

## Hyperspectral image reconstruction based on the fusion of diffracted rotation blurred and clear images

Hao Xu<sup>a</sup>, Haiquan Hu<sup>a</sup>, Shiqi Chen<sup>a</sup>, Zihai Xu<sup>a</sup>, Qi Li<sup>a</sup>, Tingting Jiang<sup>b</sup>, Yueling Chen<sup>a,\*</sup>

<sup>a</sup> State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China

<sup>b</sup> Research Center for Intelligent Sensing Systems, Zhejiang Laboratory, Hangzhou 311100, China



### ARTICLE INFO

#### Keywords:

Hyperspectral image reconstruction  
Hyperspectral imaging  
Diffraction  
Image fusion  
Convolutional neural network

### ABSTRACT

To overcome the problems of imaging speed and bulky volume of the traditional hyperspectral imaging systems, the recently proposed compact, snapshot hyperspectral imaging system with diffracted rotation has attracted a lot of interest. Due to the severe degradation of the diffracted rotation blurred image, the restored hyperspectral image (HSI) suffers from a lack of spatial detail information and spectral accuracy. To improve the quality of the reconstructed HSI, we present a joint imaging system of diffractive imaging and clear imaging as well as a convolutional neural network (CNN) based method with two input branches for HSI reconstruction. In the reconstruction network, we develop a feature extraction block (FEB) to extract the features of the two input images, respectively. Subsequently, a double residual block (DRB) is designed to fuse and reconstruct the extracted features. Experimental results show that HSI with high spatial resolution and spectral accuracy can be reconstructed. Our method outperforms the state-of-the-art methods in terms of quantitative metrics and visual quality.

### 1. Introduction

Spectral imaging can simultaneously obtain 3D data cubes including the target's 2D spatial and 1D spectral information. Since different objects have varied spectral characteristics, spectral imaging has been widely used in scientific research and industry, such as agriculture [1,2], remote sensing [3,4], food safety inspection [5], and cultural heritage research [6,7]. Traditional hyperspectral imaging systems typically rely on bulky optical systems, movable parts, and long exposure time, resulting in limited application scenarios. Recent years have seen the development of snapshot hyperspectral imaging systems that can obtain spectral images without any temporal or spatial scanning, especially computed tomography imaging spectrometry (CTIS) [8,9], coded aperture snapshot spectral imaging (CASSI) [10,11] systems. These systems, however, are still based on traditional geometric optical systems; moreover, their bulky volume and weight limit their implementation in many scenarios with high portability and stability requirements. Jeon *et al.* [12] proposed a diffraction-based snapshot hyperspectral imaging system to address the limits of mobility, weight, and volume. A diffractive optical element (DOE) with spectrally-varying point spread functions (PSFs) is utilized to replace the common optical elements (prism, relay lenses, imaging lenses, etc.) in conventional systems, and the dispersion, as well as imaging, are accomplished simultaneously by the single thin DOE. Since the PSFs in different spectral channels have similar shapes

but different rotation angles, providing the possibility of reconstructing HSIs from diffraction blurred images.

However, the reconstruction of HSIs demands a considerable computational cost, and the quality of the reconstruction results usually suffers from the severe degradation of the captured images. At the same time, clear RGB imaging has been well developed and contains rich spatial information. Restoring spectral images directly from RGB images is severely ill-posed and needs the help of prior knowledge [13,14]. As mentioned above, high-quality HSIs cannot be recovered from either diffraction blurred or clear images. To improve the quality of reconstructed HSIs, we propose a joint imaging system of diffractive imaging and clear imaging. In the joint imaging system, diffraction blurred images imply much spectral information but have a low spatial resolution. On the contrary, clear images have fewer spectral features but contain rich spatial detail information. Integrating the features of these two kinds of images, we present an HSIs reconstruction neural network based on the fusion reconstruction of diffraction blurred images and clear images.

The proposed network has two input branches, with diffraction blurred images and clear images as the inputs, respectively. The sub-networks of the two branches perform feature extraction on the input images, and after feature fusion and reconstruction, a 3D data cube with high quality is obtained. Compared with traditional methods, our method significantly improves the quality of reconstruction outcomes.

\* Corresponding author.

E-mail address: [chenyt@zju.edu.cn](mailto:chenyt@zju.edu.cn) (Y. Chen).

Our contributions can be summarized as follows.

- We propose a joint diffractive imaging and clear imaging system that can provide hyperspectral images with high spatial resolution and spectral precision by fusing and reconstructing the images captured by the two imaging branches.
- We propose a feature extraction block (FEB) and a double residual block (DRB) based on the two-branch network framework, which effectively improve the network's capability to reconstruct hyperspectral images.
- Extensive experiments show that our method outperforms other state-of-the-art approaches in terms of PSNR, SSIM, and SAM metrics.

The rest of the paper is organized as follows. [Section 2](#) overview hyperspectral imaging systems and some available fusion-based reconstruction methods. In [Section 3](#), we present the joint imaging system's imaging model and image formation, particularly the diffractive imaging branch. The architecture of our fusion-based reconstruction network is detailed in [Section 4](#). [Section 5](#) details our experiments on publicly available datasets and compares our method with other state-of-the-art methods. Finally, the paper is concluded in [Section 6](#).

## 2. Related work

### 2.1. Hyperspectral imaging systems

Conventional hyperspectral imaging systems can only obtain one-dimensional spectral or two-dimensional spatial information through a single-shot. Scanning-based systems typically obtain spectral data cubes by scanning with the help of slits [15,16], dispersive elements, or narrow-band filters [17–19]. Since scanning is required, the applications of these systems in dynamic scenarios are pretty limited. In contrast, our method can capture a diffraction blurred image and a clear image through a single-shot and subsequently obtain HSI by reconstruction, enabling snapshot hyperspectral imaging.

To counter the limitations of scanning systems, computed tomography imaging spectrometry (CTIS) [8,9] uses dispersive elements to project incident light with varying spectral compositions to different areas on the image plane. Tomographic reconstruction techniques can recover the spatial and spectral information mixed on the image plane. Coded aperture snapshot spectral imaging (CASSI) [10,11,20] based on compressed sensing theory was introduced to capture dynamic scenes and realize snapshot measurement. In recent years, numerous efforts have been made to optimize the CASSI system. A digital micromirror device (DMD) was proposed for aperture encoding, offering the potential for dynamic encoding [21]. Colored coded aperture has been presented as an alternative to conventional binary coded aperture, which can minimize the number of captures while improving the quality of reconstruction results [22,23]. Wang *et al.* [24] introduced a dual-camera design into CASSI, and the input beam is divided into two through a beam splitter. With the advancement of neural networks, the optimization of CASSI systems has primarily concentrated on improving the reconstruction algorithm in recent years [11,25,26]. However, due to the CASSI's sophisticated optical system, it can only be utilized for laboratory spectral imaging. In contrast, our method has a simpler and more compact imaging system with better mobility.

Recent years have seen the introduction of neural networks into computational imaging, and many compact snapshot spectral imaging systems have emerged. Jeon *et al.* [12] proposed a diffractive imaging lens that leads to an anisotropic shape of spectrally-varying PSFs and can achieve imaging and dispersion with a single DOE and a bare image sensor. Mikko *et al.* [27] proposed using a regular digital camera equipped with a diffraction grating element consisting of an array of equally spaced horizontal and vertical slits, using machine learning to reconstruct HSI. Jian *et al.* [28] proposed employing reconfigurable metasurface supercells to encode the incident spectrum and integrating

numerous pixels into a micro spectrometer to enable one-shot ultraspectral imaging. Zhang *et al.* [29] developed a deeply learned broadband encoding stochastic hyperspectral camera, in which a deep neural network is used to design filters and reconstruct spectra. Using the end-to-end design framework, Baek *et al.* [30] proposed a differentiable pipeline that includes a wave-optics simulation and a hyperspectral-depth image reconstruction. However, HSIs obtained by these methods suffer from poor spatial resolution and spectral accuracy. Our method can reconstruct HSIs with more accurate spatial and spectral information.

### 2.2. Hyperspectral image recovery based on fusion

Despite the fact that hyperspectral images contain spectral information, their spatial resolution is relatively low. Images captured by normal cameras generally have a high spatial resolution. Efforts have been made to improve the quality of hyperspectral images using high-resolution images of the same field of view (FOV).

**Pansharpening** Directly obtaining multispectral (MS) or hyperspectral (HS) images is challenging in remote sensing. Pansharpening generates high-resolution MS images from high spatial resolution single band panchromatic (PAN) images and low spatial resolution MS images captured simultaneously by various sensors. Multiscale decomposition was performed to inject high-frequency spatial details into spectral images [31,32]. Based on sparse representation theory, a pansharpening method was proposed from the perspective of compressed sensing [33]. The traditional fusion methods have demonstrated effectiveness, and the MS image pansharpening also benefits from CNN. Masi *et al.* [34] first applied CNN to pansharpening by using a simple and effective three-layer architecture. Liu *et al.* [35] proposed to develop an encoder attention module in the feature extraction part to better utilize the spectral and spatial features of MS and PAN images. The network with two branches was proposed to extract the features of PAN and MS images, respectively [36].

**Fusion-based method in CASSI** Inspired by pansharpening, the CASSI system also adds an imaging path to obtain additional spatial detail information. Dual-camera design was first introduced to CASSI by Wang *et al.* [24]. Based on the dual-camera design, Zhang *et al.* [37] proposed to use coded images and external learning to learn the spatial-spectral correlation of HS images and utilize normal images and internal learning to guarantee generalization ability. Tao *et al.* [20] proposed utilizing a dictionary-based algorithm and guided filtering to overcome the block effect and blurring common in other related works.

## 3. System Configuration

In this section, we first describe our diffractive imaging and clear imaging combined imaging system. Then we introduce the diffractive imaging model and the image formation of two branches. The schematic of the joint diffractive rotation and clear imaging system is shown in [Fig. 1](#). The imaging light path is divided into two branches by a beam splitter, one of which obtains blurred images with diffracted rotation through the diffractive imaging system, and the another obtains clear RGB images through the normal imaging system.

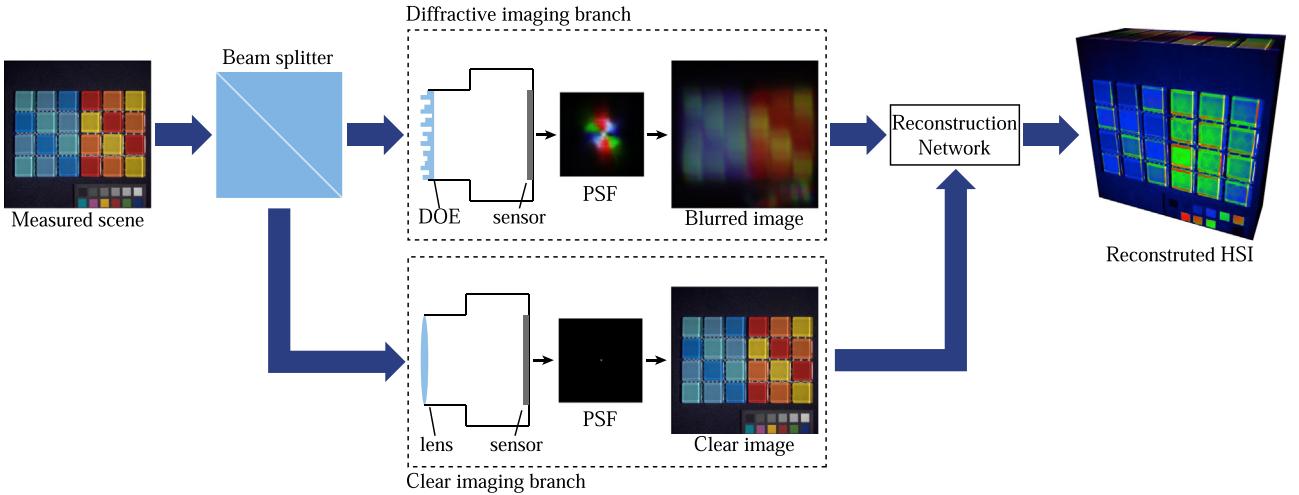
### 3.1. Diffractive imaging model

Consider the diffractive imaging model depicted in [Fig. 2](#). Assume that a point light source illuminates the imaging system consisting of a DOE and an image sensor at depth  $z$ , the intensity distribution on the sensor is the PSF of the system.

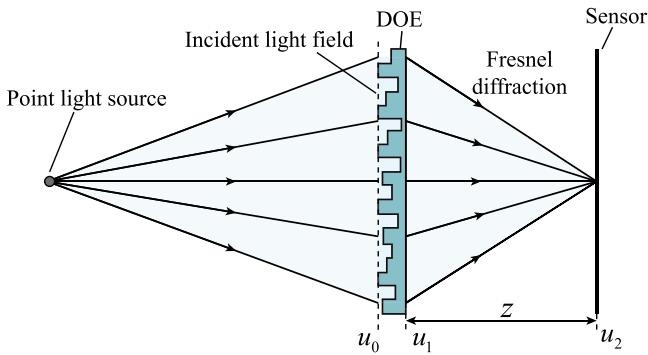
The incident light field of wavelength  $\lambda$  emitted by the point light source in front of the DOE can be expressed as

$$u_0(x, y) = A(x, y)e^{i\phi_0(x, y)}, \quad (1)$$

where  $A(x, y)$  is the amplitude,  $\phi_0(x, y)$  is the phase.



**Fig. 1.** Schematic of the joint hyperspectral imaging system. The two imaging branches can provide diffraction blurred images and normal clear images, respectively.



**Fig. 2.** Schematic diagram of forward propagation model in diffractive imaging.

After the phase modulation of the DOE, the light field passing through the DOE can be formulated as

$$u_1(x, y) = A(x, y) e^{i(\phi_0(x, y) + \phi_h(x, y))}, \quad (2)$$

where  $\phi_h(x, y)$  is the modulation term determined by the surface profile  $h(x, y)$  of the DOE, and can be expressed as

$$\phi_h(x, y) = \frac{2\pi}{\lambda} \Delta n_\lambda h(x, y), \quad (3)$$

where  $\Delta n_\lambda$  is difference between the refractive index of the substrate of DOE and air at wavelength  $\lambda$ .

Since the imaging distance  $z$  satisfies  $z \gg \lambda$ , the light field on the sensor can be obtained from the light field  $u_1(x, y)$  by the Fresnel propagation [38]:

$$u_2(x', y') = \frac{e^{ikz}}{i\lambda z} \iint u_1(x, y) \cdot e^{\frac{ik}{2z}((x-x')^2 + (y-y')^2)} dx dy, \quad (4)$$

where  $k = 2\pi/\lambda$  is the wavenumber.

Assume that the imaging system is focused at infinity, the incident light field can be approximated as a plane wave. This means that  $A(x, y)$ ,  $\phi_0(x, y)$  in Eq. 1 are constant terms. The incident plane wave can be described as  $u_0(x, y) = Ae^{i\phi_0}$ .

Substitute the simplified incident light field into Eq. 2 and Eq. 4, the Eq. 4 can be rewritten as

$$u_2(x', y') = \frac{e^{ikz}}{i\lambda z} \iint Ae^{i(\phi_0 + \phi_h(x, y))} e^{\frac{ik}{2z}((x-x')^2 + (y-y')^2)} dx dy. \quad (5)$$

The intensity distribution on the image plane is the square of the light field, which represents the point spread function  $p_\lambda(x', y')$  of the system. By Eq. 5, the point spread function can be described as the Fourier

transform of the product of the quadratic phase factor and the phase modulation term:

$$p_\lambda(x', y') \propto \left| \mathcal{F} \left[ Ae^{i\phi_0} e^{i\phi_h(x, y)} e^{\frac{ik}{2z}(x^2 + y^2)} \right] \right|^2. \quad (6)$$

### 3.2. The design of DOE

Consider the optical phase difference of two rays, a ray that passes through the DOE center along the optical axis and a ray that passes through a point on the DOE with coordinates  $(x, y)$ . Assume that  $h(x, y) = \Delta h(x, y) + h_0$ , where  $\Delta h(x, y)$  is the height difference between the DOE profile and the substrate,  $h_0$  is the height of the substrate. The optical path difference between the two rays is the outcome of the combined effect of DOE profiles and geometrical paths:

$$\delta = \Delta n_\lambda \Delta h(x, y) + \sqrt{x^2 + y^2 + z^2} - z. \quad (7)$$

To achieve the **imaging function**, the optical path difference should satisfy the constructive interference:

$$\Delta n_\lambda \Delta h(x, y) + \sqrt{x^2 + y^2 + z^2} - z = q\lambda, \quad (8)$$

where  $q$  is an integer. The  $\Delta h(x, y)$  can be simplified as

$$\Delta h(x, y) = \frac{q\lambda - (\sqrt{x^2 + y^2 + z^2} - z)}{\Delta n_\lambda}. \quad (9)$$

To achieve the **dispersion function**, the height profile of DOE needs to be specially designed for each wavelength along different angles. Convert  $(x, y)$  to polar coordinates  $(r, \theta)$ , where  $r$  is the distance from  $(x, y)$  to  $(0, 0)$  and  $\theta$  is the rotation angle of the line from  $(x, y)$  to  $(0, 0)$ . The circular DOE is divided into  $N$  equal parts to design for  $[\lambda_{\min}, \lambda_{\max}]$ . The wavelength corresponding to  $\theta$  is formed as

$$\lambda(\theta) = \begin{cases} \lambda_{\min} + (\lambda_{\max} - \lambda_{\min}) \frac{N}{2\pi} \theta & 0 \leq \theta < \frac{2\pi}{N} \\ \lambda \left( \theta - \frac{2\pi}{N} \right) & \frac{2\pi}{N} \leq \theta < 2\pi \end{cases}. \quad (10)$$

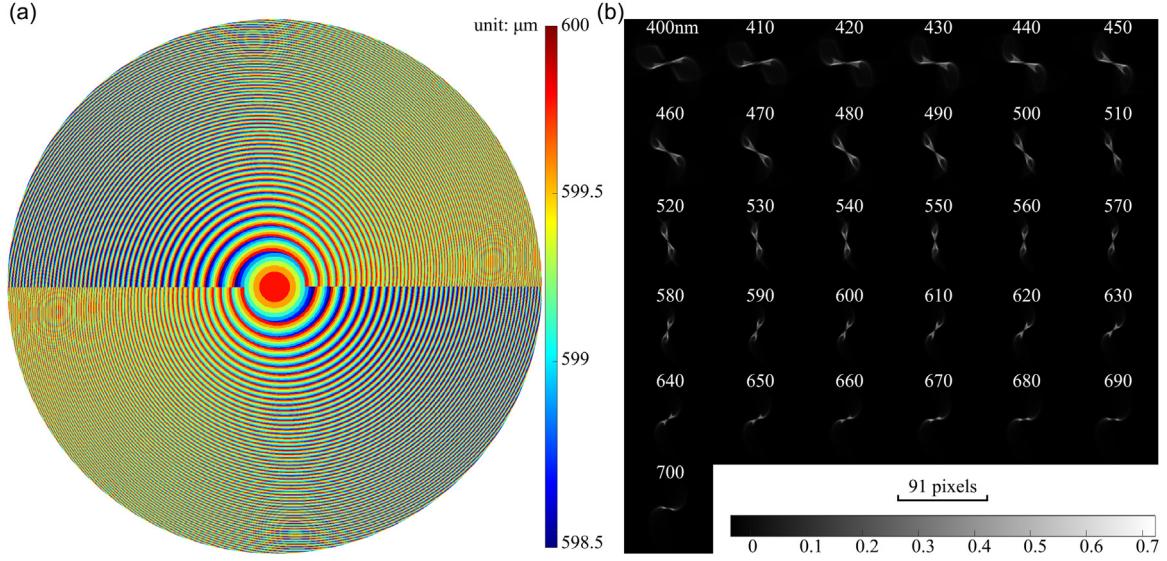
The  $N$  is also called the number of wings. Finally the height difference  $\Delta h(r, \theta)$  can be rewritten as

$$\Delta h(r, \theta) = \frac{q\lambda - (\sqrt{r^2 + z^2} - z)}{\Delta n_\lambda}. \quad (11)$$

Further, the height profile of the DOE can be expressed as

$$h(r, \theta) = h_0 + \Delta h(r, \theta). \quad (12)$$

**Fig. 3** illustrates the DOE design schematic and the corresponding spectrally-varying PSFs. As shown in Fig. 3(a), the entire DOE is composed of two identical parts. Each part's height profile at various angles



**Fig. 3.** (a) is our DOE design, the diameter of DOE is 5mm. We utilize different colors to indicate the height of the DOE surface profile. (b) is the corresponding spectrally-varying PSFs in wavelength range from 400nm to 700nm.

is designed for the wavelength range 400nm-700nm. The PSFs shown in Fig. 3(b) have a similar shape, but instead, the PSFs rotate as the wavelength changes. The PSF rotates around 180° between 400nm and 700nm, which also confirms that the number of wings used in our design is two.

### 3.3. Image Formation

**In the diffractive imaging branch**, a RGB diffraction blurred image  $J_c(x', y')$  can be obtained from a hyperspectral image  $I_\lambda(x', y')$ , where the subscript  $c \in \{r, g, b\}$  represents the color channel. Suppose that  $\omega_c(\lambda)$  is the spectral response of the sensor and the spectrally-varying PSF can be represented as  $p_\lambda(x', y')$ . The blurred image  $J_c(x', y')$  can be represented as

$$J_c(x', y') = \int \omega_c(\lambda) \cdot (I_\lambda \otimes p_\lambda)(x', y') d\lambda, \quad (13)$$

where  $\otimes$  is defined as convolution operator.

**In the clear imaging branch**, considering that when we construct the dataset for network training, the publicly available hyperspectral image datasets themselves carry the effect of optical aberration, the collected RGB clear image is the integral of the product of spectral response  $\omega_c(\lambda)$  and the HSI  $I_\lambda(x', y')$ , without any blur or dispersion. The clear image can be represented as

$$P_c(x', y') = \int \omega_c(\lambda) \cdot I_\lambda(x', y') d\lambda. \quad (14)$$

## 4. Spectral image reconstruction neural network

A fusion reconstruction network with two input branches is presented to restore HSIs with high spatial resolution and spectral accuracy from diffraction blurred images and clear images. The network is built on an encoder-decoder architecture, and its workflow can be divided into three stages: feature extraction, feature fusion, and HSI reconstruction.

The network collects input from two branches during the feature extraction stage. The input image for one branch is a diffraction blurred image, while the input image for the other is a normal clear image. The sub-networks of the two branches have the same structure but different weights. Following that, HSIs can be obtained through fusion and reconstruction. Fig. 4 illustrates the structure of our proposed network

in detail, with different types of layers represented by various colored blocks. "Conv(2,2)" in Fig. 4 denotes a convolutional layer with kernel size = 2 and stride = 2. Similarly, "Conv(3,1)" denotes a convolutional layer with kernel size = 3 and stride = 1.

### 4.1. Feature extraction block

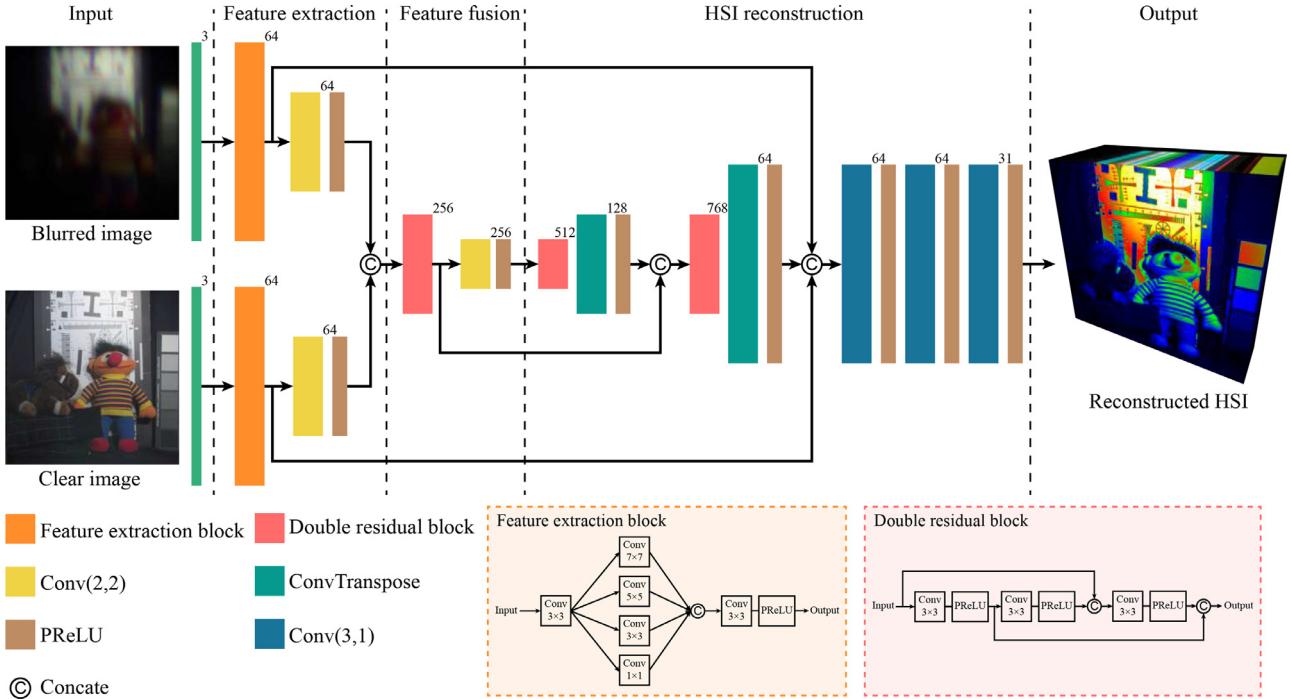
Due to the severe degradation of the blurred images (typically about 40-50 pixels), we propose a feature extraction block (FEB) built of convolution layers with varying kernel sizes to extract features of different scales. As shown in the bottom right corner of Fig. 4, the input of the FEB first passes through a 3×3 convolutional layer, followed by four convolutional layers with kernel sizes of 7, 5, 3, 1 respectively. Employing convolution kernels with diverse scales can fully perceive the features of the image at the early stage of feature extraction, avoiding image structure degradation caused by multiple downsampling procedures. To obtain the FEB output, the output of these four convolutional layers is concatenated and then fed into a 3×3 convolutional layer and a Parametric Rectified Linear Unit (PReLU) [39] layer.

### 4.2. Double residual block

We propose using double residual blocks (DRBs) in the feature fusion and hyperspectral image reconstruction stage to further increase the network's capability. Since the difference between input and output in DRB might be enormous, we employ concatenation rather than element-wise addition as in the traditional residual unit [40]. As shown in the bottom right corner of Fig. 4, assuming that the DRB's input has the shape of  $W \times H \times C$ , where  $W$ ,  $H$ , and  $C$  are width, height, and the number of channels, respectively, the input features of the second and third convolutional layers have the shape of  $W \times H \times C$  and  $W \times H \times 2C$ . The output scale of the DRB is  $W \times H \times 2C$ .

### 4.3. Loss function

The loss between the reconstructed hyperspectral images and the corresponding ground truth is minimized to optimize the parameters of the network. We use the mean absolute error (MAE) as the loss function. Specifically, assuming that the input blurred image is expressed as  $X_B^i$ , the clear input image is expressed as  $X_C^i$ . The corresponding ground truth is expressed as  $I_G^i$ , the mapping of the network is denoted as  $\mathcal{H}(\cdot)$ ,



**Fig. 4.** The proposed neural network architecture. Different layers are represented by various colored blocks (bottom left). "Conv(2,2)" denotes a convolutional layer with kernel size = 2 and stride = 2 and "Conv(3,1)" denotes a convolutional layer with kernel size = 3 and stride = 1. The feature extraction block and double residual block are detailed in the bottom right corner. The number of channels per layer is marked. We use MAE loss to optimize the network parameters.

then the  $\ell_1$  loss can be defined as:

$$\ell_1 = \frac{1}{N} \sum_i^N |\mathcal{H}(X_B^i, X_C^i) - I_G^i|, \quad (15)$$

where  $N$  is the number of samples in a mini-batch used for training.

## 5. Experiments and Discussion

In this section, we first introduce the dataset construction method and the metrics for quantitative evaluation employed in our experiment. Then the experimental implementation details are described, and the reconstruction results are presented in terms of image details and spectral accuracy. Finally, our method is compared with several state-of-the-art methods, including fusion and non-fusion.

### 5.1. Datasets and metrics

For training and testing, we selected hyperspectral images from publicly available ICVL [41], KAIST [42] and CAVE [43] datasets. The spectral range of ICVL and CAVE datasets are from 400nm to 700nm, while the KAIST dataset ranges from 400nm to 720nm. All hyperspectral images have a spectral interval of 10nm. ICVL dataset has a spatial resolution of about  $1300 \times 1392$  and covers numerous outdoor scenes with a shape of about  $1300 \times 1392 \times 31$ . KAIST dataset has a spatial resolution of  $2704 \times 3376$ , and all the scenes are captured indoors with a black background. The hyperspectral image in the KAIST dataset has a shape of  $2704 \times 3376 \times 33$ . CAVE dataset has a spatial resolution of  $512 \times 512$  and contains many comparisons between real and fake objects. The hyperspectral image in the CAVE dataset has a shape of  $512 \times 512 \times 31$ .

We randomly selected 31 images from the CAVE dataset, 21 from the KAIST dataset, and 158 images from the ICVL dataset for training. We sampled 8,980 tensor patches of size  $256 \times 256 \times 31$  from the selected hyperspectral images. The other nine images in the KAIST dataset, as well as 20 images randomly selected from the ICVL dataset, were used for testing. Gaussian noise was applied to all blurred and clear images

during training and testing with a mean value of 0 and a standard deviation of 0.01.

When constructing the dataset for training and testing, the pipeline of diffraction blurred images and clear images is shown in Fig. 5. We typically first convert the spectral image into CIEXYZ color space and subsequently into sRGB color space. Eq. 13 can be rephrased as

$$\begin{cases} J_X(x', y') = \sum_{i=1}^m X(\lambda_i) \cdot (I_{\lambda_i} \otimes p_{\lambda_i})(x', y') \\ J_Y(x', y') = \sum_{i=1}^m Y(\lambda_i) \cdot (I_{\lambda_i} \otimes p_{\lambda_i})(x', y') \\ J_Z(x', y') = \sum_{i=1}^m Z(\lambda_i) \cdot (I_{\lambda_i} \otimes p_{\lambda_i})(x', y') \end{cases} \quad (16)$$

where  $\{J_X, J_Y, J_Z\}$  is the XYZ representation of the blurred image,  $m$  is the number of channels of the HSI,  $\lambda_i$  is the wavelength at the  $i$ -th channel,  $X(\lambda)$ ,  $Y(\lambda)$  and  $Z(\lambda)$  are CIE 1931 color matching functions. The same as with Eq. 14:

$$\begin{cases} P_X(x', y') = \sum_{i=1}^m X(\lambda_i) \cdot I_{\lambda_i}(x', y') \\ P_Y(x', y') = \sum_{i=1}^m Y(\lambda_i) \cdot I_{\lambda_i}(x', y') \\ P_Z(x', y') = \sum_{i=1}^m Z(\lambda_i) \cdot I_{\lambda_i}(x', y') \end{cases} \quad (17)$$

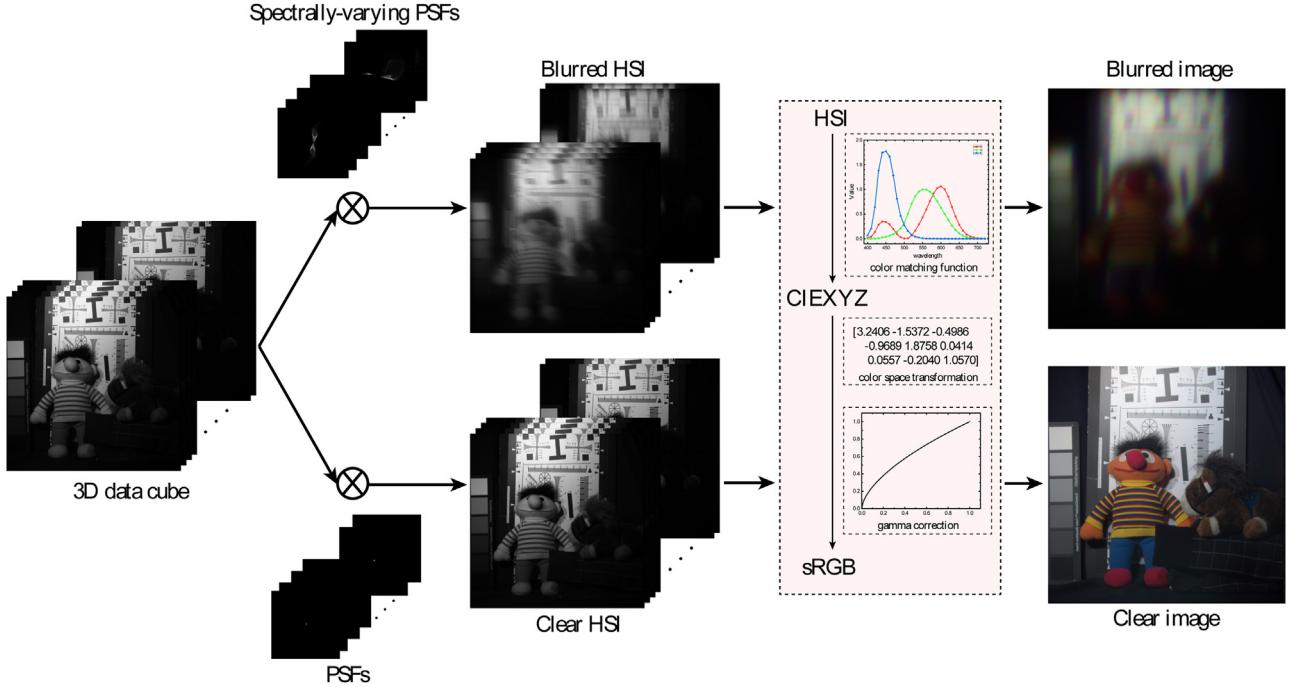
where  $\{P_X, P_Y, P_Z\}$  is the XYZ representation of the clear image.

The coordinates in the CIEXYZ color space are converted to the sRGB color space, and the color images are obtained after gamma correction. For more details, please refer to the tutorial [44].

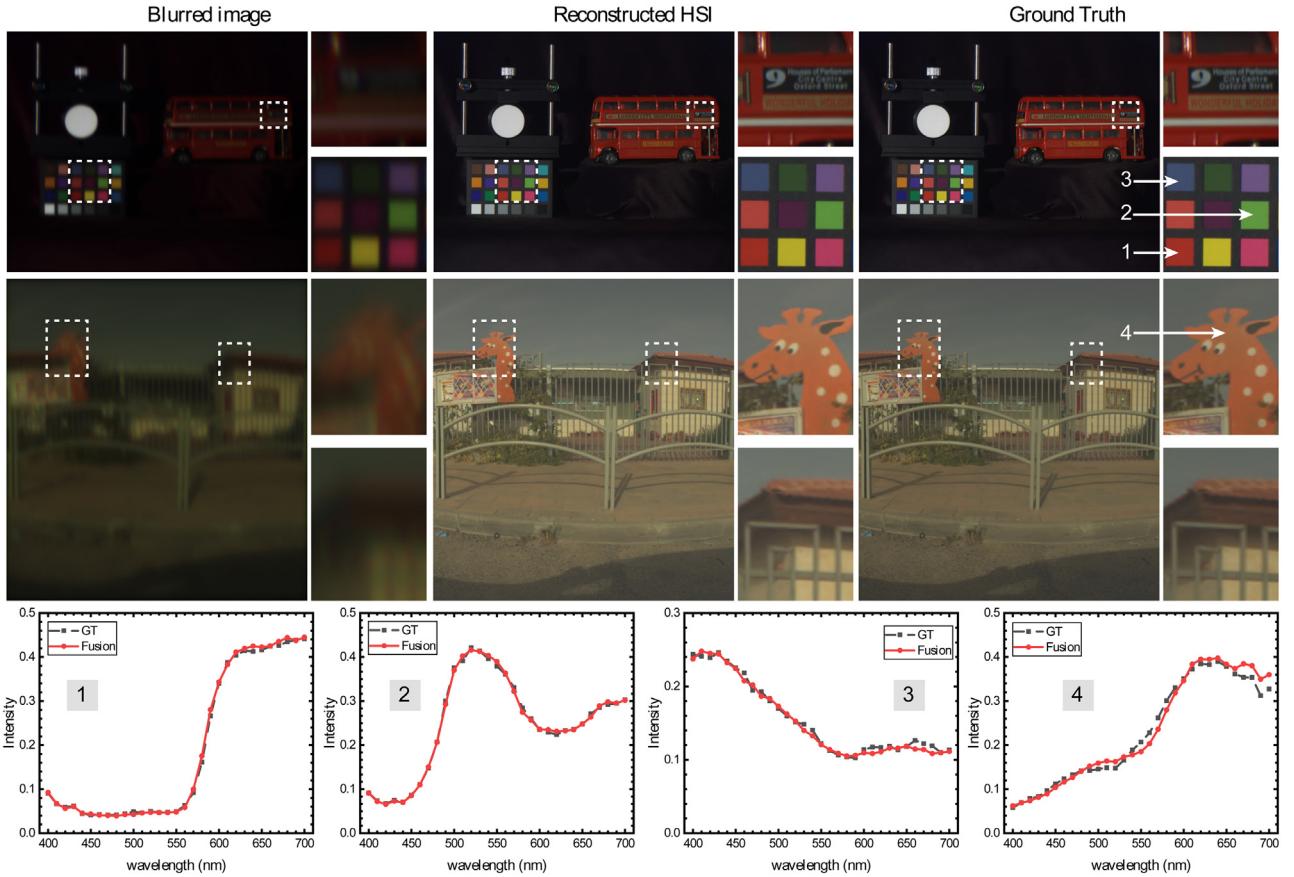
To quantitatively evaluate the performance of all methods, we calculate the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [45], and spectral angle mapping (SAM) [46] of these methods in each spectral channel respectively, and averaging along the channel as the metrics. Higher PSNR and SSIM mean better performance, while a smaller SAM indicates a higher spectral precision.

### 5.2. Implementation details and reconstruction results

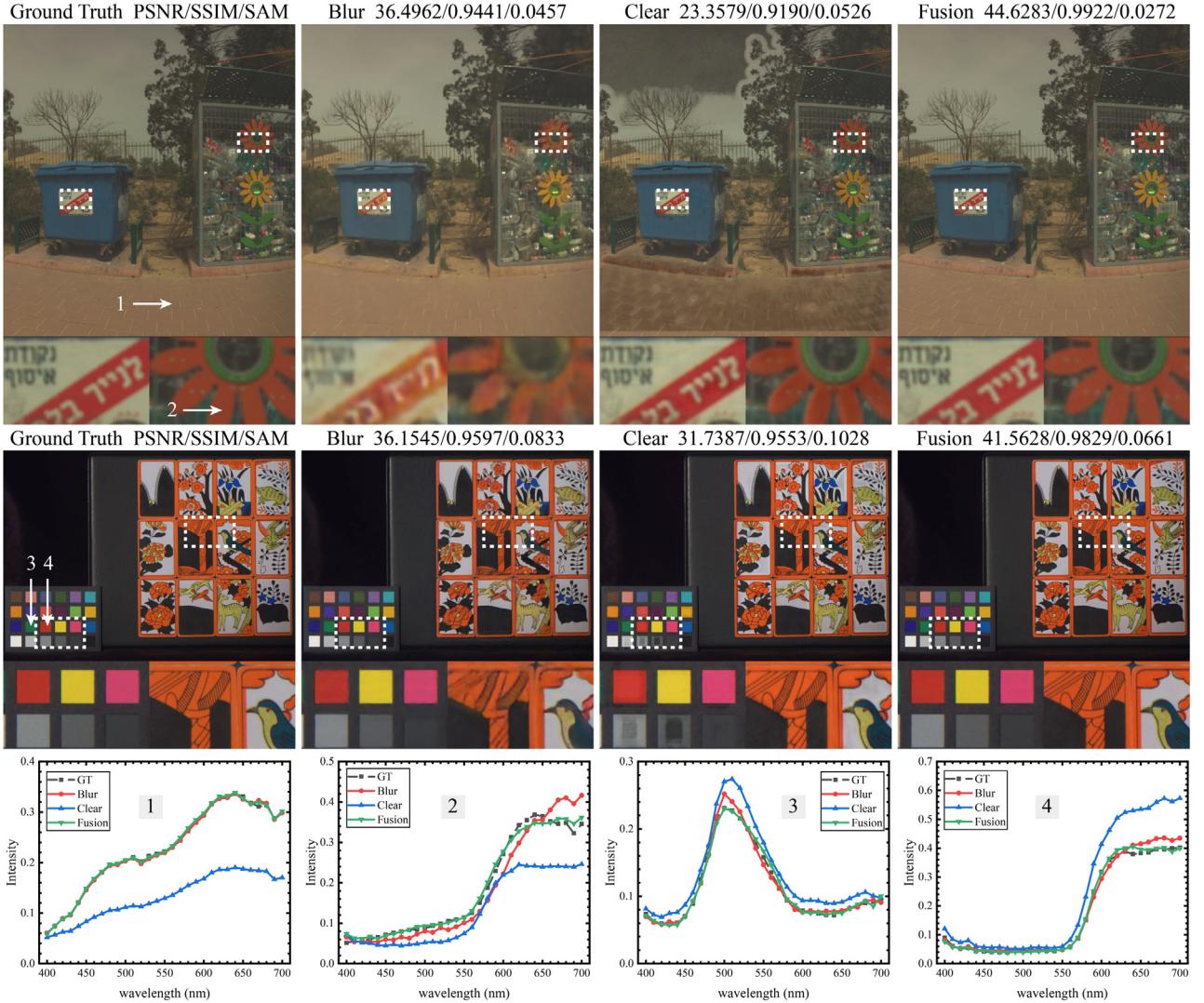
We design our DOE from 400nm to 700nm. The DOE has a focal length of 100mm, a diameter of 5mm, and a substrate thickness



**Fig. 5.** The pipeline of generating diffraction blurred images and clear images. The input image is a 3D data cube from hyperspectral image dataset. Since the 3D data cube carries the effect of optical aberration, the PSFs of the clear image do not introduce any image degradation. "⊗" denotes convolution.



**Fig. 6.** The reconstruction results of our method. The hyperspectral images are represented as RGB color. The diffraction blurred images are generated by convolution of ground truth and spectrally-varying PSFs. The RMSE of the four selected points' spectral curves are 0.005761, 0.005027, 0.005612 and 0.016279 respectively.



**Fig. 7.** We compare the reconstruction results while only the diffraction blurred images (denoted as "Blur"), the clear images (denoted as "Clear"), and both are used (denoted as "Fusion"). The hyperspectral images are represented as RGB color and the spectral curves of the three methods at the points of interest are also shown in detail.

of 0.6mm. We adjust the number of wings to two to obtain a balanced performance in both spatial resolution and spectral accuracy [47]. The DOE is discretized into  $5000 \times 5000$ , and the height difference  $\Delta h(x, y)$  is discretized into eight steps to enable calculation and fabrication.

We use the PyTorch framework to implement our network framework design. The model is optimized by using the stochastic gradient method with the ADAM optimizer [48]. The mini-batch size is set to 16, and the weight decay is set to 0. The learning rate is initially set to be  $10^{-4}$ , and will be divided by two every 20 epochs. It took about 17 hours to train the network using a workstation equipped with AMD EPYC 7543 and NVIDIA RTX A6000. During the testing phase, it took about 10ms to reconstruct a hyperspectral image with a scale of  $2704 \times 3376 \times 31$ .

To verify the effectiveness of our method, we fed the blurred and clear images into the two branches of the network, and the reconstruction results are displayed in Fig. 6. Hyperspectral images are represented as RGB color to facilitate observing spatial details and reconstruction quality. It can be seen from Fig. 6 that the spatial details of blurred images with spectrally-varying PSFs can not be distinguished. The spatial details and spectral precision of the hyperspectral images reconstructed by our method can be remarkably consistent with the ground truth. The root mean square error (RMSE) of these four spectral curves

and ground truth are 0.005761, 0.005027, 0.005612, and 0.016279, respectively.

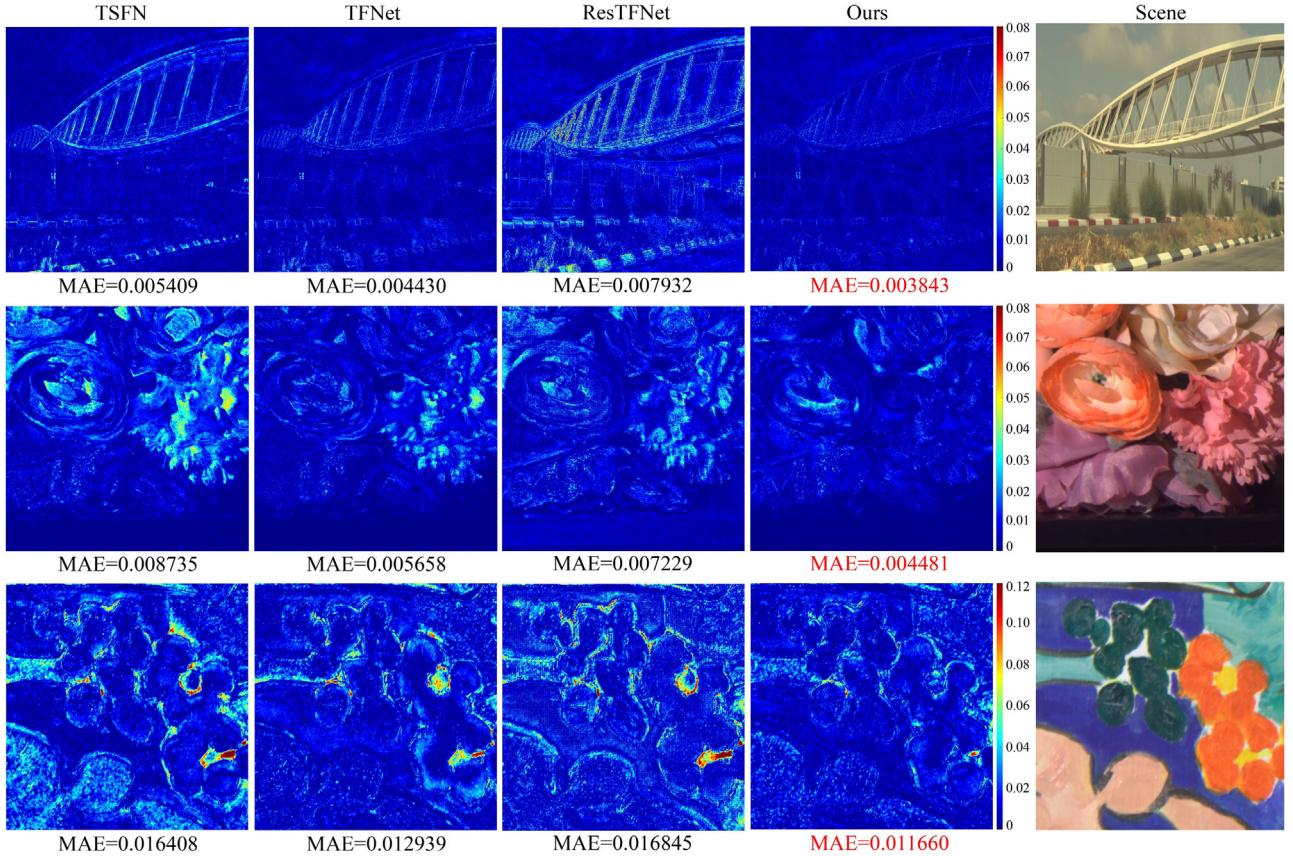
### 5.3. Compared with other methods

To confirm that clear images contain more spatial details while blurred images have latent spectral features, we compared the performance of three methods:

- Only blurred images are used for reconstruction.
- Only clear images are used for reconstruction.
- Both blurred and clear images are used for reconstruction.

We adjust the network structure in the experiment with only one image as input, which means that only one of the two sub-networks is retained in the feature extraction stage. The reconstruction results and spectral curves of these three methods are shown in Fig. 7 respectively. The RMSEs of the four spectral curves on the selected points are shown in Table 1.

The spatial details depicted in Fig. 7 show that the spatial details of the result reconstructed from the clear images are rich, which is near to the resolution of the ground truth. However, there are numerous error messages in the flat place, and the spectral curves of the selected



**Fig. 8.** Visual quality comparison on three typical scenes in HSI datasets. The error maps of various fusion-based methods at 500 nm are displayed. From left to right correspond to the reconstruction results of TSFN/TFNet/ResTFNet/ours respectively.

**Table 1**

The RMSE of the four spectral curves on the selected points.

	1	2	3	4
Blur	0.0032	0.0321	0.0065	0.0176
Clear	0.1039	0.0664	0.0212	0.0919
Fusion	<b>0.0027</b>	<b>0.0110</b>	<b>0.0053</b>	<b>0.0078</b>

point deviate significantly from the ground truth. On the other hand, the results reconstructed from the blurred images have a poor spatial resolution, and many details are indistinguishable. However, the spectral curves are closer to the ground truth than the result reconstructed from the clear images. Using both blurred and clear images for reconstruction combines the benefits of the other two methods and has the best spatial

resolution and spectral precision performance. The quantitative performance of these three methods on the two datasets is shown in [Table 2](#). The method of using only clear images is denoted as "ours-clear," the method of using only blurred images is denoted as "ours-blur," and the method of using both blurred and clear images is denoted as "ours."

Since the PSNR/SSIM/SAM of all channels are calculated and averaged, the incorrect information of several spectral channels has a significant detrimental impact on the metric calculation. This is also why we think that, while the spatial resolution of the reconstruction results from clear images is better, the PSNR/SSIM/SAM is inferior.

We compared our method with other state-of-the-art hyperspectral image reconstruction methods on KAIST and ICVL datasets. These methods are mainly divided into three categories. The first category is the spectral super-resolution networks such as HSCNN-R [\[49\]](#), HSCNN-D [\[49\]](#) and previously mentioned "ours clear," which only use clear im-

**Table 2**  
Quantitative evaluation indexes of different methods.

Dataset		KAIST			ICVL		
		PSNR↑	SSIM↑	SAM↓	PSNR↑	SSIM↑	SAM↓
Clear→HSI	HSCNN-R	31.5312	0.9339	0.1943	29.1905	0.9467	0.0373
	HSCNN-D	30.8985	0.9425	0.1964	27.7471	0.8977	0.0433
	our-clear	32.0981	0.9442	0.1844	29.1490	0.9526	0.0382
Blur→HSI	Jeon's	32.3672	0.9352	0.1764	32.2034	0.9099	0.0554
	our-blur	37.3970	0.9704	0.1104	38.3968	0.9440	0.0370
Fusion→HSI	PNN	30.0369	0.8940	0.3320	32.6725	0.9326	0.0818
	TFNet	40.7239	0.9834	0.0910	45.9293	0.9911	0.0251
	ResTFNet	38.5182	0.9695	0.1257	42.1191	0.9847	0.0309
	TSFN	39.0797	0.9748	0.1328	42.9550	0.9878	0.0342
	ours	<b>41.3280</b>	<b>0.9850</b>	<b>0.0817</b>	<b>47.4110</b>	<b>0.9932</b>	<b>0.0213</b>

ages to reconstruct hyperspectral images. We implement HSCNN-R with 12 residual blocks (256 filters in each layer) and HSCNN-D with 12 dense blocks. The second category is to reconstruct hyperspectral images only by using blurred images. Specifically the optimization-based unrolled network proposed by Jeon *et al.* [12], and the previously mentioned "ours blur." The third category is the methods that utilize clear and blurred images to reconstruct hyperspectral images simultaneously. These methods notably PNN [34], TFNet [36] and TFNet's modified network ResTFNet [36], TTSFN [50], are mostly employed for pansharpening.

**Table 2** shows the quantitative evaluation performance of these reconstruction methods on the two datasets. From **Table 2**, we can see that the majority of the fusion methods outperform the traditional methods, and the fusion reconstruction method we proposed has achieved the best performance in terms of PSNR/SSIM/SAM. To assess the visual quality of hyperspectral reconstruction, **Fig. 8** shows the error maps of various fusion-based well-performed methods at 500 nm. The error maps illustrate that our method has less error than the baseline TFNet and has the best performance compared with other state-of-the-art methods.

## 6. Conclusion

In this paper, we proposed a joint hyperspectral imaging system of clear and diffractive imaging branches. The imaging system captures clear images with abundant spatial information and diffracted rotation blurred images with latent spectral features simultaneously. Furthermore, we proposed a hyperspectral image fusion and reconstruction network with a two-branch input. Utilize the two images captured by the joint imaging system as input; the proposed network effectively improves the quality of the spectral images through feature extraction, feature fusion, and image reconstruction. The reconstruction method proposed in this work outperforms other state-of-the-art methods in terms of spatial resolution and spectral precision.

In the following, we may further build a joint imaging system and apply it in real-world imaging experiments. In addition, we will explore more feasible imaging systems that can simultaneously acquire diffraction blurred images and normal clear images in our subsequent work.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Hao Xu:** Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Haiquan Hu:** Software, Validation, Data curation, Writing – review & editing. **Shiqi Chen:** Software, Validation. **Zhihai Xu:** Resources, Validation, Project administration, Funding acquisition. **Qi Li:** Resources, Supervision, Project administration. **Tingting Jiang:** Resources, Supervision, Project administration. **Yueling Chen:** Supervision, Formal analysis, Project administration, Writing – review & editing.

## Acknowledgments

This project is supported by Key Research Project of Zhejiang Lab (No. 2021MH0AC01), National Natural Science Foundation of China (No. 62275229) and Civil Aerospace Pre-Research Project (No. D040104). We thank Meijuan Bian from the facility platform of optical engineering of Zhejiang University for instrument support. We gratefully acknowledge Jiaqi Yang and Zhaojie Chen for fruitful discussion.

## References

- [1] Sun W, Liu S, Zhang X, Li Y. Estimation of soil organic matter content using selected spectral subset of hyperspectral data. *Geoderma* 2022;409:115653. doi:[10.1016/j.geoderma.2021.115653](https://doi.org/10.1016/j.geoderma.2021.115653).
- [2] Wang F, Yi Q, Hu J, Xie L, Yao X, Xu T, Zheng J. Combining spectral and textual information in uav hyperspectral images to estimate rice grain yield. *International Journal of Applied Earth Observation and Geoinformation* 2021;102:102397. doi:[10.1016/j.jag.2021.102397](https://doi.org/10.1016/j.jag.2021.102397).
- [3] He J, Barton I. Hyperspectral remote sensing for detecting geotechnical problems at ray mine. *Engineering Geology* 2021;292:106261. doi:[10.1016/j.enggeo.2021.106261](https://doi.org/10.1016/j.enggeo.2021.106261).
- [4] Zhang Y, Migliavacca M, Penuelas J, Ju W. Advances in hyperspectral remote sensing of vegetation traits and functions. *Remote Sensing of Environment* 2021;252:112121. doi:[10.1016/j.rse.2020.112121](https://doi.org/10.1016/j.rse.2020.112121).
- [5] Feng Y-Z, Sun D-W. Application of hyperspectral imaging in food safety inspection and control: A review. *Critical Reviews in Food Science and Nutrition* 2012;52(11):1039–58. doi:[10.1080/10408398.2011.651542](https://doi.org/10.1080/10408398.2011.651542). PMID: 22823350
- [6] Costanzo A, Ebbole D, Ruffolo SA, Falcone S, la Piana C, La Russa MF, Musacchio M, Buongiorno MF. Detection of the tio2 concentration in the protective coatings for the cultural heritage by means of hyperspectral data. *Sustainability* 2021;13(1). doi:[10.3390/su13010092](https://doi.org/10.3390/su13010092).
- [7] Picollo M, Cucci C, Casini A, Stefani L. Hyper-spectral imaging technique in the cultural heritage field: New possible scenarios. *Sensors* 2020;20(10). doi:[10.3390/s20102843](https://doi.org/10.3390/s20102843).
- [8] Descour MR, Volin CE, Derenik EL, Thome KJ, Schumacher AB, Wilson DW, Maker PD. Demonstration of a high-speed nonscanning imaging spectrometer. *Opt Lett* 1997;22(16):1271–3. doi:[10.1364/OL.22.001271](https://doi.org/10.1364/OL.22.001271).
- [9] Han W, Wang Q, Cai W. Computed tomography imaging spectrometry based on superiorization and guided image filtering. *Opt Lett* 2021;46(9):2208–11. doi:[10.1364/OL.418355](https://doi.org/10.1364/OL.418355).
- [10] Gehm ME, John R, Brady DJ, Willett RM, Schulz TJ. Single-shot compressive spectral imaging with a dual-disperser architecture. *Opt Express* 2007;15(21):14013–27. doi:[10.1364/OE.15.014013](https://doi.org/10.1364/OE.15.014013).
- [11] Meng Z, Ma J, Yuan X. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In: Vedaldi A, Bischof H, Brox T, Frahm J-M, editors. *Computer Vision – ECCV 2020*. Cham: Springer International Publishing; 2020. p. 187–204. ISBN 978-3-030-58592-1.
- [12] Jeon DS, Baek S-H, Yi S, Fu Q, Dun X, Heidrich W, Kim MH. Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Trans Graph* 2019;38(4). doi:[10.1145/3306346.3322946](https://doi.org/10.1145/3306346.3322946).
- [13] Arad B, Ben-Shahar O. Sparse recovery of hyperspectral signal from natural rgb images. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Computer Vision – ECCV 2016*. Cham: Springer International Publishing; 2016. p. 19–34. ISBN 978-3-319-46478-7.
- [14] Akhtar N, Mian A. Hyperspectral recovery from rgb images using gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2020;42(1):100–13. doi:[10.1109/TPAMI.2018.2873729](https://doi.org/10.1109/TPAMI.2018.2873729).
- [15] Mouroulis P, Green RO, Chrien TG. Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information. *Appl Opt* 2000;39(13):2210–20. doi:[10.1364/AO.39.002210](https://doi.org/10.1364/AO.39.002210).
- [16] Kaiser S, Sang B, Schubert J, Hofer S, Stoeffler T. Compact prism spectrometer of pushbroom type for hyperspectral imaging. In: Mazuray L, Wartmann R, Wood A, Tissot J-L, Raynor JM, editors. *Optical Design and Engineering III*, vol. 7100. International Society for Optics and Photonics; 2008. p. 398–408. doi:[10.1117/12.797177](https://doi.org/10.1117/12.797177).
- [17] Brauers J, Schulte N, Aach T. Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms. *IEEE Transactions on Image Processing* 2008;17(12):2368–80. doi:[10.1109/TIP.2008.2006605](https://doi.org/10.1109/TIP.2008.2006605).
- [18] Gebhart SC, Thompson RC, Mahadevan-Jansen A. Liquid-crystal tunable filter spectral imaging for brain tumor demarcation. *Appl Opt* 2007;46(10):1896–910. doi:[10.1364/AO.46.001896](https://doi.org/10.1364/AO.46.001896).
- [19] Kasli PM, Vo-Dinh T. Hyperspectral imaging system using acousto-optic tunable filter for flow cytometry applications. *Cytometry Part A* 2006;69A(8):835–41. doi:[10.1002/cyto.a.20307](https://doi.org/10.1002/cyto.a.20307).
- [20] Tao C, Zhu H, Sun P, Wu R, Zheng Z. Hyperspectral image recovery based on fusion of coded aperture snapshot spectral imaging and rgb images by guided filtering. *Optics Communications* 2020;458:124804. doi:[10.1016/j.optcom.2019.124804](https://doi.org/10.1016/j.optcom.2019.124804).
- [21] Wu Y, Mirza IO, Arce GR, Prather DW. Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Opt Lett* 2011;36(14):2692–4. doi:[10.1364/OL.36.002692](https://doi.org/10.1364/OL.36.002692).
- [22] Correa CV, Arguello H, Arce GR. Snapshot colored compressive spectral imager. *J Opt Soc Am A* 2015;32(10):1754–63. doi:[10.1364/JOSAA.32.001754](https://doi.org/10.1364/JOSAA.32.001754).
- [23] Parada-Mayorga A, Arce GR. Colored coded aperture design in compressive spectral imaging via minimum coherence. *IEEE Transactions on Computational Imaging* 2017;3(2):202–16. doi:[10.1109/TCI.2017.2692649](https://doi.org/10.1109/TCI.2017.2692649).
- [24] Wang L, Xiong Z, Gao D, Shi G, Wu F. Dual-camera design for coded aperture snapshot spectral imaging. *Appl Opt* 2015;54(4):848–58. doi:[10.1364/AO.54.000848](https://doi.org/10.1364/AO.54.000848).
- [25] Liu Y, Yuan X, Suo J, Brady DJ, Dai Q. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2019;41(12):2990–3006. doi:[10.1109/TPAMI.2018.2873587](https://doi.org/10.1109/TPAMI.2018.2873587).
- [26] Jeon DS, Choi I, Kim MH. Multisampling compressive video spectroscopy. *Computer Graphics Forum* 2016;35(2):467–77. doi:[10.1111/cgf.12847](https://doi.org/10.1111/cgf.12847).
- [27] Toivonen ME, Rajani C, Klami A. Snapshot hyperspectral imaging using wide dilation networks. *Machine Vision and Applications* 2020;32(1):9. doi:[10.1007/s00138-020-01136-8](https://doi.org/10.1007/s00138-020-01136-8).

- [28] Xiong J, Cai X, Cui K, Huang Y, Yang J, Zhu H, Li W, Hong B, Rao S, Zheng Z, Xu S, He Y, Liu F, Feng X, Zhang W. Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces. *Optica* 2022;9(5):461–8. doi:[10.1364/OPTICA.440013](https://doi.org/10.1364/OPTICA.440013).
- [29] Zhang W, Song H, He X, Huang L, Zhang X, Zheng J, Shen W, Hao X, Liu X. Deeply learned broadband encoding stochastic hyperspectral imaging. *Light: Science & Applications* 2021;10(1):108. doi:[10.1038/s41377-021-00545-2](https://doi.org/10.1038/s41377-021-00545-2).
- [30] Baek S-H, Ikoma H, Jeon DS, Li Y, Heidrich W, Wetzstein G, Kim MH. Single-shot hyperspectral-depth imaging with learned diffractive optics. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021. p. 2651–60.
- [31] Vivone G, Restaino R, Chanussot J. Full scale regression-based injection coefficients for panchromatic sharpening. *IEEE Transactions on Image Processing* 2018;27(7):3418–31. doi:[10.1109/TIP.2018.2819501](https://doi.org/10.1109/TIP.2018.2819501).
- [32] Aiazzi B, Alparone L, Baronti S, Garzelli A, Selva M. Mtf-tailored multiscale fusion of high-resolution ms and pan imagery. *Photogrammetric Engineering & Remote Sensing* 2006;72(5):591–6.
- [33] Li S, Yang B. A new pan-sharpening method using a compressed sensing technique. *IEEE Transactions on Geoscience and Remote Sensing* 2011;49(2):738–46. doi:[10.1109/TGRS.2010.2067219](https://doi.org/10.1109/TGRS.2010.2067219).
- [34] Masi G, Cozzolino D, Verdoliva L, Scarpa G. Pansharpening by convolutional neural networks. *Remote Sensing* 2016;8(7). doi:[10.3390/rs8070594](https://doi.org/10.3390/rs8070594).
- [35] Liu Q, Han L, Tan R, Fan H, Li W, Zhu H, Du B, Liu S. Hybrid attention based residual network for pansharpening. *Remote Sensing* 2021;13(10). doi:[10.3390/rs13101962](https://doi.org/10.3390/rs13101962).
- [36] Liu X, Liu Q, Wang Y. Remote sensing image fusion based on two-stream fusion network. *Information Fusion* 2020;55:1–15. doi:[10.1016/j.inffus.2019.07.010](https://doi.org/10.1016/j.inffus.2019.07.010).
- [37] Zhang T, Fu Y, Wang L, Huang H. Hyperspectral image reconstruction using deep external and internal learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2019.
- [38] O’Shea D, Suleski T, of Photo-optical Instrumentation Engineers S, Kathman A, Prather D. Diffractive Optics: Design, Fabrication, and Test. Society of Photo Optical; 2004. ISBN 9780819451712.
- [39] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2015.
- [40] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.
- [41] Arad B, Ben-Shahar O. Sparse recovery of hyperspectral signal from natural rgb images. In: European Conference on Computer Vision. Springer; 2016. p. 19–34.
- [42] Choi I, Jeon DS, Nam G, Gutierrez D, Kim MH. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics (Proc SIGGRAPH Asia 2017)* 2017;36(6):218:1–13. doi:[10.1145/3130800.3130810](https://doi.org/10.1145/3130800.3130810).
- [43] Yasuma F, Mitsunaga T, Iso D, Nayar S. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Tech. Rep.; 2008.
- [44] Foster DH, Amano K. Hyperspectral imaging in color vision research: tutorial. *J Opt Soc Am A* 2019;36(4):606–27. doi:[10.1364/JOSAA.36.000606](https://doi.org/10.1364/JOSAA.36.000606).
- [45] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 2004;13(4):600–12. doi:[10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861).
- [46] Kruse F, Lefkoff A, Boardman J, Heidebrecht K, Shapiro A, Barloon P, Goetz A. The spectral image processing system (sips)interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment* 1993;44(2):145–63. doi:[10.1016/0034-4257\(93\)90013-N](https://doi.org/10.1016/0034-4257(93)90013-N). Airbone Imaging Spectrometry
- [47] Hu H, Zhou H, Xu Z, Li Q, Feng H, Chen Y, Jiang T, Xu W. Practical snapshot hyperspectral imaging with doe. *Optics and Lasers in Engineering* 2022;156:107098. doi:[10.1016/j.optlaseng.2022.107098](https://doi.org/10.1016/j.optlaseng.2022.107098).
- [48] Kingma D.P., Ba J.. Adam: A method for stochastic optimization. 2014. 10.48550/ARXIV.1412.6980.
- [49] Shi Z, Chen C, Xiong Z, Liu D, Wu F. Hscn+: Advanced cnn-based hyperspectral recovery from rgb images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; 2018.
- [50] Wang X, Chen J, Wei Q, Richard C. Hyperspectral image super-resolution via deep prior regularization with parameter estimation. *IEEE Transactions on Circuits and Systems for Video Technology* 2022;32(4):1708–23. doi:[10.1109/TCSVT.2021.3078559](https://doi.org/10.1109/TCSVT.2021.3078559).