IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 July 2024, accepted 21 August 2024, date of publication 29 August 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3451560

RESEARCH ARTICLE

High Accurate and Efficient 3D Network for Image Reconstruction of Diffractive-Based Computational Spectral Imaging

HAO FAN[1,2], CHENXI LI[1], HUANGRONG XU[1], LVRONG ZHAO[1,2], XUMING ZHANG[3], HENG JIANG[3], AND WEIXING YU[1,2]

[1]Key Laboratory of Spectral Imaging Technology of Chinese Academy of Sciences, Xi'an Institute of Optics and Precision Mechanics, Xi'an 710119, China
[2]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[3]Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong, China

Corresponding author: Weixing Yu (yuwx@opt.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62061160488, Grant 61975231, and Grant 62305376; in part by China National Key Research and Development Program under Grant 2021YFC2202002; in part by the Natural Science Basic Research Program of Shaanxi Province under Grant 2021JQ-130; and in part by the CAS Specific Research Assistant Funding Program under Grant E229431101.

ABSTRACT Diffractive optical imaging spectroscopy as a promising miniaturized and high throughput portable spectral imaging technique suffers from the problem of low precision and slow speed, which limits its wide use in various applications. To reconstruct the diffractive spectral image more accurately and fast, a three-dimensional spectrum recovery algorithm is proposed in this paper. The algorithm takes advantage of a neural network for image reconstruction which consists of a U-Net architecture with 3D convolutional layers to improve the processing precision and speed. Numerical experiments are conducted to prove its effectiveness. It is shown that the mean peak signal-to-noise ratio (MPSNR) of the recovered image relative to the original image is improved by 1.8 dB in comparison to other traditional methods. In addition, the obtained mean structural similarity (MSSIM) of 0.91 meets the standard of discrimination to human eyes. Moreover, the algorithm runs in just 0.36 s, which is faster than other traditional methods. 3D convolutional networks play a critical role in performance improvement. Improvements in processing speed and accuracy have greatly benefited the realization and application of diffractive optical imaging spectroscopy. The new algorithm with high accuracy and fast speed has a great potential application in diffraction lens spectroscopy and paves a new way for emerging more portable spectral imaging technique.

INDEX TERMS Computational imaging, spectral imaging, inverse problems, diffractive lenses.

I. INTRODUCTION

Hyperspectral imaging, which acquires information about the spatial and spectral dimensions of a scene simultaneously, has a wide range of applications in various fields such as medical diagnosis [1], art conservation [2], food quality analysis [3], agriculture [4], and water resources mapping [5]. However, traditional hyperspectral imaging systems typically possess the large system size and poor mobility since their spatial scanning or spectral scanning relying on complex prisms or gratings. It is time-consuming that spatial scanning acquires the complete spectrum of each pixel through a point-scanning instrument or a line-scanning instrument. Spectral scanning methods capture the entire scene in a single exposure using an array of photodetector detectors, and then progressively scan through the wavelengths to complete the data cube. However, Conventional hyperspectral imaging with the spectral separation typically exhibits the low energy utilization efficiency and longer scan times, facing limitations in miniaturized and high throughput portable spectral imaging technique [6].

In recent years, computational spectral imaging methods receive more and more attentions form researchers due to its advantages on the high energy utilization and working efficiencies over traditional spectral imaging methods [7], [8], [9]. Coded aperture snapshot spectral imagers (CASSI) as one of the computational spectral imaging techniques

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Zuo.

employ a mask-encoded aperture and dispersion elements to obtain spectral aliasing information, and thus have a higher energy utilization in comparison with traditional slit-grating based spectral imager. However, effective and accurate compressed sensing algorithms are critical for CASSI to reconstruct multi-spectral images quickly and easily [7], [8]. The diffractive optical imaging spectroscopy (DOIS) as another computational spectral imaging technique has also been practically verified in various applications such as gas detection [10], military identification [11], and gas leakage [12]. Different from the CASSI system using coded aperture and dispersive elements like prisms and gratings, the DOIS employs diffractive lenses and takes advantage of its unique axial dispersion property to acquire images of different wavelengths at different focusing distances by moving the focal plane imaging detector. Similarly, the image quality of the DOIS system is also determined by good image recovery algorithms [13]. Traditionally, iterative algorithms based on inverse filtering method are used, but they suffer from long iteration times and produce poor recovery results. More recently, researchers reported the application of neural network algorithms in spectral image reconstruction that can achieve a higher spectral resolution with better spatial imaging quality as well [14], [15], [16]. These algorithms typically include four groups: deep networks unfold the iterative process into multiple stages of the deep network; joint networks simultaneously learn mask coding and reconstruction processes, and untrained networks use a wide range of neural network as priors and end-to-end networks to direct reconstruct aliasing spectral images. In DOIS, spectral features propagated by point spread functions require spectral reconstruction to recover the convolution results of large wavelength-dependent convolution kernels. This causes a great mathematic challenge because the reconstruction of DOIS is a pathological deconvolution problem. Traditional spectral reconstruction algorithms based on optimization algorithms, utilize the image prior to constrain the solution space and exhibit mutual constraints on spectral and spatial resolution when solving the reconstruction problem of DOIS [17], [18]. The accuracy and speed of the image restoration have been greatly improved via deep learning methods [14], [15], [19]. However, these methods still apply the two-dimensional convolution to fabricate the backbone of neural network computation, resulting in the underutilization of information between spectral channels and low energy utilization.

In order to resolve this problem, we propose a new high accurate and efficient 3D Network "HAS3D" for image restoration to enhance the utilization of cross-information between spectral channels in image restoration by 3D convolution. The algorithm employs a 3D U-Net structure that the encoding and decoding paths are composed of 3D convolutional blocks. Numerical experiments are conducted to validate its effectiveness. The experimental results show that the mean peak signal-to-noise ratio (MPSNR) of the recovered image relative to the original image can be improved by 1.8 dB in comparison to other traditional methods. The new method possesses a mean structural similarity (MSSIM) of around 0.91, which meets the standard of indistinguishableness to human eyes. Moreover, the proposed algorithm runs in just 0.36 s that is faster than other traditional methods. The effectiveness of 3D convolution is also verified and compared with the SCA in our HAS3D.

II. FORWARD MODEL

Fig. 1(a) shows the schematic of the diffractive spectral imaging system that is typically consisted of a diffractive lens and a movable sensor. The optical field from the object such as a parrot is dispersed by the diffractive lens along the optical axis (z-axis), where the focal length f of the diffractive lens is inversely proportional to the wavelength as $f(\lambda) = \lambda_0 f_0/\lambda$ with λ_0 and f_0 is the designed wavelength and focal length respectively. Considering the wavelength-dependent behavior of the diffractive lens, the multispectral light field is divided into slices at different focal positions of the measurement end. By moving the position of the sensor, different slices of the multispectral light field are measured. Typically, the forward model is utilized to emulate the process of the diffractive multispectral imaging. As is shown in Fig. 1(a), the discretized true spectral distribution $\mathrm{I}(x, y, \lambda)$ corresponding to the object is convolved with the point spread function (PSF) $\mathrm{H}(x, y, \lambda)$ of the diffractive lens at different wavelength channels separately. Then, the simulated result $\mathrm{Y_c}(x, y)$ of a certain position can be obtained by summing these convolved results of different wavelength channels and adding gaussian white noise η. Calculating the PSFs of DOIS is essential for high accurate deconvolution. In this work, the diffraction model of DOIS is applied to simulate the imaging process. According to the classic scalar diffraction theory [19], [20], the function of the incident light field P can be expressed by (1),

$$\mathrm{P}(s, t, \lambda) = \mathrm{A}(s, t)e^{ik(n(\lambda)-1)\mathrm{h}(s,t)} \tag{1}$$

where $\mathrm{A}(s, t)$ represents the function of the aperture, $\mathrm{h}(s, t)$ is the function of the surface height of the diffractive lens, (s, t) is the spatial coordinates in the plane of the diffractive lens, λ and k donate the wavelength and wave number respectively. The axial PSF of diffractive lens can be determined by (2),

$$\mathrm{H}(x, y, \lambda) = \left| \frac{1}{\lambda f} e^{\frac{ik}{2f}(x^2+y^2)} \int\int \mathrm{P}(s, t, \lambda) e^{\frac{ik}{2f}(s^2+t^2) - \frac{ik}{f}(xs+yt)} dsdt \right|^2 \tag{2}$$

where here (x, y) is the spatial coordinate in the plane of the image sensor, f donates the distance between the lens and the sensor, and $n(\lambda)$ is the refractive index of the material that varies with the wavelength [21], [22]. Then, the image information at the sensor $\mathrm{Y_c}$ can be calculated by (3),

$$\mathrm{Y_c}(x, y) = \int_{\lambda_{\min}}^{\lambda_{\max}} [\mathrm{H}(x, y, \lambda) \star \mathrm{I}(x, y, \lambda)] d\lambda + \eta \tag{3}$$

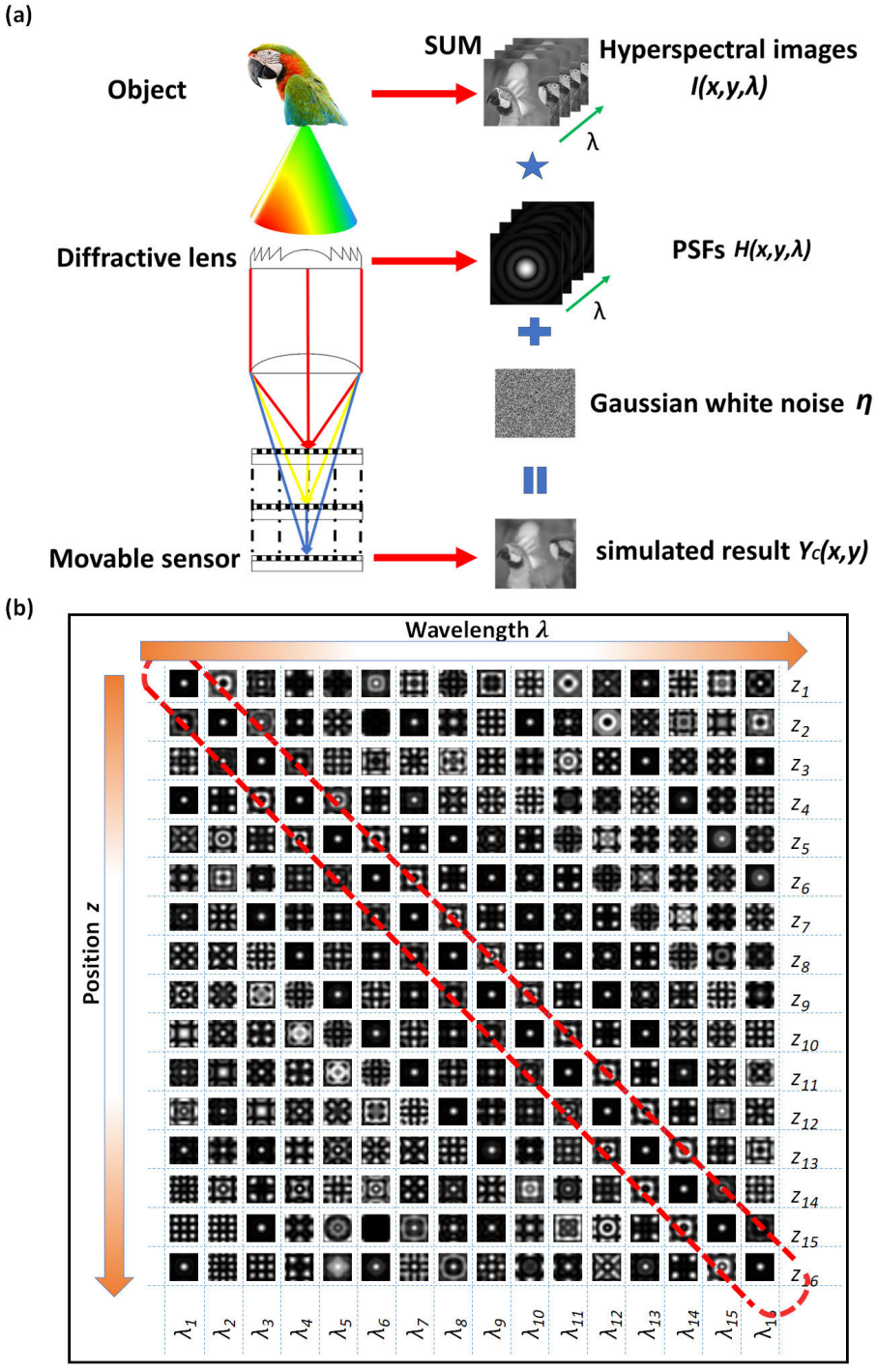

FIGURE 1. (a) The schematic of a typically diffractive spectral imaging system and its physical process of the multispectral imaging (b) Theoretical PSFs(x, y, z_i, λ_j) of 16 different positions for $z_1 \sim z_{16}$ and wavelength $\lambda_1 \sim \lambda_{16}$, where each subgraph represents a 2D PSF in x and y direction. The circled red dashed box exhibits the Airy spot of PSFs(x, y, z_i, λ_j) at diagonal position $i = j$.

where $[\lambda_{min}, \lambda_{max}]$ denotes the spectral range and η denotes the sensor noise that is simulated by Gaussian noise $\eta \sim N(0, \sigma^2)$. Discretizing (3) and denoting the pixel size by Δ, then the optical transfer function can be obtained and expressed by (4),

$$\mathrm{H}(x, y, \lambda) = \sum_{m,n} \mathrm{H}(m, n, \lambda)\, \mathrm{rect}\left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n\right) + \eta \quad (4)$$

where m and n donate the discretizing position of the center of each pixel in x and y direction. Finally, the signal g in each pixel can be obtained by sampling the accumulation of the image information of each spectral segment and expressed by (5),

$$\begin{aligned} &\mathrm{g}(m,n) \\ &= \int\int \mathrm{Y_c}(x,y)\,\mathrm{rect}\left(\frac{x}{\Delta}-m,\frac{y}{\Delta}-n\right)dxdy+\eta \\ &= \int\int \mathrm{rect}\left(\frac{x}{\Delta}-m,\frac{y}{\Delta}-n\right)dxdy \\ &\quad \int_{\lambda_{\min}}^{\lambda_{\max}} \sum_{m'n'} \mathrm{I}(x,y,\lambda) \\ &\quad * \mathrm{PSF}\left(m',n',\lambda\right)\mathrm{rect}\left(\frac{x}{\Delta}-m',\frac{y}{\Delta}-n'\right)d\lambda\Big]+\eta \end{aligned} \tag{5}$$

when the image applied for simulation is discrete, (5) can be simplified by using the intensity $\mathrm{I}(\xi,\zeta,\lambda)$ at the central pixel that represents the light intensity near the (x, y) position of the original image approximately, and can be calculated by,

$$\mathrm{g}(m,n)=\sum_{\lambda}\sum_{m,n}\mathrm{I}(\xi,\zeta,\lambda)\mathrm{H}(m-\xi,n-\zeta,\lambda)+\eta \tag{6}$$

the PSFs $\left(x,y,z,\lambda\right)$ can be calculated for the waveband ranging from 400 to 760 nm at full space with 360 sampled imaging positions according to (2) and (4). Fig. 1(b) shows the calculated $\mathrm{PSF_{cal}}(x,y,z_i,\lambda_j)$ at 16 different positions for z_i (i = 1-16) and wavelength λ_j (j = 1-16), where each wavelength λ_j is associated with a diffractive specific imaging position z_i.

As illustrated by the red dashed circular in Fig. 1(b), the PSFs $\left(x,y,z,\lambda\right)$ at diagonal positions with $i=j$ appears as a converging Airy spot. Otherwise, the PSFs exhibits as intricate out-of-focus diffraction patterns. For a given wavelength λ_i, the concentrated converging Airy spot emerges only at the corresponding focusing position z_i, and the energy is divergent with increasing distance, forming various diffuse diffraction patterns. Representing the $\mathrm{PSF_{cal}}$ as a matrix form $\mathrm{PSF_{mat}}$, $\mathrm{PSF_{mat}}$ is traditionally treated as a block circulant matrix according to the principle of spatial translation invariance. However, Fig. 1(b) reveals that the sub-matrix rows do not follow a simple circulation pattern in z direction. In accordance with (6), the simulation results can be computed by using the approximation of the block loop matrix, which shows a difference in MPSNR of 19.33 dB compared with the result of the simulation using the real PSFs matrix. These above results indicate the approximation hinders the image restoration with a high accuracy. Therefore, the 3D convolution is used to extract features of the cyclic matrices and reserve the discreteness at the same time.

III. SPECTRAL IMAGE RECONSTRUCTION

According to the analysis results in the previous section, the imaging process of the diffractive lens in (6) is a convolution of a three-dimensional PSF function of the incident light field. Then, the diffractive imaging process can be expressed as $Y=A\otimes I+N$, where Y denotes $\mathrm{Y_C}$ (x, y), A denotes the $\mathrm{PSF}(x,y,\lambda)$, I denote $\mathrm{I}(\xi,\zeta,\lambda)$,and N denotes η. The process of the spectral reconstruction is a typically invert method that the incident light information I can be formulated by the captured image Y and the PSF function A as an optimized problem of (7) [21],

$$\hat{I}=\arg\min_I \frac{1}{2}\|y-AI\|_2^2+\lambda\mathrm{R}(I) \tag{7}$$

Traditionally, iterative methods with constraint terms such as TV regularization and l_2 regularization are commonlly used to achieve convergence and obtain the optimal solution. However, methods with iteration are time-consuming and greatly rely on the selection of appropriate priori factors. When chosen priori factors deviate from the real data distribution, severe distortions in the recovery results could be caused. Inspired from the learning properties of neural networks, convolutional neural networks (CNNs) have been shown to have a great potential to improve the quality of the reconstruction of the diffractive spectral imaging [17], [23]. Particularly, Jeon et al. developed a U-net based end-to-end network to enhance the accuracy and speed of the derived imaging recovery, making it more suitable for practical applications [23]. Similarly, Oktem et al. demonstrated that the combination of model-based deep learning methods with traditional approaches can achieve significant performance improvements [14]. Aforementioned methods effectively utilize the feature extraction capabilities of neural networks and employ the simple algorithm design to invert complex function mappings. However, their networks adopt 2D convolution and leave out the learning of inter-channel information between spectra, causing the wastage of acquired spectral energy and inter-channel spectral information. To resolve this issue, we propose a method that combines 3D convolution with inter-channel attention to extract the spatial-spectral structure of the captured image data. Moreover, it is important to note that utilizing 3D convolution typically requires higher computational power and larger data storage. In order to address this problem, a baseline for 3D image reconstruction network is also proposed by the reported 2D reconstruction network [24].

Fig. 2(a) shows the U-net structure employed in our design [25]. This structure made by extended modules possesses strong capabilities in integrating contextual information and processing detailed information, and requires fewer parameters and computational resources, which could be a popular choice for spectral reconstruction problems. PSFs in Fig. 1(b) shows the similarity of internal structure and the features of adjacent local regions as mentioned in previous section. The 3D convolution template could act on different pixels (m, n) and wavelengths λ_j in the neighborhood simultaneously, and extract the 3D context features more directly and efficiently. The aliasing diffraction spectral image at z_j is constituted by a combination of the image of focused wavelength λ_j and images of other wavelengths λ_k

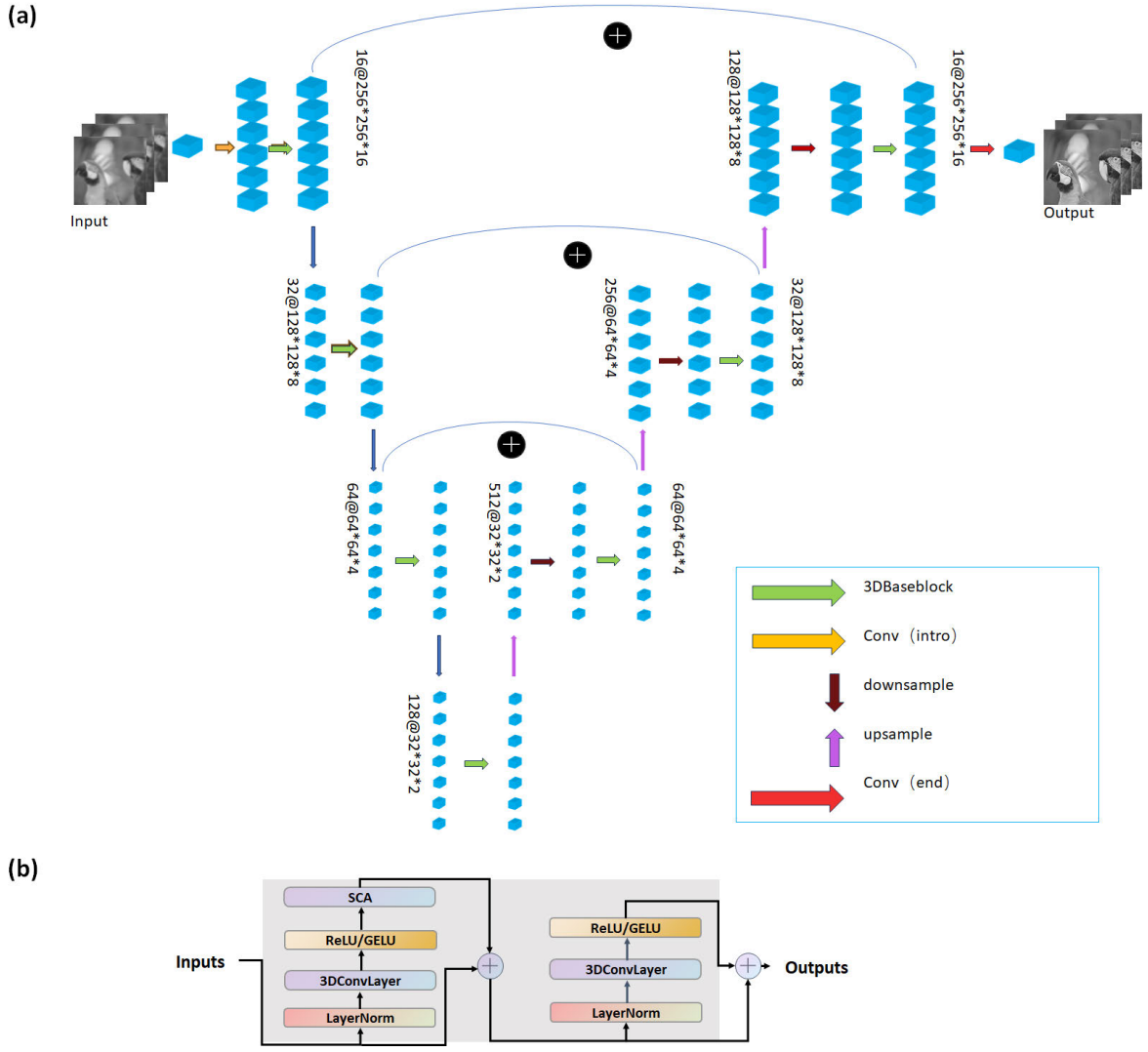

FIGURE 2. (a) Structure of the U-net network and (b) its 3D block.

(k = 1 …, j−1, j+1 …,16). The influence of the nearest wavelengths λ_{j-1} and λ_{j+1} on the aliasing image at z_j is particularly pronounced. It is a great challenge for characterizing multispectral relationship through interchannel attention methods and two-dimensional convolution. In contrast, 3D convolution offers a straightforward and precise means of characterizing this feature. From Fig. 2(b), 3D-Block based on the baseline is designed for the reconstruction of 3D data cubes. Additionally, to overcome the limitation of the small receptive field of the convolutional network, the inter-channel attention module is introduced to obtain the spectral channel correlation. The inter-channel attention module adjusts the inter-channel weights of the spectral channels to enhance the expression capability of features. The 3D convolution layer is used to capture the characteristics of the spectral dimension, and the simplified channel attention (SCA) layer is used to adjust the channel weight of the spectral dimension. The SCA enables cross-channel attention extraction, and reduces computational complexity. The 3D convolution dominates the learning of three-dimensional feature while SCA emphasizes the relationship between spectral steps, so that these two steps can complement each other. The layer normalization is applied to resolve the problem of small batchsize division arising from the small number of data volume groups and the large memory occupancy of each group. Furthermore, the layer normalization within each group of 3D data cubes is normalized to preserve the differences between the data. Considering the nonlinear function relationship between wavelengths and positions factor of PSFs, ReLU and GELU are employed to accelerate the initial learning process and improve the learning quality in the later stage respectively [26], [27].

IV. RESULTS AND DISCUSSIONS

We implemented python programs to execute the aforementioned network structure. All experiments were performed on a server that had an Intel(R) Xeon(R) Platinum 8255C CPU and an RTX 3080 (10GB) GPU. The used operating system

TABLE 1. Comparison of recovery effects of different algorithms.

Method	Gap-tv	Autoencoder	ISTA-Net	Unrolled Network	**HAS3D (ours)**
MPSNR (dB)	22.77	28.22	33.37	35.88	**37.68**
MSSIM	0.17	0.72	0.88	**0.93**	0.91
Rebuilding time(s)	51.20	2.03	1.28	0.58	**0.36**
Training times(h)	×	**19.4**	23.6	32.9	26.5

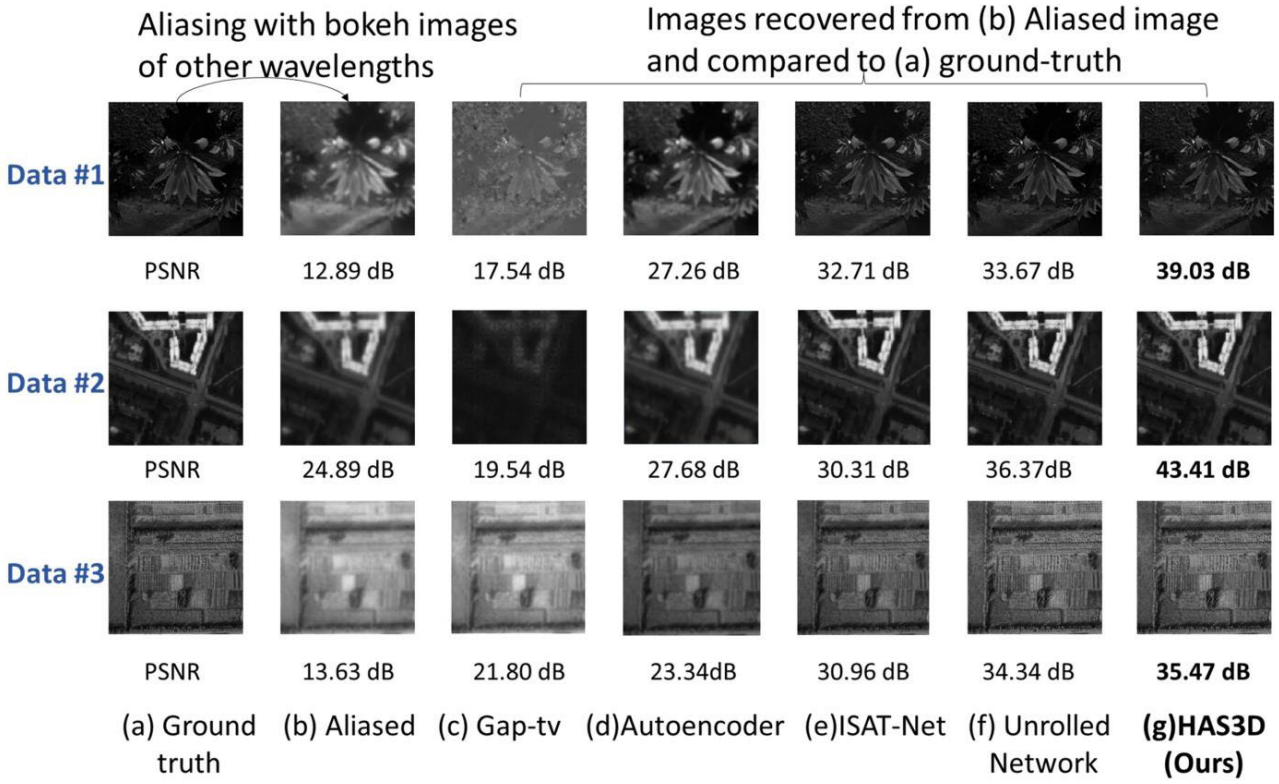

FIGURE 3. Comparison of space performance (PSNR) of single spectral image unmixing results using different algorithm. (a) ground-truth for reference, (b) aliased image influenced by other wavelengths, (c) image recovered by gap-tv algorithm, (d) image recovered by autoencoder algorithm (e) image recovered by ISTA-Net algorithm (f) image recovered by Unrolled network algorithm (g) image recovered by our HAS3D algorithm.

was Ubuntu 20.04.4 LTS. The model was developed by using the pytorch version 1.11.0 framework, and was built on the open-source BasicSR library [28].

The training data was obtained by simulating the model and selecting a number of publicly available spectral datasets as the original spectral data [29], [30]. The simulated 3D signal data was obtained by using the forward model simulation mentioned earlier, resulting in 1,000 sets of 3D data cubes. The ratio of the training set, test set, and validation set was 6:2:2.

In order to prevent the problem of convergence difficulties of the 3D network, the 2D network was firstly trained by using the same structure as shown in Fig. 2. The 3D block was then converted to 2D, and the parameters of the 2D convolution was then imported into 3D along the spectral dimension, where these training results of 2D network was used as the pre-trained network parameter inputs. Model optimization was performed by using the Pytorch framework and the AdamW solver [31]. All convolutional layers except the 1×1 convolutional layer used a 3×3 filter with zero padding to keep the feature map constant. The whole network contained around 589.1×10^6 parameters, and took 26.5 hrs. to train.

Two parameters of MPSNR and MSSIM are used the main evaluation metrics to measure the effectiveness of image restoration [32]. MPSNR measures the similarity between the restored image and the original image by comparing the pixel difference, where the higher value indicates a higher similarity. Typically, a restored image with a MPSNR over 30 dB is considered acceptable for human eye recognition. On the other hand, MSSIM characterizes the influence of brightness, contrast, and structure, which is in consistent with the subjective perception of image quality from human eye. The MSSIM value commonly ranges from 0 to 1,

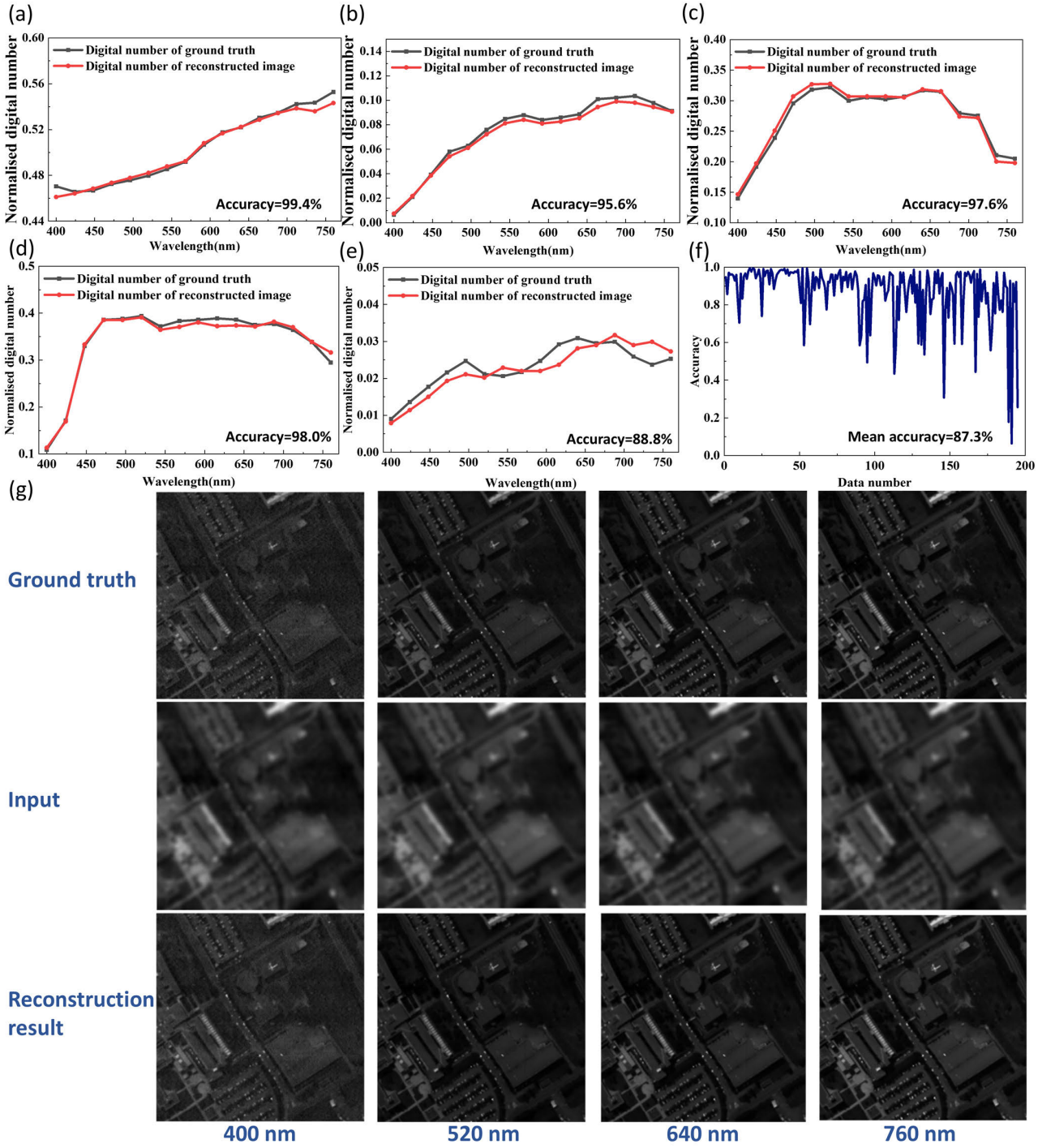

FIGURE 4. Comparison of spectral performance of multispectral image unmixing results. Reconstructed spectral data curve and accuracy (a)-(e) are spectral curves extracted from different data, (f) is the average of the accuracy of the spectral curves for all data. Red curve is the target spectral curve extracted from ground-truth. Black curve is the spectral curve recovered by our algorithm. (g) is a comparison of ground truth, network input and reconstructed outputs at different wavelengths.

where 1 denotes identical images and 0 signifies completely different images. As shown in Table 1, the recovery effect of our HAS3D network is also tested and compared with typical Gap-tv, Autoencoder, ISTA-Net and Unrolled network on various datasets. The HAS3D network exhibits a MPSNR of around 37.68 dB, showing an improvement of 1.8 dB in comparison with the highest MPSNR of typical Unrolled Network (around 35.88 dB). The MSSIM of our designed network is around 0.91 that meets the standard of indistinguishableness to human eyes (MSSIM>0.9). However, in terms of improving MPSNR, the improvement in MSSIM is only marginal, possibly due to the network's

oversimplified structure lacking depth in processing image information [33]. The execution time of the designed program for a single run only takes 0.36 s that is faster than above algorithms in Table 1, showing the time efficiency is improved for around 37.9% in comparison with that of the typical Unrolled Network (0. 58 s). The rebuilding time t_r of the proposed HAS3D is around 0.36 s that is faster than other algorithms. Furthermore, the proposed HAS3D needs a shorter training time of $t_t \sim 26.5$ h than that of the Unrolled Network ($t_t \sim 32.9$ h) but longer than that of the Autoencoder ($t_t \sim 19.4$ h) and ISTA-Net ($t_t \sim 23.6$ h). However, the Autoencoder and ISTA-Net need longer rebuilding time of around 2.03 s and 1.28 s respectively. In practice, different spectral image systems only need to train once to learn parameters of the model but have to take the rebuilding time every time for different scenes to obtain multispectral images, and thus shorter rebuilding time is apparently more important for a good algorithm for the multispectral imaging application. As a result, the proposed HAS3D shows a faster rebuilding speed and an acceptable training speed in comparison with other traditional algorithms. In view of this, we believe the proposed HAS3D has a great potential to be an alternative algorithm for the diffractive multispectral imaging under different scenes.

The restoration results of computing spectral imaging are also analyzed in spatial and spectral dimension as shown in Fig. 3 and Fig. 4, and are carefully compared with the other traditional methods. From the reconstruction results of the first set of data (Data #1, first row) in Fig. 3, the aliased image exhibits a MPSNR around 12.89 dB in contrast to the ground truth image, and the restoration results calculated by gap-tv, autoencoder, ISTA-Net, Unrolled network and HAS3D algorithm, whose MPSNR is around 17.58 dB, 27.68 dB, 32,71 dB, 33.67 dB and 39.03 dB respectively, showing our 3D convolutional network could efficiently rebuild aliased spectral images with the highest PSNR 39.03 dB. Similarly, the second (Data #2, second row) and third set of data ((Data #3 third row) also show the highest value of MPSNR 43.41 dB and 35.47 dB that can be achieved in comparison with other networks, confirming our HAS3D could efficiently rebuild aliased spectral images in various scenes and may provide a useful method for highly accurate reconstruction of diffractive-based computational spectral imaging. The spectral curves of the recovered images are further extracted and compared with the original image. From Figs.4(a)-(b), the accuracy of the reconstructed spectral data curve of five different original image is 97.3%, 99.4%, 97.4%, 98.1% and 92.0% respectively, and the average accuracy in Fig. 4(f) is around 86.7%, showing a high accuracy of the reconstructed method. The reconstruction multispectral results for wavelengths of 400 nm, 520 nm,640 nm and 760 nm are given in Fig. 4(g). It is obvious that the reconstructed multispectral images in the last row are clearer than the input images and are similar with the ground truth images, which confirms the accuracy of our network in the computational spectral imaging.

TABLE 2. The effectiveness of 3D convolution and Simplified Channel Attention (SCA) have been verified.

	2D→3D Block	SCA	PSNR	SSIM	Rebuilding time(s)
HAS3D	√	√	37.68	0.91	0.36
	√	×	35.70	0.90	0.36
Pre-trained 2D	×	√	29.70	0.89	0.31
	×	×	29.29	0.88	0.30

The effectiveness of the 3D convolution and SCA is also verified by ablation experiments in our HAS3D. As illustrated in Table 2, simulated PSNR from HAS3D algorithm with 3D convolution and SCA modules is around 37.68 dB, while the simulated of PSNR from HAS3D algorithm only with 3D convolution is around 35.70 dB, confirming SCA is more effective in integrating adjacent band information. Similarly, HAS3D with 3D convolution and SCA modules has a higher $PSNR_{3D\&SCA}$ of around 37.68 dB. In comparison with that of the pre-trained 2D algorithm only with SCA modules (around 29.70 dB), the 3D convolution plays a critical role in improving the PSNR. On the other hand, simulated SSIM and times of HAS3D with 3D convolution and SCA modules are similar to HAS3D of 3D convolution, which is around 0.9 and 0.36 respectively and is in accordance with the SSIM and times of Pre-trained 2D algorithm. Ablation experiment confirms the important role of the inter-spectral communication in improving the quality of spectral reconstruction.

V. CONCLUSION

In summary, we propose a highly accurate and efficient 3D network with encoding and decoding paths from 3D convolutional blocks for spectral image restoration of the DOIS system. Our algorithm shows significant improvements in the increase of PSNR$\sim$ 1.8 dB and the enhancement speed around 37.9% at the same time. The shallow network with a small convolutional kernel is applied to enhance speed, and its MSSIM of 0.91 meets the standard of indistinguishableness to human eyes, and meanwhile the algorithm runs in just 0.36 s that is faster than other traditional methods. However, the shallow network with a small convolutional kernel exhibits the indistinct improvement of MSSIM, which may attribute to the oversimplified structure lacking depth of network. Of course, the limitation of MSSIM should be addressed in future research. Simulated experiments confirm the utilization of 3D convolution can effectively integrate information from different spectral bands and play a critical role in optimizing the spectral image recovery. Furthermore, the designed algorithm can be physically validated by applying it to a diffraction spectral imager. The application of our 3D network algorithm to diffraction lens spectroscopy may benefit more areas such as Chemical Identification, medical diagnosis and food monitoring in future [34], [35], [36].

REFERENCES

[1] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, Jan. 2014, Art. no. 010901.
[2] W. S. Z. Lifu, Z. Yan, Y. Deshuai, S. Ruoxi, Q. Wenchao, Q. U. Liang, L. U. Zhiyong, and T. Qingxi, "Progress of hyperspectral remote sensing applications on cultural relics protection," *Acta Geodaetica Et Cartographica Sinica*, vol. 52, no. 7, pp. 1126–1138, 2023.
[3] Y. Liu, H. Pu, and D.-W. Sun, "Hyperspectral imaging technique for evaluating food quality and safety during various processes: A review of recent applications," *Trends Food Sci. Technol.*, vol. 69, pp. 25–35, Nov. 2017.
[4] B. Lu, P. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, p. 2659, Aug. 2020.
[5] C. Fan, "Spectral analysis of water reflectance for hyperspectral remote sensing of water quailty in estuarine water," *J. Geosci. Environ. Protection*, vol. 2, no. 2, pp. 19–27, 2014.
[6] Z. Yang, T. Albrow-Owen, W. Cai, and T. Hasan, "Miniaturization of optical spectrometers," *Science*, vol. 371, no. 6528, Jan. 2021, Art. no. eabe0722.
[7] L. Huang, R. Luo, X. Liu, and X. Hao, "Spectral imaging with deep learning," *Light, Sci. Appl.*, vol. 11, no. 1, p. 61, Mar. 2022.
[8] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 65–88, Mar. 2021.
[9] J. Bacca, E. Martinez, and H. Arguello, "Computational spectral imaging: A contemporary overview," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 40, no. 4, p. C115, 2023.
[10] M. Hinnrichs, "Infrared hyperspectral imaging sensor for gas detection," in *SPIE Proc.*, Nov. 2000, p. 12.
[11] M. Hinnrichs, "Remote sensing for gas plume monitoring using state-of-the-art infrared hyperspectral imaging," in *SPIE Proc.*, Feb. 1999, p. 12.
[12] M. Hinnrichs, "Imaging spectrometer for fugitive gas leak detection," in *Proc. Photon. East '99*, 1999, p. 10.
[13] K. Whitcomb, D. Lyons, and S. Hartnett, *DOIS: A Diffractive Optic Image Spectrometer (Aerospace/Defense Sensing and Controls)*. Bellingham, WA, USA: SPIE, 1996.
[14] F. S. Oktem, O. F. Kar, C. D. Bezek, and F. Kamalabadi, "High-resolution multi-spectral imaging with diffractive lenses and learned reconstruction," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 489–504, 2021.
[15] Y. Peng, Q. Fu, F. Heide, and W. Heidrich, "The diffractive achromat full spectrum computational imaging with diffractive optics," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–11, Jul. 2016.
[16] C. G. Z. Ming-Qi, C. Qiang, and S. Quan-Sen, "Image restoration method based on improved inverse filtering for diffractive optic imaging spectrometer," *Comput. Sci.*, vol. 46, no. 1, pp. 86–93, 2019.
[17] W. Xu, C. Qiang, and S. Quansen, "Multichannel spectral–spatial total variation model for diffractive spectral image restoration," *J. Comput. Res. Develop.*, vol. 57, p. 413, Aug. 2020.
[18] Z. Qiao, X. Wen, X. Zhou, F. Qin, S. Liu, B. Gao, W. Liu, D. Chi, and Z. Liu, "Adaptive iterative guided filtering for suppressing background noise in ptychographical imaging," *Opt. Lasers Eng.*, vol. 160, Jan. 2023, Art. no. 107233.
[19] M. A. X. Born and E. Wolf, *Chapter VIII-Elements of the Theory of Diffraction*, 6th ed., New York, NY, USA: Pergamon, 1980, pp. 370–458.
[20] J. W. Goodman, *Introduction to Fourier Optics*, 4th ed. W.H. Freeman, 2017.
[21] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering* (Fundamentals of Algorithms). Society for Industrial and Applied Mathematics, 2006, pp. 90–93.
[22] P. B. M. Bertero, *Christine De Mol, Introduction to Inverse Problems in Imaging*. Boca Raton, FL, USA: CRC Press, 2021.
[23] D. S. Jeon, S.-H. Baek, S. Yi, Q. Fu, X. Dun, W. Heidrich, and M. H. Kim, "Compact snapshot hyperspectral imaging with diffracted rotation," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–13, Aug. 2019.
[24] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 17–33.
[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2015, pp. 234–241.
[26] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," presented at the *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 1–19.
[27] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.
[28] X. Wang, L. Xie, K. Yu, K. C. K. Chan, C. C. Loy, and C. Dong, *BasicSR: Open Source Image and Video Restoration Toolbox*. Github, 2022.
[29] Y. Cena L. F. Zhang, X. Zhang, Y. M. Wang, W. C. Qi, S. L. Tang, and P. Zhang, "Aerial hyperspectral remote sensing classification dataset of Xiongan new area (matiwan village)," *J. Remote Sens. Chin.*, vols. 11–24, pp. 1299–1306, Jun. 2020.
[30] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 19–34.
[31] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–22.
[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[33] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2015, pp. 730–734.
[34] A. Bodkin, A. Sheinis, A. Norton, J. Daly, C. Roberts, S. Beaven, and J. Weinheimer, "Video-rate chemical identification and visualization with snapshot hyperspectral imaging," in *SPIE Proc.*, May 2012, pp. 1–24.
[35] X. Zhao, J. Chang, Y. Hu, S. Du, and D. Li, "Optical design of a snapshot hyperspectral imaging spectrometer for facial colour diagnosis in traditional Chinese medicine," in *Proc. 7th Symp. Novel Photoelectronic Detection Technol. Appl.*, Mar. 2021, pp. 1–28.
[36] M. Al-Sarayreh, M. M. Reis, W. Q. Yan, and R. Klette, "Potential of deep learning and snapshot hyperspectral imaging for classification of species in meat," *Food Control*, vol. 117, Nov. 2020, Art. no. 107332.

HAO FAN received the B.S. degree from the College of Optical Science and Engineering, Zhejiang University. He is currently pursuing the Ph.D. degree with Xi'an Institute of Optics and Precision Mechanics, University of Chinese Academy of Sciences. His research interests include diffractive optics, spectroscopy, and computational imaging.

CHENXI LI, photograph and biography not available at the time of publication.

HUANGRONG XU, photograph and biography not available at the time of publication.

LVRONG ZHAO, photograph and biography not available at the time of publication.

XUMING ZHANG, photograph and biography not available at the time of publication.

HENG JIANG, photograph and biography not available at the time of publication.

WEIXING YU received the B.S. degree from Northwestern Polytechnical University, in 1998, the M.Sc. degree from Changchun Institute of Optics, Fine Mechanics, and Physics, in 2001, and the Ph.D. degree from Nanyang Technological University, in 2005. Currently, he is working as a Full Professor with Xi'an Institute of Optics and Precision Mechanics (XIOPM). Before, he joined XIOPM, he also worked as a Full Professor at Shenzhen University for one year and Changchun Institute of Optics, Fine Mechanics, and Physics for five years. He once worked as a Postdoctoral at Heriot-Watt University for three years, from 2006 to 2009. His research interests include hyperspectral imaging based on micro/nano optics, bionic optical imaging, micro-optics, and nanophotonic.