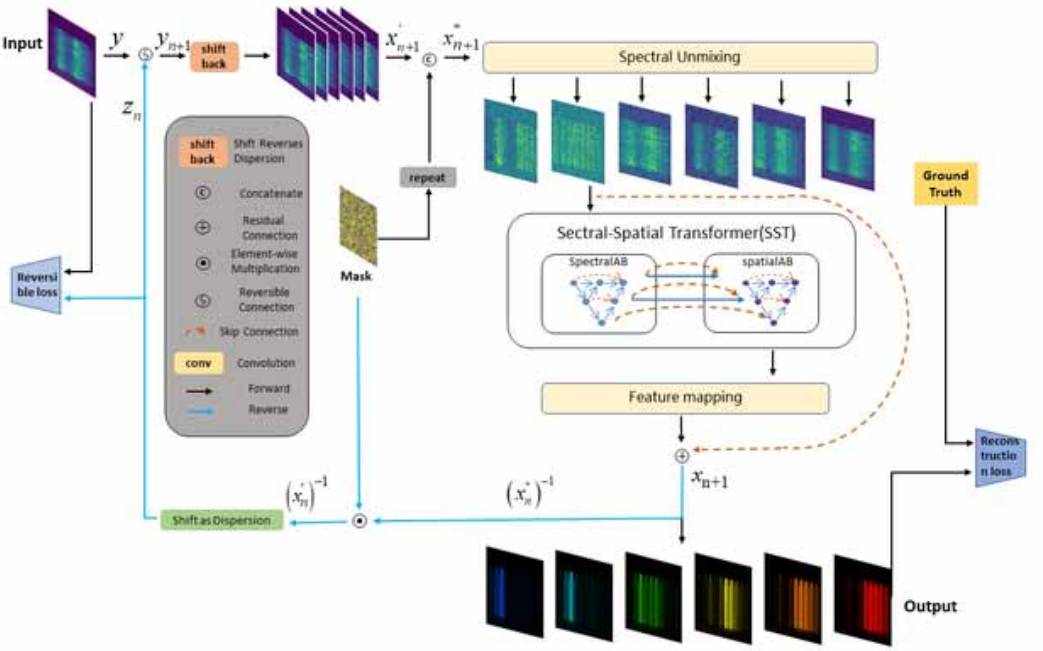
The physical meaning of the reversible module is that the reconstruction network takes the spectral image x_n obtained from the n-th reconstruction and compresses it into the measurement space z_n following the left-to-right process of Figure 2, expressed as shown in Equation 6.

$$z_n = G\left(x_n\right) \tag{6}$$

where $\mathcal{G}$ is the mapping of the spectral 3D cube to the 2D measurement, z_n is the output of the n-stage inverse process, x_n is the reconstruction result of the n-stage. Thereafter, based on the results of the reversible process, the residuals $\left(y - z_n\right)$ of the true measurements and the reprojected data can be constructed. Thus, the forward reconstruction process can be expressed as an iterative process, as shown in Equation 7.

$$x_{n+1} = \mathcal{F}_{n+1}\left(y - z_n\right) + x_n \tag{7}$$

Figure 3. Diagram of the Framework Structure Based on the Reversible Prior



where y is the actual measurement from the charge-coupled device (CCD) camera, $\mathcal{F}$ is the mapping from the input to the spectral 3D cube, and x_{n+1} is the reconstruction result of the $(n+1)$-stage.

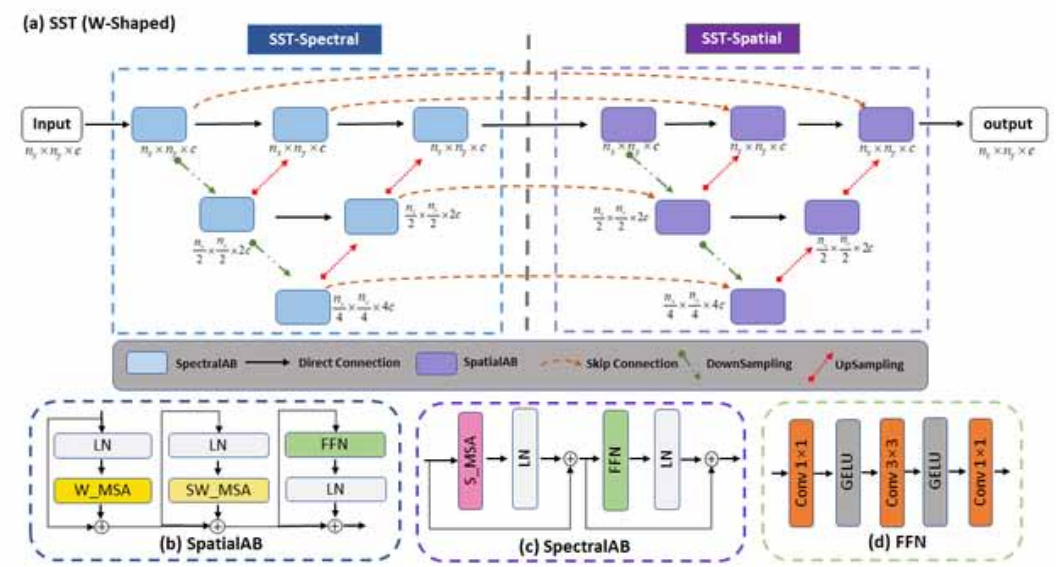
Figure 3 shows a diagram of our framework structure based on the reversable prior. The upper half is the forward process, reconstruction of the subnet including unmixing block, SST and mapping block. The blue line in the lower half indicates the inverse process, the reversible module. The reconstruction subnet corresponds to the inverse of the CASSI optical path and the reversible module corresponds to the forward direction of the CASSI optical path, with both directions allowing the SST to form a closed-loop iterative capability. A reversible loss is proposed on the inverse of the network.

As shown in Figure 4, we have created four SST-ReversibleNet models (SST), each with a different number of stages in the output (x_n). These models are named SST-S (n=1), SST-M (n=2), SST-L (n=4), and SST-LPlus (n=9). The models vary in their parameter sizes and computing costs, ranging from tiny to extremely big. In SST-S, we use reversible prior between SpatialAB and SpectralAB, while in other networks, we only use reversible prior between SpatialAB and SpectralAB modules.

Reversible Module

The inverse procedure is executed by utilizing the output of the spectral reconstruction network at the n-th stage to derive the anticipated value of the spectral cube x_n. According to Equations 3 and 4, the predicted value $\left(x_n''\right)^{-1} = x_n$ can be projected back into the measurement space after mask encoding, dispersion, and blending. As shown in the blue line in Figure 3, $\left(x_n''\right)^{-1}$, $\left(x_n'\right)^{-1}$ correspond to the inverse predicted values of x_n'' and x_n' in the forward process, respectively, and the inverse process is described as shown in Equations 8 and 9.

Figure 4. Diagram of Spectral-Spatial Transformer (Note: (a) SST adopts a W-shaped structure; (b) SpatialAB consists of a window multihead-self-attention (MSA), a shifted-window-MSA, a feature forward network (FFN), and three layer normalization; (c) SpectralAB consists of a spectral-MSA, a FNN, and two layer normalization; (d) components of FFN)



$$\left(x_n'\right)^{-1}\left(x,y,:\right)=\left(x_n''\right)^{-1}\left(x,y,:\right)\odot M \tag{8}$$

$$z_n\left(x,y\right)=\sum_{m=1}^{c}\left(x_n'\right)^{-1}\left(x,y+d_m,m\right) \tag{9}$$

After obtaining z_n, our reconstruction network reconstructs a residual $y-z_n$ of the measurements y and feeds it again into the reconstruction network to relearn the mapping of $y-z_n$ to $x_{Truth}-x_n$. The input y_{n+1} to the reconstruction subnet is represented as shown in Equation 10.

$$y_{n+1}=\begin{cases}y & \text{if } n=0\\ y-z_n & \text{otherwise}\end{cases} \tag{10}$$

Reconstruction Subnet

The function of the reconstruction subnet is to establish a mapping between the different inputs and outputs, mainly consisting of a spectral-spatial transformer. In addition, considering that the measurement space is a compressed 2D space and that there is aliasing of data from different channels, we introduced a module for unaliasing and feature mapping before the mapping network input and after the output.

Given the measurement after initialization $x_{n+1} \in \mathbb{R}^{n_x \times \left(n_y + d_m\right)}$, firstly, we divide the aliased data into input signals with different wavebands according to the backlight propagation. The initialized signal $X_{n+1}^{'} \in \mathbb{R}^{n_x \times n_y \times c}$ can be obtained as shown in Equation 11.

$$x_{n+1}^{'}\left(x,y,m\right) = y_{n+1}\left(x, d_m : d_m + h\right), m = 1,2,\ldots,c \tag{11}$$

where x, y are the spatial coordinates of a point on the 3D cube, d_m is the offset of the spectral image on the m-channel, and h is the height of the 3D cube.

Spectral Unmixing

Afterwards, we utilize the mask's prior information to direct the input for unmixing. This is done by passing the shifted y, which is combined with the mask M, then through convolution with $conv1*1$ kernel to back to input signal $X_n^{''} \in \mathbb{R}^{n_x \times n_y \times 2c} \rightarrow X_n \in \mathbb{R}^{n_x \times n_y \times c}$. The spectral unmixing is realized through the convolution layer with varying sizes to solve the aliasing problem under different receptive fields ($conv3*3$, $conv5*5$, $conv7*7$).

SST

The suggested reconstruction subnet is designed to generate high-quality HSIs from the spectral images obtained after unmixing. We use a W-shaped spectral-spatial transformer module (Figure 4), which is composed of encoding and decoding of spectral features and the encoding and decoding between spatial channels.

The SST-Spectral and SST-Spatial models employ an encoder-decoder architecture similar to an unet. These models are coupled through a sequence of nested dense SpatialAB and SpectralAB blocks, respectively. This architecture is specifically designed to merge the discontinuities between the feature maps of the encoder and decoder that correspond to the same feature but in different dimensions.

Spectral-Spatial-Wise Multihead Self-Attention (MSA)

The Cube of the spectrum has a spatial correlation in the spatial dimension, which is related to the target's properties and the surface's reflectivity. While, in the spectral dimension, the continuity of the spectrum determines that the adjacent spectra are similar, and the farther the spectral distance is, the more ranges are complementary. And since $W = H >> M$, modelling spatial-wise correlations will be more cost-effective than capturing spectral-wise interactions. However, when the model reaches a certain scale, a single method cannot continue to mine the information of spectral Cube. Therefore, we consider learning the global correlation of spatial features based on the spectral attention map to improve the learning ability.

SpectralAB is consistent with multihead self-attention block (MSAB) in mask-guided spectral-wise transformer (MST) (Cai et al., 2022). In order to compute self-attention along the spectral channel, SpectralAB's objective is to consider each spectral feature map as an individual token. And the formula for each $head_j^{Spectral}$ and SpectralAB is shown in Equations 12 and 13.

$$head_j^{Spectral} = Softmax\left(\sigma_j Q_j K_j^T\right) V_j \tag{12}$$

$$SpectralAB\left(X\right) = \underset{j=1}{\overset{N}{concat}}\left(head_j\right) W + f\left(V\right) \tag{13}$$

SpatialAB makes improvements based on swim-transformer (Liu et al. 2021). We remove Avgpooling blocks, and add a feature forward network (FFN) module and a LayerNorm layer (Figure 4). The objective of SpatialAB is to compute self-attention along the spatial dimension by treating each local spatial feature map as a token. Input $x_n \in \mathbb{R}^{n_x \times n_y \times c}$ is changed into tokens $x \in \mathbb{R}^{\frac{n_x}{s}\cdot\frac{n_y}{s} \times s \cdot s \times c}$, s represents the window-size (set to 8 by default) of each window. The following module adopts a windowing configuration that is moved from that of the preceding layer, by moving the windows by $\left(\frac{s}{2}, \frac{s}{2}\right)$ pixels from the regularly partitioned windows. The formula for spatial Attention is shown in Equation 14.

$$Attention = SoftMax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{14}$$

Loss Function

Our network has reversible module and reconstruction subnet. Therefore, we design the reversible loss according to this structure. The previous work only establishes the L2 loss of the reconstruction results with the true result. While in the reversible light path, since the reconstruction result is a biased estimation of the true value, by constraining the data in the measurement space, the constraint ability of the network can be improved. The outputting loss is calculated as the L2 loss of $x_{out} - x_{truth}$. The reversible loss calculation x_{out} is mapped back to the CCD under the nature of the reversible optical path to obtain the L2 loss of the $\mathcal{G}\left(x_{out}\right)$ value to the actual measurement y. We defined the loss function as shown in Equation 15.

$$L = x_{out} - x_{truth2}^{\ 2} + \xi \cdot \mathcal{G}\left(x_{out}\right) - y_2^2 \tag{15}$$

where x_{out} represents the final predicted network values. $\mathcal{G}$ means the process of mask coding and spreading out the expected values. Meanwhile, y represents the observation. ξ is penalty coefficient, has a default value of 0.2.

EXPERIMENTS AND ANALYSIS

Experiment Setup

In our implementation, the spectral channels are represented by a total of 28 distinct wavelengths ranging from 450 nm to 650 nm. We conduct tests on both simulated and real HSI datasets

Two datasets of simulated hyperspectral images, CAVE (Yasuma et al., 2010) and KAIST (Choi et al., 2017), were utilized. The CAVE dataset has 32 hyperspectral pictures with a 512×512 spatial resolutions. The KAIST collection is made up of 30 hyperspectral images with a size of 2704×3376 pixels. Following the schedule set by spectral-spatial transformer (TSA-Net, Meng et al., 2020), we use CAVE as our training set. Ten scenes from KAIST will be tested. For this, we rely on the real-world HSIs dataset amassed by the TSA-Net CASSI system. We measure the HSI reconstruction performance using PSNR and structural similarity (SSIM), which were proposed by Wang et al. (2004).

We use Pytorch to implement SST-ReversibleNet. Our SST-S, SST-M, and SST-L are trained on a single RTX 3090 graphics processing unit (GPU), while SST-LPlus is trained on two RTX 3090

GPUs. We use the Adam optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) for 300 epochs. The learning rate is initially set to 4×10^{-4} and half every 50 epochs throughout training. The reconstruction quality is evaluated using the PSNR and SSIM measures.

Results on CAVE

We compare the Params, GFLOPs, PSNR, and SSIM of our SST-ReversibleNet with several SOTA HSI reconstruction algorithms, including λ -net, alternating direction method of multipliers (ADMM-Net, (Ma et al., 2019), TSA-Net, deep image priors pre-trained hyperspectral images (DIP-HIS, Meng et al., 2021), deep gaussian scale mixture prior (DGSMP, Huang et al., 2021), bidirectional recurrent neural networks with adversarial training (BIRNAT), MST series, coarse-to-fine sparse transformer (CST) series (Cai et al., 2022), high-resolution dual-domain learning (HD-Net), and degradation-aware unfolding half-shuffle transformer (DAUHST) series. Table 1 lists the parameters and GFLOPs examined with the identical settings (test size = 256×256), as well as the PSNR and SSIM findings of different approaches on all simulated datasets.

Here, you need to briefly introduce Table 1 in the text (callout) before including it in the paper.

The most effective of our models, SST-LPlus, produces extremely impressive outcomes: 97.4% in SSIM and 39.16 dB in PSNR, which is more than 3 dB than the best PSNR of the SOTA published models, and the SSIM is more than 1.5%. SST-LPlus significantly outperforms DAUHST-9stg, BIRNAT, MST++, MST-L, HD-Net, TSA-Net and λ -Net of PSNR by 0.80, 1.58, 3.82, 4.07, 4.19, 6.86 and 7.39 dB, and 0.7%, 1.4%, 2.1%, 2.4%, 3.1%, 5.8% and 8.4% improvement of SSIM, indicating the efficacy of our approach.

Comparisons of our SST-LPlus algorithm and other SOTA algorithms on Scene 5 using four (out of twenty-eight) spectral channels are illustrated in Figure 5. Zoomed-in regions of white boxes are displayed in the upper-right corner of the HSIs; compared to previous SOTA techniques, reconstructed HSIs generated by SSTs have more distinct spatial characteristics and a more distinct texture across various spectral channels.

Furthermore, as shown in Figure 6, in positions A, B, and C, while all restoration techniques can better capture the qualitative trend of spectrum changes, SST spectral curves have superior spectral accuracy and perceptual quality.

Observably, our SST-ReversibleNet outperforms SOTA methods by a substantial margin, while requiring considerably less memory and computation power. Compared with other Transformer-based method CST-L and MST-L, our SST-S outperforms CST-L by 0.59 dB but only costs 35.3% (1.06/3.00) Params and 71.3% (19.83/27.81) GFLOPs, and SST-S exhibits a 1.62 dB improvement over MST-L while requiring only 1.06/3.66 parameters for 29.0% and 19.83/28.15 GFLOPs for 70.4%. Likewise, our SST-M outperforms DAUHST-5stg by 0.13 dB but only costs 61.3% (2.11/3.44) Params and 78.5% (35.03/44.61) GFLOPs and SST-M outperform CST-L-plus by 1.76 dB, but only costs 70.3% (2.11/3.00) Params and 87.4% (35.03/40.1) GFLOPs. More specifically, our SST-M acquires the equivalent SSIM (96.7%) of DAUHST-9stg (the best model at present), but only costs 34.3% (2.11/6.15) Params and 44.1% (35.03/79.5) GFLOPs. In addition, our SST-L and SST-LPlus outperform other competitors by very large margins. Figure 7 presents comparisons of various reconstruction algorithms based on PSNR, parameters, and GFLOPs.

Real Data Results

We employ five compressive measurements obtained from the operational spectral SCI system to assess the proposed technique and evaluate its performance on actual data. The CAVE datasets are utilized to train all of the approaches, employing a consistent real mask and introducing 11-bit shot noise, in order to guarantee equitable comparisons. See how the suggested SST-M stacks up against the current SOTA methods in Figure 8.

Table 1. Comparisons Between SSTs and SOTA Methods on 10 Simulation Scenes (S1-S10), Where Params, GFLOPS, PSNR and SSIM are Reported

Scene		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg	Params	GFLOPs
λ -net	PSNR	32.50	31.23	33.89	40.28	29.86	30.27	30.33	28.98	31.98	28.36	31.77	62.6	118.0
	SSIM	0.892	0.854	0.930	0.965	0.889	0.893	0.975	0.880	0.891	0.834	0.890		
ADMM-Net	PSNR	34.12	33.62	35.04	41.15	31.82	32.54	32.42	30.74	33.75	30.68	33.58	4.27	78.58
	SSIM	0.918	0.902	0.931	0.966	0.922	0.924	0.896	0.907	0.915	0.895	0.918		
TSA-Net	PSNR	32.95	31.69	33.01	41.24	30.12	31.89	30.75	29.89	31.61	29.9	32.30	44.2	135.2
	SSIM	0.913	0.884	0.932	0.975	0.911	0.929	0.895	0.912	0.920	0.890	0.916		
GAP-Net	PSNR	26.82	22.89	26.31	30.65	23.64	21.85	23.76	21.98	22.63	23.10	24.36	4.27	84.47
	SSIM	0.754	0.610	0.802	0.852	0.703	0.663	0.688	0.655	0.682	0.584	0.669		
DGSMP	PSNR	33.26	32.09	33.06	40.54	28.86	33.08	30.74	31.55	31.66	31.44	32.63	3.76	646.7
	SSIM	0.915	0.898	0.925	0.964	0.882	0.937	0.886	0.923	0.911	0.925	0.917		
DIP-HSI	PSNR	32.68	27.26	31.30	40.54	29.79	30.39	28.18	29.44	34.51	28.51	31.26	33.9	64.42
	SSIM	0.890	0.833	0.914	0.962	0.900	0.877	0.913	0.874	0.927	0.851	0.894		
BIRNAT	PSNR	36.79	37.89	40.61	46.94	35.42	35.30	36.58	33.96	39.47	32.80	37.58	4.35	2131
	SSIM	0.951	0.957	0.971	0.985	0.964	0.959	0.955	0.956	0.970	0.938	0.960		
HD-Net	PSNR	35.14	35.67	36.03	42.30	32.69	34.46	33.67	32.48	34.89	32.38	34.97	2.37	154.8
	SSIM	0.935	0.940	0.943	0.969	0.946	0.952	0.926	0.941	0.942	0.937	0.943		
MST-L	PSNR	35.29	35.48	36.72	42.68	32.55	34.67	33.53	32.50	34.98	32.45	35.09	3.66	28.15
	SSIM	0.945	0.944	0.956	0.980	0.947	0.957	0.929	0.953	0.948	0.945	0.950		
MST++	PSNR	35.53	35.68	35.99	42.78	32.71	35.14	34.24	33.30	35.13	32.86	35.34	1.33	19.64
	SSIM	0.946	0.946	0.954	0.977	0.949	0.959	0.938	0.957	0.951	0.948	0.953		
CST-L	PSNR	35.96	36.84	38.16	42.44	33.25	35.72	34.86	34.34	36.51	33.09	36.12	3.00	40.10
	SSIM	0.949	0.955	0.962	0.975	0.955	0.963	0.944	0.961	0.957	0.945	0.957		
DAUHST-2st	PSNR	35.93	36.70	37.96	44.38	34.13	35.43	34.78	33.65	37.42	33.07	36.34	1.40	18.44
	SSIM	0.943	0.946	0.959	0.978	0.954	0.957	0.940	0.950	0.955	0.941	0.952		
DAUHST-3st	PSNR	36.59	37.93	39.32	44.77	34.82	36.19	36.02	34.28	38.54	33.67	37.21	2.08	27.17
	SSIM	0.949	0.958	0.964	0.980	0.961	0.963	0.950	0.956	0.963	0.947	0.959		
DAUHST-5st	PSNR	36.92	38.52	40.51	45.09	35.33	36.56	36.28	34.74	38.71	34.27	37.75	3.44	44.61
	SSIM	0.955	0.962	0.967	0.980	0.964	0.965	0.958	0.959	0.963	0.952	0.962		
DAUHST-9st	PSNR	37.25	39.02	41.05	46.15	35.80	37.08	37.57	35.10	40.02	34.59	38.36	6.15	79.50
	SSIM	0.958	0.967	0.971	0.983	0.969	0.970	0.963	0.966	0.970	0.956	0.967		
SST-S	PSNR	36.55	37.27	38.49	44.50	34.30	36.18	35.35	34.03	37.21	33.19	36.71	1.06	19.83
	SSIM	0.953	0.955	0.964	0.984	0.960	0.966	0.948	0.959	0.955	0.949	0.959		
SST-M	PSNR	37.32	38.60	40.76	45.73	35.56	37.01	36.44	34.69	38.63	34.09	37.88	2.11	35.03
	SSIM	0.961	0.965	0.973	0.987	0.968	0.972	0.955	0.966	0.966	0.959	0.967		
SST-L	PSNR	37.77	39.56	41.87	46.72	36.50	37.54	37.28	35.11	39.80	34.83	38.70	4.25	72.52
	SSIM	0.966	0.972	0.977	0.990	0.972	0.975	0.962	0.968	0.972	0.963	0.972		
SST-LPlus	PSNR	38.24	40.05	42.45	47.87	37.02	37.59	37.20	35.42	40.54	35.25	39.16	9.64	167.5
	SSIM	0.968	0.974	0.979	0.992	0.975	0.975	0.960	0.971	0.975	0.962	0.974		

Figure 5. Visual Comparisons of Our SST-ReversibleNet and Other SOTA Methods of Scene 5 with 4 Out of 28 Spectral Channels on the KAIST Dataset

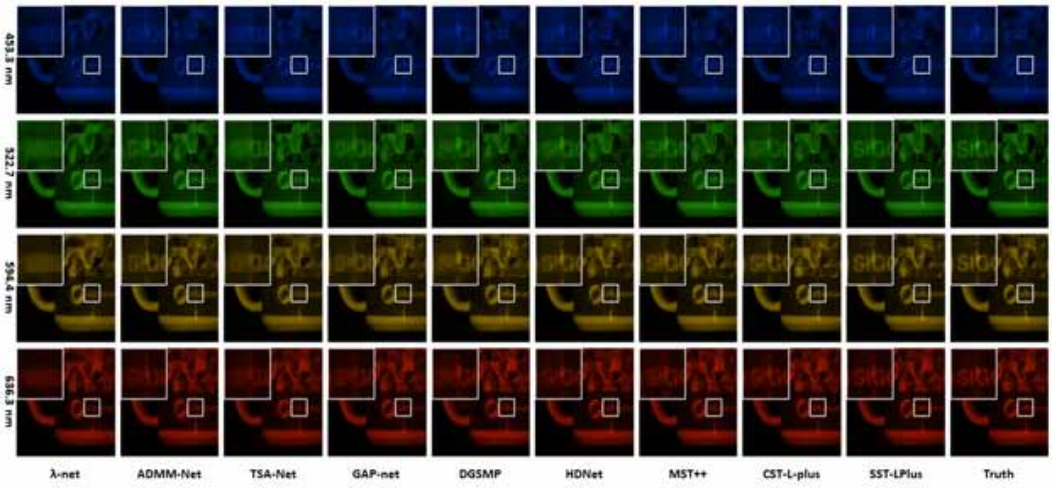


Figure 6. Spectral Curves of the SOTA Methods in Figure 5 on the Randomly Selected Regions A, B and C

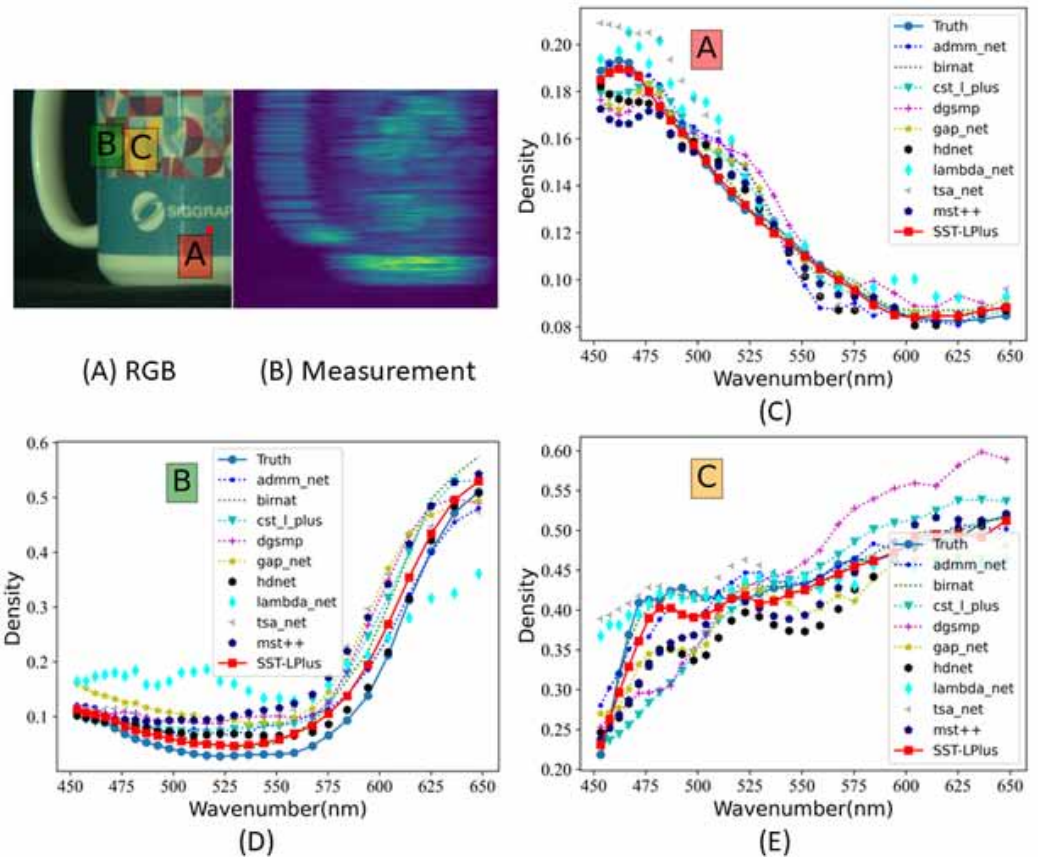
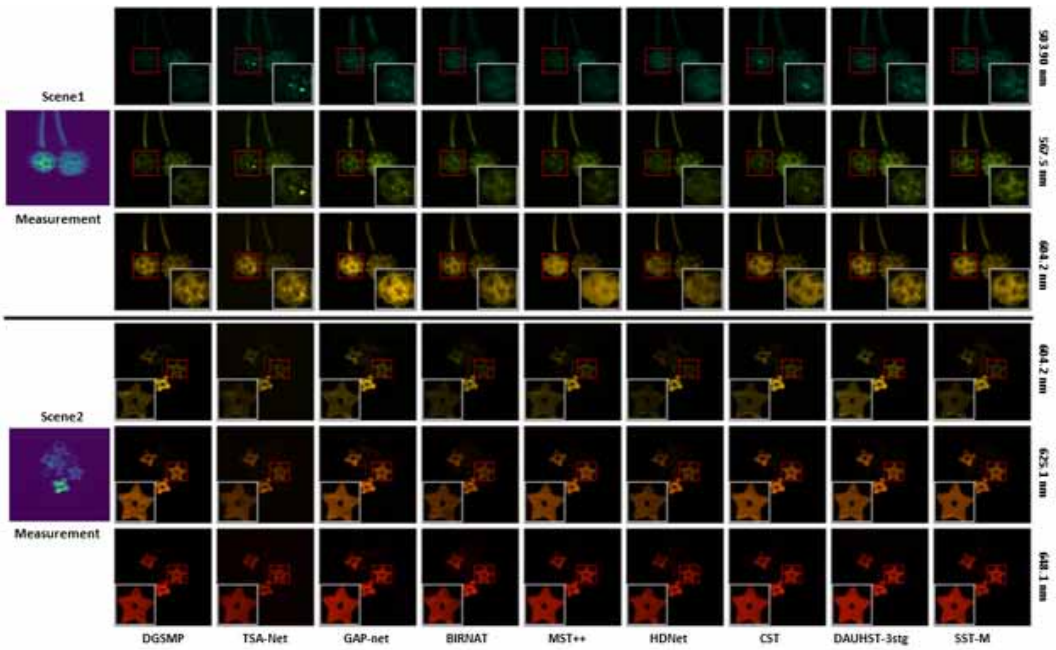


Figure 7. PSNR-Params-GFLOPs Comparisons of Our SST-ReversibleNet and SOTA HSI Reconstruction Methods. Note. The Vertical Axis is PSNR (dB), the Horizontal Axis is GFLOPs (Computational Cost), and the Circle Radius is Params (Memory Cost).

Figure 8. Real HSI Reconstruction Comparison of Two Scenes. 6 Out of 28 Spectra are Randomly Selected



These methods include DGSMP, TSA-Net, GAP-Net (Meng et al., 2020), BIRNAT, MST++, HDNet, CST, and DAUHST. When compared to earlier algorithms, our SST-S achieves better results in restoring structural details at high frequencies and suppressing actual noise. More texture and detail, particularly around blossom margins, can be restored in Scene 2 using the suggested method.

ABLATION STUDY AND ANALYSIS

An ablation study is undertaken on the CAVE and KAIST datasets to assess the contribution of different components in the proposed SST-ReversibleNet (SST). Our primary emphasis is on the four components, i.e. whether to use the reversible prior, whether to use the reversible loss, the structural shape of the feature extraction network, and the effect of the combination of spectral self-attentive blocks (SpectralAB) and spatial self-attentive blocks (SpatialAB) on the model. Tables 2–5 2 to Table 5 show the results of the comparison between PSNR and SSIM at different settings.

In Table 5, we build two networks A and B with similar number of parameters and GFLOPs as SST-S. In this case, A uses only the self-attention of the spectral channels and B only looks for correlations in the spatial dimension.

Table 2. Ablation of Use Prior

Baseline	Use Prior	Params(M)	GFLOPs(G)	PSNR	SSIM
SST-S	×	1.03	17.98	33.52	92.2%
SST-S	√	1.06	19.83	36.71	95.9%

Table 3. Ablation of Use Reversible Loss

Baseline	Use reprojection loss	Params(M)	GFLOPs(G)	PSNR	SSIM
SST-S	×	1.06	19.83	36.66	95.7%
SST-S	√	1.06	19.83	36.71	95.9%

Table 4. Ablation of SST Structures

Baseline	shape of spectral spatial Transformer	Params (M)	GFLOPs (G)	PSNR	SSIM
SST-S	Unet-like	1.01	16.99	35.42	95.3%
SST-S	Unet++-like	1.06	19.83	36.71	95.9%
SST-M	W-shaped	2.11	35.03	37.86	96.6%

Table 5. Ablation of Use SpectralAB and SpatialAB

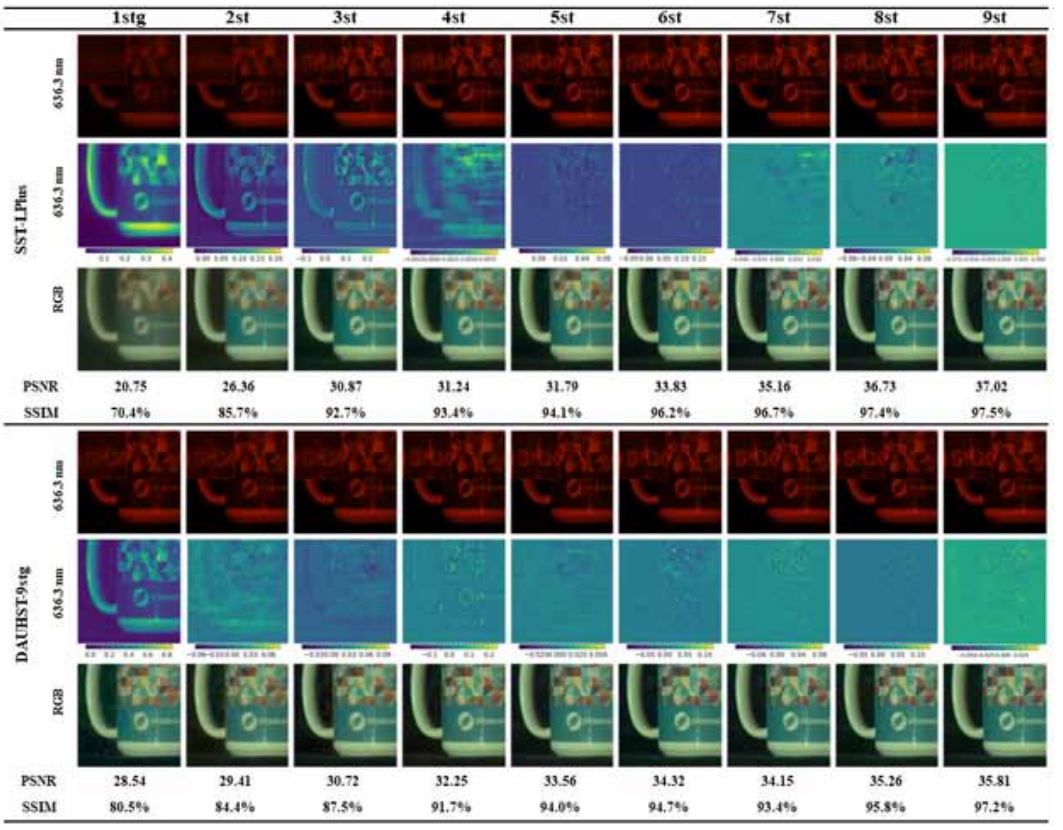
Baseline	Use SpectralAB	Use SpatialAB	Params (M)	GFLOPs (G)	PSNR	SSIM
A	√	×	0.97	19.61	34.83	93.2%
B	×	√	1.13	21.74	35.77	94.7%
SST-S	√	√	1.06	19.83	36.71	95.9%

The results show that 1) the impact of reversible prior on the model is crucial, comparing two networks with similar number of parameters for deepening the network depth and using reversible prior in Table 2, utilizing a reversible prior can significantly enhance the model's reconstruction capability, with PSNR and SSIM improving by 3.19 dB and 3.7%, respectively. 2) The reversible loss can be a constraint on the model, and without changing the number of parameters or operations, the PSNR and SSIM are able to improve by 0.05db and 0.2% respectively. 3) The shape of the network structure also has a more obvious improvement on the reconstruction effect, the W-shaped structure can improve the PSNR and SSIM by 2.44 db/1.3% and 1.15 db/0.7%, respectively, compared to Unet-like and Unet++like shape. 4) The spectral-space transformer is a huge advantage over the spectral-transformer and spatial-transformer, especially the results for model B vs. SST-s. Model B has 7% higher number of parameters and 10% higher GFLOPs than SST-S, but the reconstruction results are 0.94 dB less than SST-S.

Difference With DU

To explore the differences between our iterative method and the DU, we compared the iterative process of SST-LPlus and DAUHST-9st on the simulated dataset. A randomly selected scene is visualized in both spectral channels and RGB changes. In addition, below the 636.3 nm visualization image, we have extracted the changes at different stages of each feature learning. Below the image we list the PSNR and SSIM changes from 1stg to 9stg, as shown in Figure 9.

Figure 9. Comparison Graph of E2E Iterative Method (SST-LPlus) and DU (DAUHST-9stg) in Different Stages (Note. Comparison of the differences on the 636.3 nm spectral image, the 636.3 nm spectrally learned feature map, the RGB image, the PSNR and the SSIM from 1 stage to 9 stage)



Visualizing the analysis of the results, we believe that the reversible framework benefits from the learning of residuals to effectively improve the learning of features, with DAUHST learning a large number of global features at 1stg and fine-tuning from 2stg onwards.

Although SST-ReversibleNet only starts fine-tuning at 4stg, our SST-ReversibleNet learns more global features from 1stg to 4stg in the early stage and achieves results beyond DAUHST at 7stg to 9stg. We therefore believe that the feature learning capabilities of SST-ReversibleNet and DU are not the same.

Analysis

With the above experiments, we analyze the reasons why SST-ReversibleNet achieves good results and has a smaller number of parameters. The most important reason is the architecture of SST-ReversibleNet ‘s reversible prior. Its architecture of reversible prior has more iterations compared to the E2E method, which can learn detailed features that were not learned in the previous stages from multiple residual learning. And compared with the DU method, which has two or more iterative steps in a single stage, SST-ReversibleNet has only one iterative step in a single stage. Thus, SST-ReversibleNet has fewer parameters with similar reconstruction accuracy as DU. In addition, the designed structure of W-shape can learn features in both spatial and spectral dimensions, which greatly improves the performance of the network.

Although it is of note that the SST-ReversibleNet residuals are learned in such a way that they cannot be targeted to estimate the noise of the measurement. Considering that denoising of hyperspectral images can be a separate task, the network of SST-ReversibleNet can be designed as a two-stage network, and by introducing a denoising module (Sun et al., 2023; Li et al., 2023) at the input of the network, the noise as well as the reconstruction task can be solved simultaneously.

CONCLUSION

Inspired by the reversible light path, this paper proposes a novel SST-ReversibleNet for CASSI. The new framework significantly improves the reconstruction metrics and can be used for other algorithms. We use a W-shaped spectral-spatial transformer module to improve spatial and spectral feature extraction. In addition, we design a reversible loss. Using these innovative methods, we create a collection of exceptionally effective SST-ReversibleNet models. Quantitative experiments demonstrate that our approach surpasses state-of-the-art algorithms by at least 0.8 dB and only uses 34.3% Params and 44.1% GFLOPs. Furthermore, we believe that the proposed a priori reversible optical path can be used equally well for tasks such as super-resolution, video compression, de-fogging, and denoising. For all tasks with explicit optical significance, we consider the decomposition of the task into forward and backward processes to improve network performance by modeling the physical processes of the system, either explicitly or implicitly. Since SST-ReversibleNet and DU have a similar iterative structure, SST-ReversibleNet will face the problem of increasing parameters and arithmetic power as the number of iterations increases, and our future work will consider sharing weights or improving the reconstruction effect of a single stage to alleviate this problem.

COMPETING INTERESTS

We declare that none of the work contained in this manuscript has been published in any other language or is currently under consideration at any other journal, and there are no conflicts of interest to declare.

FUNDING

The authors would like to thank editor and anonymous reviewers who gave valuable suggestion that has helped to improve the quality of the paper. This work was supported in part by the National

Natural Science Foundation of China under Grants 32171911; by the National Key R\&D Program of China under Grants 2021YFD2000503; by the Special Project on Basic Research of Frontier Leading Technology of Jiangsu Province of China under Grant BK20192004C.

PROCESS DATES

Received: 3/12/2024, Revision: 4/15/2024, Accepted: 4/15/2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Zheng Liu (China, liuzheng9012@163.com)

REFERENCES

Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lucic, M., & Schmid, C. (2021). Vivit: A video vision transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6836-6846.

Cai, Y., Lin, J., Hu, X., Wang, H., Yuan, X., Zhang, Y., Timofte, R., & Gool, L. V. (2022). Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17502-17511. doi:10.1109/CVPR52688.2022.01698

Cai, Y., Lin, J., Lin, Z., Wang, H., Zhang, Y., Pfister, H., Timofte, R., & Gool, L. V. (2022). Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 745-755. doi:10.1109/CVPRW56347.2022.00090

Cai, Y., Lin, J., Wang, H., Yuan, X., Ding, H., Zhang, Y., Timofte, R., & Gool, L. V. (2022). Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. *Advances in Neural Information Processing Systems*, *35*, 37749–37761.

Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Carin, L., & Brady, D. J. (2016). Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, *33*(5), 95–108. doi:10.1109/MSP.2016.2582378

Cheng, Z., Chen, B., Lu, R., Wang, Z., Zhang, H., Meng, Z., & Yuan, X. (2022). Recurrent neural networks for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(2), 2264–2281. doi:10.1109/TPAMI.2022.3161934 PMID:35324434

Choi, I., Jeon, D. S., Nam, G., Gutierrez, D., & Kim, M. H. (2017). High-quality hyperspectral reconstruction using a spectral prior. [TOG]. *ACM Transactions on Graphics*, *36*(6), 1–13. doi:10.1145/3130800.3130810

Chopra, M., Singh, S. K., Sharma, A., & Gill, S. S. (2022). A comparative study of generative adversarial networks for text-to-image synthesis. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–12. doi:10.4018/IJSSCI.300364

Chu, J., Zhao, X., Song, D., Li, W., Zhang, S., Li, X., & Liu, A. (2022). Improved semantic representation learning by multiple clustering for image-based 3D model retrieval. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–20. doi:10.4018/IJSWIS.297033

De Angelis, S., Ammannito, E., di Iorio, T., De Sanctis, M. C., Manzari, P., Liberati, F., Tarchi, F., Dami, M., Olivieri, M., Pompei, C., & Mugnuolo, R. (2015). The spectral imaging facility: Setup characterization. *The Review of Scientific Instruments*, *86*(9), 093101. doi:10.1063/1.4929433 PMID:26429423

Deng, S., Deng, L., Wu, X., Ran, R., Hong, D., & Vivone, G. (2023). PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–15. doi:10.1109/TGRS.2023.3244750

Du, H., Tong, X., Cao, X., & Lin, S. (2009). A prism-based system for multispectral video acquisition. *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 175-182.

Ghoneim, A., Ghulam, M., Amin, S. U., & Gupta, B. B. (2018). Medical image forgery detection for smart healthcare. *IEEE Communications Magazine*, *56*(4), 33–37. doi:10.1109/MCOM.2018.1700817

Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A. J., & Chanussot, J. (2021). SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–15. doi:10.1109/TGRS.2022.3172371

Hu, B., Gaurav, A., Choi, C., & Almomani, A. (2022). Evaluation and comparative analysis of semantic web-based strategies for enhancing educational system development. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–14. doi:10.4018/IJSWIS.302895

Hu, X., Cai, Y., Lin, J., Wang, H., Yuan, X., Zhang, Y., Timofte, R., & Gool, L. V. (2022). Hdnet: High-resolution dual-domain learning for spectral compressive imaging. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17542-17551. doi:10.1109/CVPR52688.2022.01702

Huang, T., Dong, W., Yuan, X., Wu, J., & Shi, G. (2021). Deep gaussian scale mixture prior for spectral compressive imaging. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 16216-16225. doi:10.1109/CVPR46437.2021.01595

Ishida, T., Kurihara, J., Viray, F. A., Namuco, S. B., Paringit, E. C., Perez, G. J., Takahashi, Y., & Marciano, J. J. Jr. (2018). A novel approach for vegetation classification using UAV-based hyperspectral imaging. *Computers and Electronics in Agriculture*, *144*, 80–85. doi:10.1016/j.compag.2017.11.027

Jia, S., Zhu, S., Wang, Z., Xu, M., Wang, W., & Guo, Y. (2023). Diffused convolutional neural network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–15. doi:10.1109/TGRS.2023.3250640

Kim, M. H., Rushmeier, H. E., Dorsey, J., Harvey, T. A., Prum, R. O., Kittle, D. S., & Brady, D. J. (2012). 3D imaging spectroscopy for measuring hyperspectral patterns on solid objects. [TOG]. *ACM Transactions on Graphics*, *31*(4), 1–11. doi:10.1145/2185520.2185534

Li, D., Deng, L., Gupta, B. B., Wang, H., & Choi, C. (2019). A novel CNN based security guaranteed image watermarking generation scenario for smart city applications. *Information Sciences*, *479*, 432–447. doi:10.1016/j.ins.2018.02.060

Li, J., Hu, Q., & Ai, M. (2019). RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Transactions on Image Processing*, *29*, 3296–3310. doi:10.1109/TIP.2019.2959244 PMID:31869789

Li, M., Fu, Y., & Zhang, Y. (2023). Spatial-spectral transformer for hyperspectral image denoising. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(1), 1368–1376. doi:10.1609/aaai.v37i1.25221

Li, Y., Shi, Y., Wang, K., Xi, B., Li, J., & Gamba, P. (2022). Target detection with unconstrained linear mixture model and hierarchical denoising autoencoder in hyperspectral imagery. *IEEE Transactions on Image Processing*, *31*, 1418–1432. doi:10.1109/TIP.2022.3141843 PMID:35038293

Liang, J., Zeng, H., & Zhang, L. (2022). Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5657-5666. doi:10.1109/CVPR52688.2022.00557

Lin, J., Cai, Y., Hu, X., Wang, H., Yuan, X., Zhang, Y., Timofte, R., & Gool, L. V. (2022). Coarse-to-fine sparse transformer for hyperspectral image reconstruction. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 686-704.

Liu, T., Liu, H., Li, Y., Chen, Z., Zhang, Z., & Liu, S. (2019). Flexible FTIR spectral imaging enhancement for industrial robot infrared vision sensing. *IEEE Transactions on Industrial Informatics*, *16*(1), 544–554. doi:10.1109/TII.2019.2934728

Liu, Y., Hu, J., Kang, X., Luo, J., & Fan, S. (2022). Interactformer: Interactive transformer and CNN for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–15. doi:10.1109/TGRS.2022.3183468

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF international conference on computer vision,* 10012-10022. doi:10.1109/ICCV48922.2021.00986

Ma, J., Liu, X., Shou, Z., & Yuan, X. (2019). Deep tensor admm-net for snapshot compressive imaging. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10223-10232. doi:10.1109/ICCV.2019.01032

Mandle, A. K., Sahu, S. P., & Gupta, G. P. (2022). CNN-based deep learning technique for the brain tumor identification and classification in MRI images. [IJSSCI]. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–20. doi:10.4018/IJSSCI.304438

Meng, Z., Ma, J., & Yuan, X. (2020). End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. *European conference on computer vision*. Cham: Springer International Publishing, 187-204. doi:10.1007/978-3-030-58592-1_12

Meng, Z., Yu, Z., Xu, K., & Yuan, X. (2021). Self-supervised neural networks for spectral snapshot compressive imaging. *Proceedings of the IEEE/CVF international conference on computer vision*, 2622-2631. doi:10.1109/ICCV48922.2021.00262

Miao, X., Yuan, X., Pu, Y., & Athitsos, V. (2019). l-net: Reconstruct hyperspectral images from a snapshot measurement. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4059-4069. doi:10.1109/ICCV.2019.00416

Nhi, N. V., Le, T. M., & Van, T. T.Thanh The Van. (2022). A model of semantic-based image retrieval using C-tree and neighbor graph. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–23. doi:10.4018/IJSWIS.295551

Qian, W., Li, H., & Mu, H. (2022). Circular lbp prior-based enhanced GAN for image style transfer. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(2), 1–15. doi:10.4018/IJSWIS.315601

Ramírez, J. M., Torre, J. I., & Arguello, H. (2021). LADMM-Net: An unrolled deep network for spectral image fusion from compressive data. *Signal Processing*, *189*, 108239. doi:10.1016/j.sigpro.2021.108239

Sun, L., Cao, Q., Chen, Y., Zheng, Y., & Wu, Z. (2023). Mixed noise removal for hyperspectral images based on global tensor low-rankness and nonlocal SVD-aided group sparsity. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–17. doi:10.1109/TGRS.2023.3335418

Udin, W. S., Norazami, N. A., Sulaiman, N., Zaudin, N. A., Marail, S., & Nor, A. S. (2019). UAV based multi-spectral imaging system for mapping landslide risk area along Jeli-Gerik highway, Jeli, Kelantan. *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE, 162-167.

Wagadarikar, A. A., John, R., Willett, R. M., & Brady, D. J. (2008). Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, *47*(10), B44–B51. doi:10.1364/AO.47.000B44 PMID:18382550

Wang, H., Li, Z., Li, Y., Gupta, B. B., & Choi, C. (2020). Visual saliency guided complex image retrieval. *Pattern Recognition Letters*, *130*, 64–72. doi:10.1016/j.patrec.2018.08.010

Wang, L., Xiong, Z., Gao, D., Shi, G., & Wu, F. (2015). Dual-camera design for coded aperture snapshot spectral imaging. *Applied Optics*, *54*(4), 848–858. doi:10.1364/AO.54.000848 PMID:25967796

Wang, L., Xiong, Z., Shi, G., Wu, F., & Zeng, W. (2016). Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(10), 2104–2111. doi:10.1109/TPAMI.2016.2621050 PMID:28113743

Wang, M., Wang, Q., & Chanussot, J. (2021). Tensor low-rank constraint and l_0 total variation for hyperspectral image mixed noise removal. *IEEE Journal of Selected Topics in Signal Processing*, *15*(3), 718–733. doi:10.1109/JSTSP.2021.3058503

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. doi:10.1109/TIP.2003.819861 PMID:15376593

Wright, S. L., Levermore, J. M., & Kelly, F. J. (2019). Raman spectral imaging for the detection of inhalable microplastics in ambient particulate matter samples. *Environmental Science & Technology*, *53*(15), 8947–8956. doi:10.1021/acs.est.8b06663 PMID:31293159

Yang, Z., Wei, Y., & Yang, Y. (2021). Associating objects with transformers for video object segmentation. *Advances in Neural Information Processing Systems*, *34*, 2491–2502.

Yasuma, F., Mitsunaga, T., Iso, D., & Nayar, S. K. (2010). Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum [J]. *IEEE Transactions on Image Processing*, *19*(9), 2241–2253. doi:10.1109/TIP.2010.2046811 PMID:20350852

Yu, C., Li, J., Li, X., Ren, X., & Gupta, B. B. (2018). Four-image encryption scheme based on quaternion Fresnel transform, chaos and computer generated hologram. *Multimedia Tools and Applications*, *77*(4), 4585–4608. doi:10.1007/s11042-017-4637-6

Yu, H., Xu, Z., Zheng, K., Hong, D., Yang, H., & Song, M. (2022). MSTNet: A multilevel spectral–spatial transformer network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–13. doi:10.1109/TGRS.2022.3186400

Zhang, J., & Ghanem, B. (2018). ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1828-1837. doi:10.1109/CVPR.2018.00196

Zhang, S., Wang, L., Fu, Y., Zhong, X., & Huang, H. (2019). Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10183-10192. doi:10.1109/ICCV.2019.01028

Zhang, T., Fu, Y., Wang, L., & Huang, H. (2019). Hyperspectral image reconstruction using deep external and internal learning. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8559-8568. doi:10.1109/ICCV.2019.00865