Research Article Vol. 32, No. 8 / 8 Apr 2024 / Optics Express 13918
Optics EXPRESS

Deep learning-enhanced snapshot hyperspectral confocal microscopy imaging system

SHUAI LIU,[1,†] WENZHEN ZOU,[2,†] HAO SHA,[2] XIAOCHEN FENG,[2] BIN CHEN,[3] JIAN ZHANG,[3] SANYANG HAN,[1] XIU LI,[1] AND YONGBING ZHANG[2,*]

[1]*Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China*
[2]*School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, Guangdong 518055, China*
[3]*School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen, Guangdong 518055, China*
[†]These authors contributed equally to this work
[*]*ybzhang08@hit.edu.cn*

Abstract: Laser-scanning confocal hyperspectral microscopy is a powerful technique to identify the different sample constituents and their spatial distribution in three-dimensional (3D). However, it suffers from low imaging speed because of the mechanical scanning methods. To overcome this challenge, we propose a snapshot hyperspectral confocal microscopy imaging system (SHCMS). It combined coded illumination microscopy based on a digital micromirror device (DMD) with a snapshot hyperspectral confocal neural network (SHCNet) to realize single-shot confocal hyperspectral imaging. With SHCMS, high-contrast 160-bands confocal hyperspectral images of potato tuber autofluorescence can be collected by only single-shot, which is almost 5 times improvement in the number of spectral channels than previously reported methods. Moreover, our approach can efficiently record hyperspectral volumetric imaging due to the optical sectioning capability. This fast high-resolution hyperspectral imaging method may pave the way for real-time highly multiplexed biological imaging.

1. Introduction

The spectral information of dyes or autofluorescence in cells is important for quantifying and analyzing the components within the sample [1–8]. Hyperspectral confocal microscopy is an advanced microscopy technique that combines the principles of confocal microscopy and hyperspectral imaging [9,10]. Compared to wide-field microscopy, hyperspectral confocal microscopy not only enables the detection of spectral information but also possesses enhanced optical sectioning capability. The hyperspectral laser-scanning confocal microscopy system typically consists of a confocal microscope equipped with a diffraction grating or prism. The fluorescence of the sample expands into a three-dimensional hyperspectral cube after passing through the dispersive element, which cannot be directly accessed by a two-dimensional image sensor. Therefore, the traditional method achieves the acquisition of hyperspectral images by scanning. However, the imaging speed of laser-scanning confocal microscopy is slow since it adopts the point scanning manner [11–13]. Acquisition speed can be increased by optimizing the illumination scheme like simultaneously scanning multiple points or lines [14,15]. Although line-scanning confocal microscopy has improved the temporal resolution to some extent, but it still belongs to the scanning imaging mode of using time to exchange spectral information and lead to a big data size to be stored, transmitted, and processed [16].

#519045 https://doi.org/10.1364/OE.519045
Journal © 2024 Received 15 Jan 2024; revised 3 Mar 2024; accepted 17 Mar 2024; published 1 Apr 2024

Optics EXPRESS

Code Aperture Snapshot Spectral Imaging (CASSI) based on compressive sensing (CS) [17,18] is a novel spectroscopic imaging technique that captures a compressive measurement in a snapshot and recovers the 3D hyperspectral cube by the reconstruction algorithm [19–24]. This spectral imaging approach, which utilizes coded aperture modulation instead of scanning, has great potential for applications in real-time scenes. However, the quality of CASSI reconstruction is poor when facing complex data cube reconstruction with hundreds of channels, because of the substantial compressive ratio. To overcome this obstacle, researchers have tried to reduce the compressive ratio, capturing multiple snapshots using different coded apertures or by utilizing non-diffractive images as references for reconstruction [25–27]. However, both methods either introduce extra exposures or increase the cost and complexity of the system. On the other hand, deep learning algorithms use convolutional neural networks to learn the mapping relations of hyperspectral datasets directly from coded compressive images, which can achieve good reconstruction results. Due to GPU acceleration, deep learning algorithms can achieve millisecond reconstruction time performance. However, the majority learning-based methods rely on an assumption of strong prior similarity between the training and testing data. The current deep learning-based CASSI algorithms typically utilize synthetic datasets to train networks [28–31]. In fact, the emission spectrum and central wavelength of samples are always different from the synthesized dataset, which may result in the degradation of the model and unreliable outputs. Another effective deep learning algorithm is the self-supervised network based on the image prior [32,33], but it requires an additional camera to acquire the reference image.

In this study, we realize snapshot hyperspectral confocal microscopic imaging by applying the CASSI to the confocal imaging system for the first time, to our knowledge. Our system employs a DMD-based microscope for paired hyperspectral compressive data acquisition and utilizes an interpretable deep unfolding neural network (SHCNet) for hyperspectral confocal reconstruction. The snapshot hyperspectral confocal microscopic imaging system (SHCMS) can generate different structures of illumination based on the programmable characteristics of the DMD, thus providing pairs of snapshot compressive measurement data and hyperspectral confocal images for SHCNet training. This guarantees that our trained network is more applicable to the spectral response of our hardware system, which makes the reconstruction more credible. The SHCNet is based on the generative adversarial network (GAN) framework that achieves a good balance of local detail enhancement and artifact suppression [34]. We carry out experiments on the autofluorescence of potato tuber to demonstrate the performance of our system. Our proposed method can reconstruct 160-bands hyperspectral confocal images, which achieves almost 5 times improvement in the number of spectral channels than previous reported methods. In summary, our system not only realizes high-resolution CASSI reconstruction at a high compressive ratio but also has significant potential for applications in real-time hyperspectral confocal microscopy imaging.

2. Method

2.1. Principle of SHCMS

An illustration of our experimental setup is shown in Fig. 1. Our prototype system can recover 160 spectral channels confocal images in the visible bandwidth from 409–677 nm using only a single-shot compressive measurement. The system consists of a DMD-based hyperspectral confocal microscopy and a deep learning-based reconstruction algorithm. DMD-based hyperspectral confocal microscopy can take advantage of the programmable characteristics of DMD to switch between snapshot hyperspectral imaging and line-scanning hyperspectral confocal imaging. Therefore, the SHCMS can capture snapshot compressive measurement data and hyperspectral confocal image data in pairs for network training. SHCNet is trained on the pairwise data collected by the SHCMS system rather than the synthetic data, enabling the model to consider the influence of noise distribution in real scenarios and learn the spectral response of SHCMS. This

Optics EXPRESS

provides more reliable reconstructions for microscopic hyperspectral imaging. Experimentally, we verify the snapshot hyperspectral confocal imaging performance of SHCMS with the potato tuber autofluorescence.

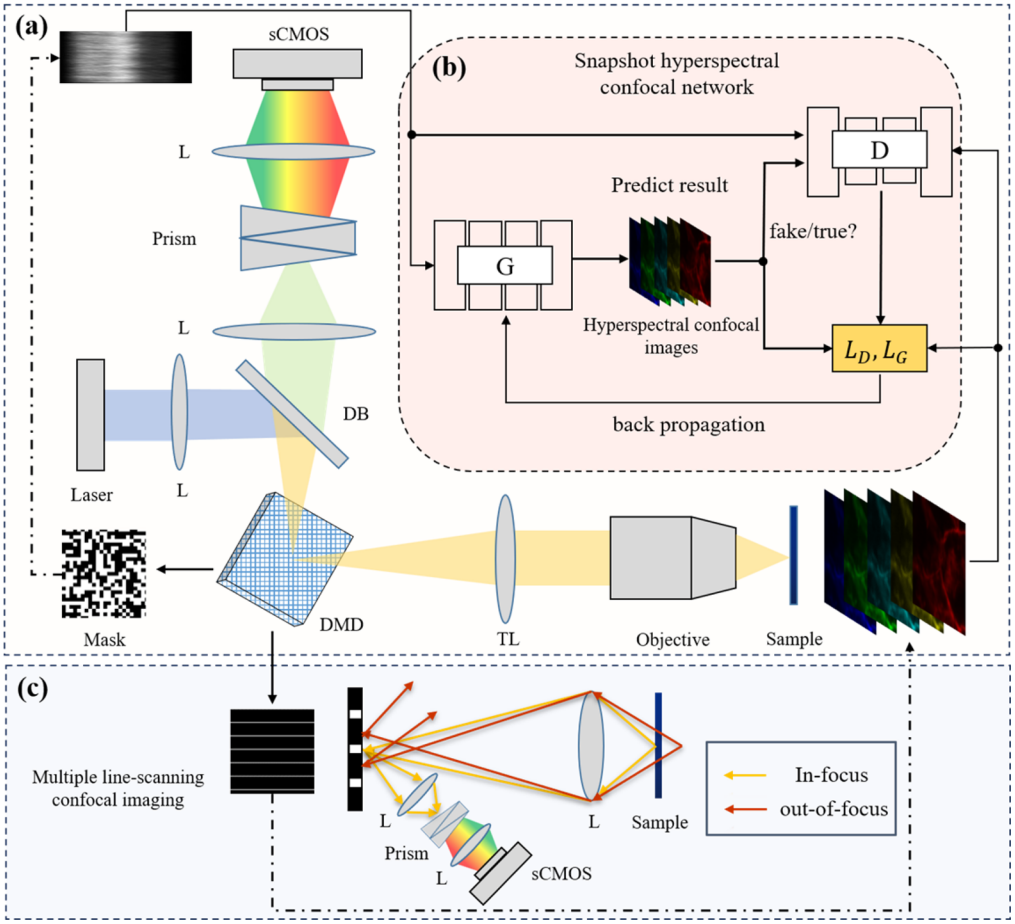

Fig. 1. Workflow of the SHCMS. Our system consists of (a) and (b). (a) Schematic of snapshot hyperspectral confocal microscopy. DMD as a coding mask in conjunction with a prism to perform the spatial and spectral modulations for compressive measurement. (b) The overall structure of the SHCNet. The network is capable of recovering 160 spectral channel confocal images. (c) Schematic of confocal imaging. Multi-line patterns are utilized to implement the pinhole function to obtain the hyperspectral confocal images, which act as the ground truth for SHCNet training. L, lens; DB, dichroic beamsplitter; TL, tube lens.

2.2. Hardware prototype implementation

Figure 1(a) shows the schematic illustration of our designed SHCMS. The power tunable laser, operating at 488 nm, is used as the light source. The laser beam is reflected by a dichroic beam splitter (DB) to the DMD (V7001 DLP7000&DLPC410), which modulates the laser with a random mask and then illuminates the sample through a tube lens and objective (Nikon 20X/0.75). The sample is selectively excited and emits fluorescence at the focal plane of the objective following the DMD modulation pattern. Subsequently, the fluorescence returns to the DMD along the excitation path and is dispersed by the prism (Shanghai Optics, custom-made) through

Optics EXPRESS

the DB. Finally, the encoded compressed fluorescence is received by the sCMOS (ORCA-Flash 4.0LT + Hamamatsu) through the 4-f system and used as an input to the SHCNet. On the other hand, when we load multiple line scan patterns on the DMD, the system becomes a parallelized line-scanning hyperspectral confocal microscopy, as shown in Fig. 1(c). The sample is selectively excited according to the multi-line pattern displayed by the DMD. The DMD only reflects the in-focus fluorescence in the white regions of the modulation patterns in Fig. 1 (c). Here, the DMD plays as a confocal pinhole because it is in the back focal plane of the objective lens. Thus, we can acquire the ground truth of hyperspectral confocal images for network training via this imaging system.

The calibration of the system consists of two main aspects, one is the calibration of the tilt angle between the DMD and sCMOS and the other is the spectral calibration. In consideration of the inherent reflective properties of the DMD, it becomes imperative to position the sCMOS orthogonally at a 24° angle relative to the optical axis. However, such positioning results in the generation of non-uniform imagery, predominantly attributable to the variances in the optical pathways induced by the DMD. To compensate this optical path difference reflected by the different DMD micromirrors, the sCMOS must be further tilted in the opposite direction by a Scheimpflug angle [35,36]. The detailed calibration method can refer to Ref. [37]. Next, we can estimate the DMD multiline scanning interval to be 160 columns based on the range of spectral shifts when the halogen lamp illuminates the DMD single column, as shown in Fig. 2(a). The scan interval at 160 columns means that there is no spectral crosstalk between adjacent scanning. Through spectral calibration experiments of a mercury lamp and a halogen lamp filtered by 460 and 655 nm band-pass filters (Chroma, ET460/36 m, and AT655/30 m), we found a nonlinear relationship between the wavelength λ and the number of sCMOS columns. The wavelength versus pixel shift calibration curve is shown in Fig. 2(b), and we used a two-order polynomial fitting model. Through this dispersion relation, we can map the spectral wavelengths to the pixel shift in the sCMOS. We also provide spatial and spectral resolution characterization in Supplement 1 Fig. S1.

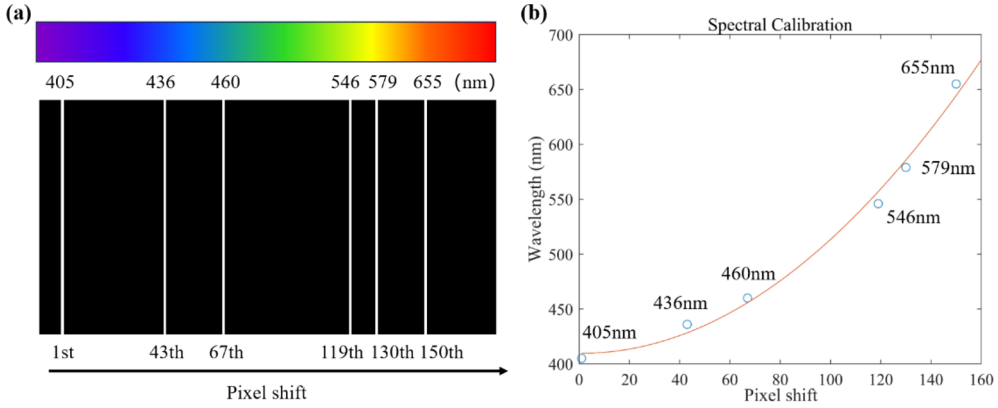

Fig. 2. Spectral calibration. (a) The spectrally dispersed image of a mercury lamp and a halogen lamp filtered by 460 and 655 nm band-pass filters caused by the deflection of the single column micromirrors. (b) The fitting result of the dispersive relation.

2.3. Reconstruction network

SHCNet is designed following the framework of GAN. It consists of a generator based on deep unfolding [38,39] and a discriminator based on U-Net, as shown in Fig. 1(b). The generator produces reconstructed results, while the discriminator assesses the authenticity of the output.

Optics EXPRESS

In generator, Deep unfolding network unrolls the iterative optimization-based reconstruction procedure into a neural network, as shown in Fig. 3. Each stage has the task of solving the iterative equations, which makes the neural network interpretable. Specifically, for a 128×287 input image, we sliced it through a sliding window with step size of 1, resulting in a total of 160 images of size 128×128, and finally acquired an initial data block of dimensions $128 \times 128 \times 160$. Subsequently, we compute results through a series of stages, where each stage comprises a gradient descent module and a proximal mapping module, corresponding to the two iterative steps in the Iterative Shrinkage-Thresholding Algorithm (ISTA) [40]. As for the discriminator, we employ a U-Net structure to discern the reconstructed results, which is advantageous for reconstructing details [34].

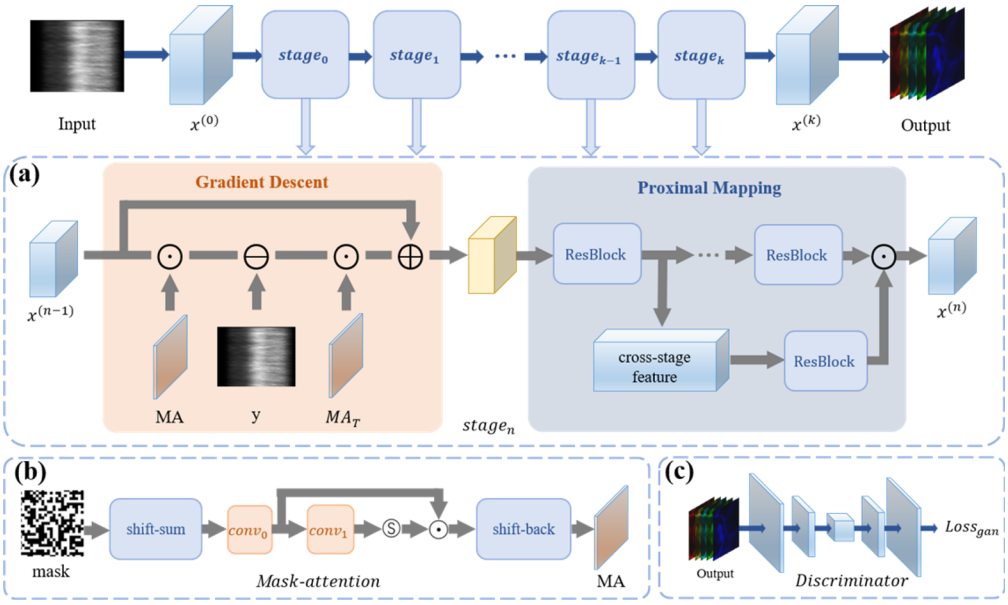

Fig. 3. The architecture of SHCNet. (a) The structure of an individual stage within iterations consists of a gradient descent module and a proximal mapping module. (b) Extracts matrix self-attention from the encoding matrix used for compression. The "S" function here refers to the sigmoid function. (c) Utilizes a U-Net-based discriminator to evaluate the reconstructed results.

Forward Model of CASSI. The snapshot hyperspectral confocal microscopy system is based on the coded aperture snapshot spectral imaging technique. The hyperspectral image can be seen as a three-dimensional data block $X \in R^{H \times W \times C}$, where H and W denote spatial dimensions and C represents the spectral dimension. The encoding process is performed using a mask $M \in \{0, 1\}^{H \times W}$, and the modulated hyperspectral image is expressed as follows:

$$X'(x, y, \lambda_i) = M(x, y) \odot X(x, y, \lambda_i), \tag{1}$$

where x, y denote spatial positions, $\odot$ is the Hadamard (element-wise) product, and λ_i represents the i-th wavelength channel. By employing a prism, the encoded spectral channels are dispersed into different spatial positions and then superimposed. The displacement distance is wavelength-dependent, denoted as $d(\lambda_i)$. Consequently, the relationship is defined as:

$$Y(x, y) = \sum_{i=1}^{C} X'(x + d(\lambda_i), y, \lambda_i). \tag{2}$$

Optics EXPRESS

By combining Eqs. (1) and (2), we obtain the complete imaging process of CASSI as follows:

$$Y(x,y) = \sum_{i=1}^{C} M(x + d(\lambda_i), y) \odot X(x + d(\lambda_i), y, \lambda_i) \,, \tag{3}$$

using vector x to represent X, vector y to represent Y, Φ as the imaging matrix, and n as noise, the expression can be written as follows:

$$y = \Phi x + n \tag{4}$$

Architecture of the SHCNet. The overall structure of SHCNet adopts the framework of GAN, we use an optimization-inspired deep unfolding network for image reconstruction, serving as the generator of the GAN network, while the discriminator is employed to distinguish between real and fake outputs.

The generator part employs a deep unfolding network. Drawing inspiration from the ISTA, we view image reconstruction as the following optimization problem:

$$x = \underset{x}{argmin} \frac{1}{2} \|y - \Phi x\|_2^2 + v\psi(x), \tag{5}$$

where the first term denotes the data fidelity term, constraining the reconstructed content, and the second term is the regularization term. The parameter v serves as the regularization coefficient to ensure generalization capability.

The traditional ISTA algorithm solves the data fidelity and regularization terms separately using gradient descent and proximal mapping as follows:

$$\begin{aligned} r^{(k)} &= x^{(k-1)} - \rho\Phi^T(\Phi x^{(k-1)} - y), \\ x^{(k)} &= \underset{x}{argmin} \frac{1}{2} \left\| x - r^{(k)} \right\|_2^2 + v\psi(x). \end{aligned} \tag{6}$$

Here, ρ represents the iteration step size, $r^{(k)}$ denotes the solution to the data fidelity term, while $x^{(k)}$ represents the solution to the regularization term obtained through proximal mapping. Based on these two steps, we designed a deep unfolding network composed of a gradient descent module based on mask self-attention and a proximal mapping module.

In gradients descent module, The ISTA-based deep unfolding networks currently employ fixed Φ and Φ^T in Eq. (6), limiting the flexibility of the network. To address this problem, another deep unfolding network called HerosNet has been proposed [41]. It replaces them with convolutions, but does not merge the spatial information from the mask. We aim to enhance network reconstruction performance by introducing spatial information from the mask while increasing the flexibility of the network's reconstruction capabilities. Consequently, we adopted a mask self-attention mechanism. The mask attention is computed with the mask self-attention module, denoted as MA. The specific structure is illustrated in Fig. 3(b). Firstly, following the principles of CASSI imaging, we transformed the mask into an equivalent M_r as follow:

$$M_r(x,y) = \sum_{i=1}^{C} M(x + d(\lambda_i), y), \tag{7}$$

here, Eq. (7) represents the shift-sum operation in Fig. 3(b). Subsequently following the mapping through a 1×1 convolution, M_r undergoes deep convolution and is transformed into depth feature information using the sigmoid function. This depth feature information is then multiplied with the mapped information to obtain the mask self-attention information M_s. Finally, we use the

Optics EXPRESS

shift-back operation to transform the M_s back into a 3D data cube to obtain the MA. The specific operation of shift-back is represented by Eq. (8):

$$MA(x, y, \lambda_i) = \sum_{i=1}^{C} M_S(x - d(\lambda_i), y). \tag{8}$$

We demonstrate the effectiveness of mask attention by ablation experiments, please refer to the Supplement 1 Fig. S2. Meanwhile, for achieving adaptive adjustment of the iterative step size, a dynamic step calculation module is introduced during gradient descent. It relies on the previous iteration's result $x^{(k-1)}$, and through residual operations and sigmoid activation, it achieves adaptive adjustment of the iteration step.

Given that the structure in deep unfolding networks will be repeated multiple times, overly complex structures can lead to excessive parameters and prolonged inference time. Therefore, we adopt simple modules to address the problem of feature domain solving as much as possible. The residual block, composed of convolutions and activation functions, is a commonly used and simple module for image feature processing. Additionally, to address the problem of information loss caused by the limited data transmission between different stages, we introduced a cross-stage feature sharing module. Representing the inter-stage features within each layer as $d^{(k)}$ and their collection as $D^{(k)}$, we obtained a deep-unfolding network composed of the following two iterative steps:

$$\begin{aligned} r^{(k)} &= x^{(k-1)} - \Lambda \odot Conv_2(MA_T \odot \left(Conv_1\left(MA \odot Conv_0(x^{(k-1)}\right)) - y\right)\right), \\ \Lambda &= Sigmoid\left(ResBlock\left(x^{(k-1)}\right)\right), \\ x^{(k)}, d^{(k)} &= ResBlocks\left(r^{(k)}, D^{(k-1)}\right), \\ D^{(k)} &= D.append\left(d^{(k)}\right). \end{aligned} \tag{9}$$

Here, MA_T shares the same structure with MA and is also trainable. The inclusion of "T" indicates its role in replacing various positions of mask attention within the model. Within this framework, the residual block applies a convolutional layer connecting a ReLU activation function, which connects another convolutional layer, followed by addition with the initial input.

For the design of loss function, L1 loss has typically been the primary method employed in the previous multispectral compressive sensing tasks. However, L1 loss outputs poorer in details in our experiments, leading to distortion and other artifacts. To address this limitation, we introduced a U-net-based discriminator to directly evaluate the reliability of the reconstruction. We demonstrate the effectiveness of GAN loss by ablation experiments, please refer to the Supplement 1 Fig. S3. By combining this discriminator evaluation with L1 loss, we derived a hybrid loss:

$$L = L_1 + \gamma L_{gan}, \tag{10}$$

using γ as the blending weight to integrate the two loss components.

Training details. In our implementation, we set the image size $N = H \times W \times C = 128 \times 128 \times 160$, where each measurement size is $H \times (W + C - 1)$. For SHCNet, we default the number of stages to 8 and the feature channels to $C = 128$. For each sample, we collected 1600 data pairs, randomly selecting 160 pairs to construct the test datasets and collecting an additional 1440 pairs to form the training datasets. The batch size was set to $B = 2$, and training was conducted using the ADAM optimizer with a momentum of 0.9 and weight decay of 0.999. Learning the sampling pattern and SHCNet required approximately one week in total on an NVIDIA RTX 3090 GPU for 300 iterations. The learning rate was initialized to 1×10^{-4} and decayed to 1×10^{-6} at the end. The Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were used as quantitative metrics for all evaluations.

Optics EXPRESS

3. Results

In this section, we verify the hyperspectral and confocal imaging capabilities of the SHCMS. Specifically, we evaluate the hyperspectral reconstruction capability by plotting spectral reconstruction curves and validate the confocal imaging capability by comparing the resolution of the reconstructed image with wide-field image. In addition, we performed a comparison of SHCNet with other reconstruction algorithms. We use quantitative evaluation metrics including PSNR and SSIM to compare the image quality of the reconstructed hyperspectral images with the ground truth.

3.1. Hyperspectral confocal experimental results

To verify the performance of the SHCMS, we perform experimental imaging of potato tuber autofluorescence. We employ SHCMS to acquire single-shot compressive measurements of fluorescence as input to the SHCNet and reconstruct hyperspectral confocal images with 160 wavelength channels. Figure 4(a) shows the spectral reconstituted results of potato tubers in three channels (513.4 nm, 559.5 nm, 599.6 nm) out of 160 wavelength channels and its merge image,

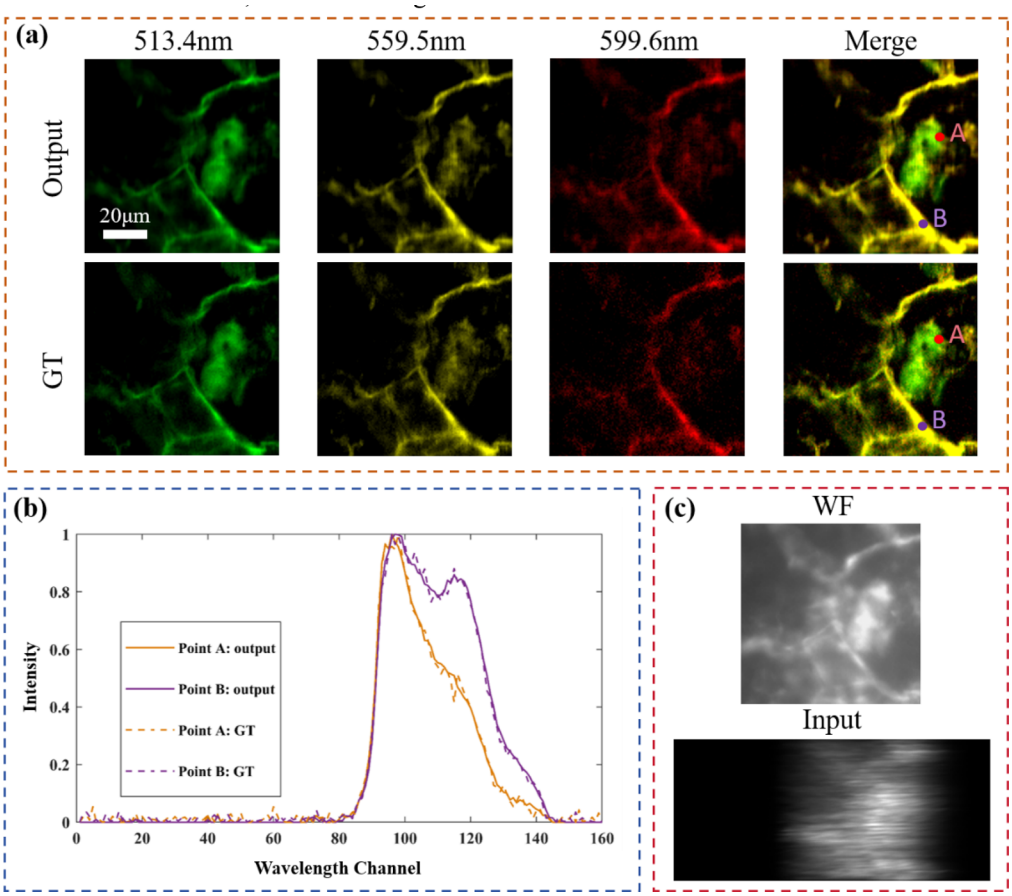

Fig. 4. The autofluorescence of potato tuber. (a) Reconstruction results in three channels (513.4 nm, 559.5 nm, 599.6 nm) out of 160 spectral channels and its merge image with SHCNet. (b) Two points marked by 'A' and 'B' in the merge image are selected to plot the spectral curves. (c) The compressive measurement and wide-field image corresponding to the same scene. GT, ground truth; WF, wide-field.

Optics EXPRESS

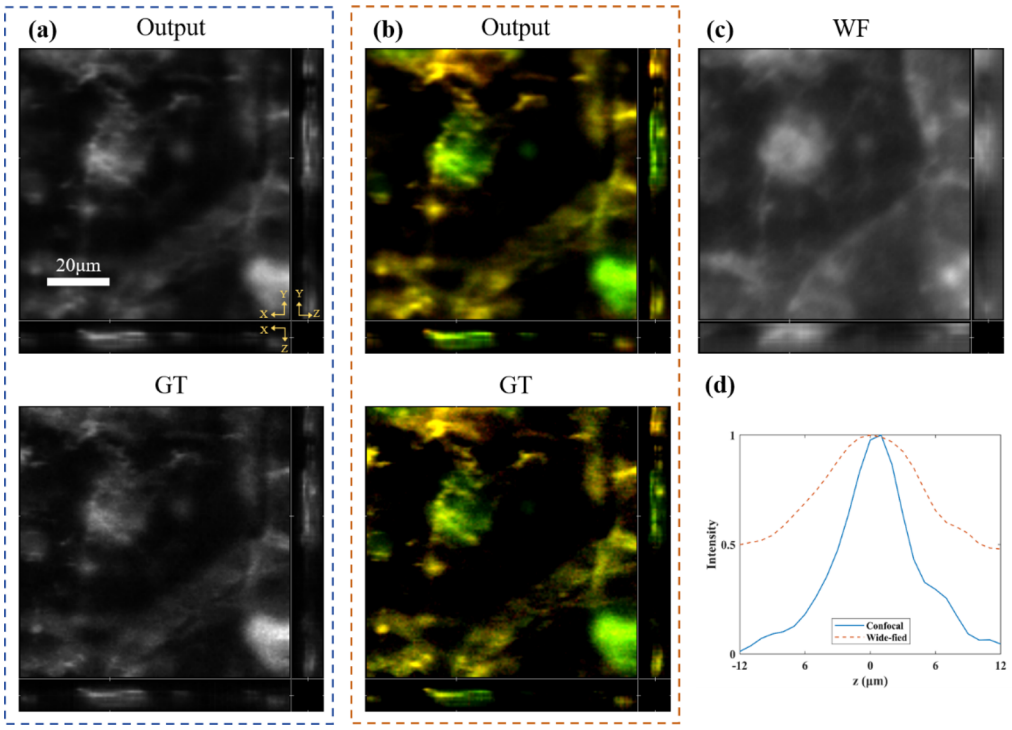

Fig. 5. Spatial 3D hyperspectral imaging of potato tubers. (a) Reconstruction results and GT in single channel (513.4 nm) with SHCNet. (b) Reconstruction merge results and GT in three channels (513.4 nm, 559.5 nm, 599.6 nm) out of 160 spectral channels with SHCNet (c) Wide-field image corresponding to the same scene. (d) Confocal and wide-field profile of the intensity value with respect to the z position, which are calculated from the intensity image of 1 μm fluorescent spheres.

which are very close to the ground truth in visualization. The size of all reconstructed images shown in this study is 128×128. More wavelength channels of potato tuber autofluorescence hyperspectral images can be found on Fig. S4. We also provide MSE and SSIM indices for the spectral data cube, which are both indicative of the high quality of our network reconstruction, as shown in Fig. S5.

Figure 4(b) shows the spectral differences of two selected point (A and B) within 160 channels between network outputs and ground truth. Two solid lines (network output) almost coincide with the dashed line (ground truth), which means that the spectral distribution of the SHCNet reconstruction is almost identical to that of the ground truth. Meanwhile, the difference between the spectral distribution curves of points A and B can indicate the different components in potato tubers. In contrast, it is difficult to select a suitable filter to observe the autofluorescence because of the broad spectral response [42]. To validate our spectral reconstruction capability in all channels, we performed hyperspectral reconstruction of Hematoxylin and Eosin (H&E) stained pathology slide under halogen illumination, the details are shown in Fig. S6.

Figure 4(c) shows the compressive measurements and wide-field images acquired by the system. The compressive measurement image is captured at the sCMOS exposure time of 20 ms with a size of 128×287. Here, the wide-field images are obtained by the sCMOS with the DMD displaying an all-"on" pattern and the prism is removed. It can be seen that the reconstructed confocal image has higher spatial resolution compared to the wide-field image.

Optics EXPRESS

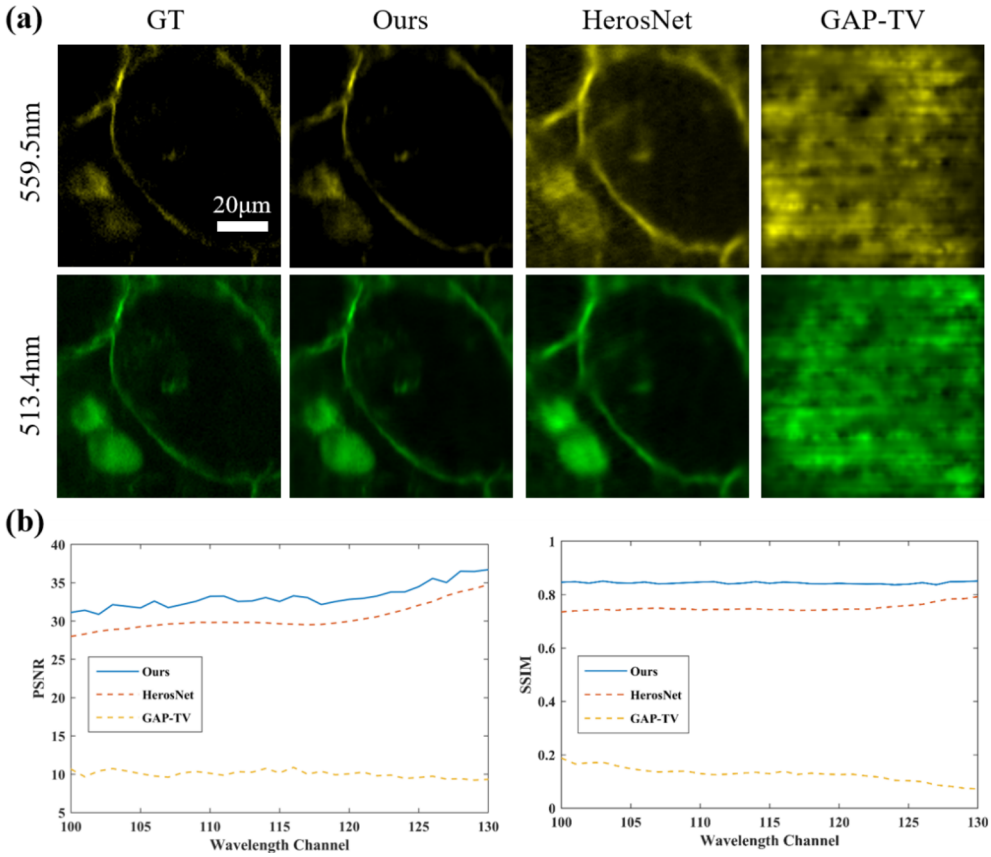

Fig. 6. Performance comparison among different reconstruction algorithms. (a) Reconstruction results in two channels (513.4 nm and 559.5 nm) with SHCNet, HerosNet, and GAP-TV. (b) PSNR and SSIM values of 100-130 wavelength channel.

3.2. 3D imaging of potato tubers

To verify the confocal imaging capability of SHCMS, we performed spatial 3D imaging of potato tubers. We capture 15 z-slices according to 5 μm step size in the z-direction. The moving mechanism is realized by a motorized translation stages (Thorlabs, model number: MT1/M-Z9). Figure 5(a) shows the triple-view maximum intensity projection (MIP) of the single-channel (513.4 nm) hyperspectral image of potato tubers. We can observe that it has a higher resolution compared to the wide-field images (Fig. 5(c)). Meanwhile, the merged hyperspectral image of the three channels (513.4 nm, 559.5 nm, 599.6 nm) has equally high resolution, as shown in Fig. 5(b).

In addition, to quantitatively analyze confocal capability, we take 1 μm fluorescent spheres as samples and measure their intensity degradation with the out-of-focus position. We acquire a series of line-scan confocal and wide-field images obtained with the sample stage moving along the z direction from −12 to +12 μm. When the depth of defocus increases, the intensity of the fluorescent spheres gradually decreases. The degradation rate of intensity can be used as a measure to evaluate the effectiveness of eliminating the "out-of-focus" light. When the degradation rate of intensity is large, it means that the image is less affected by out-of-focus light. Therefore, SHCMS has a higher axial resolution compared to the wide-field image as seen in Fig. 5(d).

Optics EXPRESS

3.3. Performance comparison among different reconstruction algorithms

We compare our SHCNet with two hyperspectral reconstruction methods, including HerosNet and GAP-TV [43]. HerosNet adopts a similar deep unfolding networks framework to ours and has shown commendable performance. GAP-TV belongs to the optimization iterative algorithm. All methods are implemented based on their official source codes and evaluated with the same parameters.

Figure 6(a) shows hyperspectral confocal reconstruction of different algorithms in two channels (513.4 nm and 559.5 nm) out of 160 wavelength channels. The reconstructed images of different reconstruction algorithms were scaled uniformly. It can be observed that our network outputs the reconstruction having finer details and sharper image edges than other methods. HerosNet is able to reconstructs the morphological features but blurs in detail presentation. This verifies that the U-Net discriminator module we introduced is favorable for reconstructing details. Meanwhile, the GAP-TV cannot reconstruct any information because the compressive rate is too substantial for reconstructing 160 channels. To quantitatively compare the reconstruction performance, we use the PSNR and SSIM as the evaluation indicator. We calculated PSNR and SSIM values for the 100-130 channel, where autofluorescence of potato tubers is concentrated. SHCNet achieves the highest PSNR and SSIM value, as shown in Fig. 6(b). It can be concluded that SHCNet presents the best in spectral results.

4. Discussion

Deep learning algorithms are data-driven and usually require pre-training process. The quality of the training datasets determines the performance of the network. However, deep learning algorithms are often trained using simulated synthetic datasets to deal with a lack of paired data. The lack of training data in experiment limits the effectiveness and generalization of the algorithms in most deep-learning application scenarios. The advantage of SHCMS is that it can acquire paired datasets of hyperspectral confocal images and snapshot compressive measurements in realistic scenarios by switching the pattern of the DMD. This means that SHCMS can continuously generate large, diverse, and paired training datasets to enhance the robustness and generalization performance of SHCNet.

We experimented to compare the reconstruction of SHCNet trained with realistic datasets (R.D.) and simulated datasets (S.D.). We simulated the physical process of snapshot compressed sampling on a computer. The 160-bands hyperspectral confocal images are multiplied by the mask and then stacked along the column direction to generate simulated snapshot compressive measurements. We trained the network with simulated data and tested the network with realistic data. A compressive measurement captured by SHCMS is the input to the network for testing, which never appears during the training of the network. Figure 7 shows the reconstruction of the network trained with R.D. and S.D., respectively. The single channel reconstruction of the network trained with S.D. only shows the contours of the sample and the entire image is blurred, as shown in Fig. 7(a). On the contrary, the reconstruction of the network trained with R.D. is the closest to the ground truth. Quantitively, we provide the PSNR and SSIM indices of these two networks in channels 100 to 130 (Spectral interval of sample fluorescence), as shown in Fig. 7(b). Both PSNR and SSIM scores of R.D. trained network are much higher than S.D. trained network.

The reason for such a result is that there is a difference between the simulated compressive measurements and the real collected compressive measurements. Many factors in the hardware including point spread function (PSF), spectral calibration and noise etc. may cause this difference. Furthermore, comprehensively accounting for all variables within the simulation process poses significant challenges, particularly due to the substantial cost associated with calibrating the PSF for each wavelength. Nevertheless, our system facilitates the acquisition of optical system properties through the training process with the realistic paired dataset. This attribute stands as a distinctive advantage of our approach, underscoring the novelty and efficacy of our methodology.

Optics EXPRESS

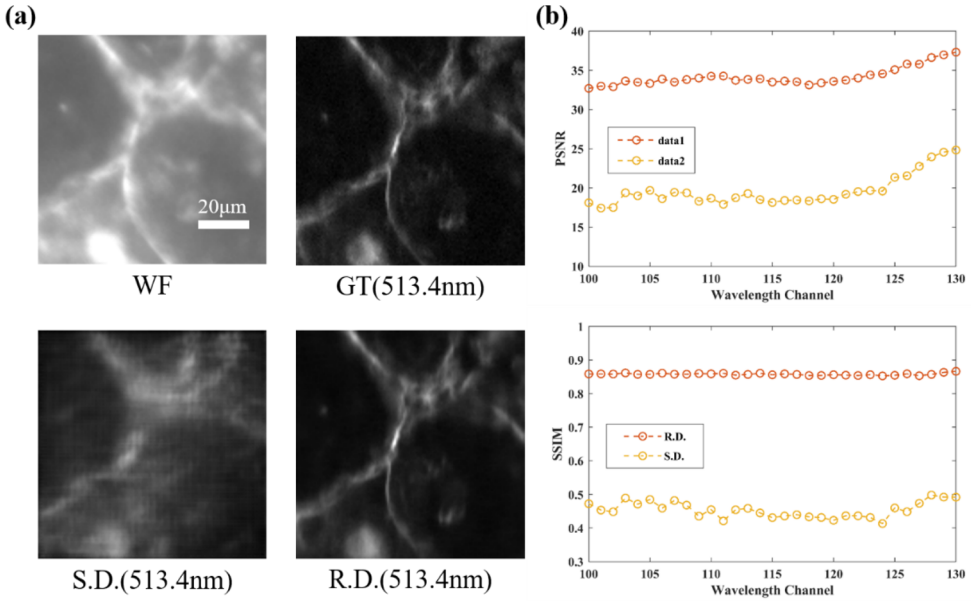

Fig. 7. Performance comparison of SHCNet trained with R.D. and simulated datasets S.D. (a) Reconstruction results in single channels (513.4 nm). (b) PSNR and SSIM values of spectral channels 100 to 130. R.D.: realistic datasets, S.D.: simulated datasets.

5. Conclusion

In this paper, we develop a novel hyperspectral confocal microscopy along with a deep learning reconstruction algorithm, achieving snapshot hyperspectral confocal microscopic imaging. Based on the programmable characteristics of DMD, we can acquire the paired datasets. These data are used as training datasets for the SHCNet to improve its robustness to realistic noise and learn the spectral response of SHCMS. The performance of our system may be continually enhanced as the dataset increases. Moreover, we propose a deep unfolding network structure based on the GAN network framework. The U-Net structured discriminators of our network that show advantages in detail reconstruction. Finally, we achieve snapshot hyperspectral confocal images of potato tuber autofluorescence, yielding high-contrast 160-bands hyperspectral fluorescence images. We hope the proposed method will benefit future works in compressive hyperspectral image reconstruction and confocal microscopy.

Funding. Shenzhen Science and Technology Research and Development Funds (WDZC20200821104802001); Fundamental Research Funds for the Central Universities (Grant No. HIT.OCEF.2023050); Shenzhen Science and Technology Project (GXWD20220818170353009, JCYJ20200109142808034); National Natural Science Foundation of China (62031023, 62331011).

Disclosures. The authors declare no conflicts of interest.

Additional information
Correspondence and requests for materials should be addressed to Y. Z. (ybzhang08@hit.edu.cn).

Data availability. For reproducible research, we will upload the complete source code of our SHCNet and a partial dataset at [44].

Supplemental document. See Supplement 1 for supporting content.

References

1. M. B. Sinclair, J. A. Timlin, D. M. Haaland, *et al.*, "Design, construction, characterization, and application of a hyperspectral microarray scanner," Appl. Opt. **43**(10), 2079–2088 (2004).

Optics EXPRESS

2. C. A. S. Monteiro, D. J. Chow, G. R. Leal, *et al.*, "Optical imaging of cleavage stage bovine embryos using hyperspectral and confocal approaches reveals metabolic differences between on-time and fast-developing embryos," Theriogenology **159**, 60–68 (2021).
3. A. M. Valm, S. Cohen, W. R. Legant, *et al.*, "Applying systems-level spectral imaging and analysis to reveal the organelle interactome," Nature **546**(7656), 162–167 (2017).
4. W. Shi, D. E. Koo, M. Kitano, *et al.*, "Pre-processing visualization of hyperspectral fluorescent data with Spectrally Encoded Enhanced Representations," Nat. Commun. **11**(1), 726 (2020).
5. Z. Liao, F. Sinjab, H. M. Elsheikha, *et al.*, "Optical sectioning in multifoci Raman hyperspectral imaging," J. Raman Spectrosc. **49**(10), 1660–1667 (2018).
6. A. J. Bares, M. A. Mejooli, M. A. Pender, *et al.*, "Hyperspectral multiphoton microscopy for in vivo visualization of multiple, spectrally overlapped fluorescent labels," Optica **7**(11), 1587–1601 (2020).
7. F. R. Bertani, E. Botti, L. Ferrari, *et al.*, "Label-free and non-invasive discrimination of HaCaT and melanoma cells in a co-culture model by hyperspectral confocal reflectance microscopy," J. Biophotonics **9**(6), 619–625 (2016).
8. T. Kubo, K. Temma, N. I. Smith, *et al.*, "Hyperspectral two-photon excitation microscopy using visible wavelength," Opt. Lett. **46**(1), 37–40 (2021).
9. W. F. Vermaas, J. A. Timlin, H. D. Jones, *et al.*, "In vivo hyperspectral confocal fluorescence imaging to determine pigment localization and distribution in cyanobacterial cells," Proc. Natl. Acad. Sci. **105**(10), 4050–4055 (2008).
10. D. M. Haaland, H. D. Jones, M. H. Van Benthem, *et al.*, "Hyperspectral confocal fluorescence imaging: exploring alternative multivariate curve resolution approaches," Appl. Spectrosc. **63**(3), 271–279 (2009).
11. J.-A. Conchello and J. W. Lichtman, "Optical sectioning microscopy," Nat. Methods **2**(12), 920–931 (2005).
12. R. H. Webb, "Confocal optical microscopy," Rep. Prog. Phys. **59**(3), 427–471 (1996).
13. C. Sheppard and A. Choudhury, "Image formation in the scanning microscope," Opt. Acta **24**(10), 1051–1073 (1977).
14. M. Liang, R. L. Stehr, and A. W. Krause, "Confocal pattern period in multiple-aperture confocal imaging systems with coherent illumination," Opt. Lett. **22**(11), 751–753 (1997).
15. Q. S. Hanley, P. J. Verveer, M. J. Gemkow, *et al.*, "An optical sectioning programmable array microscope implemented with a digital micromirror device," J. Microsc. **196**(3), 317–331 (1999).
16. S. Li and R. Liang, "DMD-based three-dimensional chromatic confocal microscopy," Appl. Opt. **59**(14), 4349–4356 (2020).
17. D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006).
18. E. Dandes, "Near-optimal signal recovery from random projections," IEEE Trans. Inform. Theory **52**(12), 5406–5425 (2006).
19. G. R. Arce, D. J. Brady, L. Carin, *et al.*, "Compressive coded aperture spectral imaging: An introduction," IEEE Signal Process. Mag. **31**(1), 105–115 (2014).
20. M. E. Gehm, R. John, D. J. Brady, *et al.*, "Single-shot compressive spectral imaging with a dual-disperser architecture," Opt. Express **15**(21), 14013–14027 (2007).
21. Z. Yu, D. Liu, L. Cheng, *et al.*, "Deep learning enabled reflective coded aperture snapshot spectral imaging," Opt. Express **30**(26), 46822–46837 (2022).
22. W. Zhang, J. Suo, K. Dong, *et al.*, "Handheld snapshot multi-spectral camera at tens-of-megapixel resolution," Nat. Commun. **14**(1), 5043 (2023).
23. M. Yako, Y. Yamaoka, T. Kiyohara, *et al.*, "Video-rate hyperspectral camera based on a CMOS-compatible random array of Fabry–Pérot filters," Nat. Photonics **17**(3), 218–223 (2023).
24. A. Wagadarikar, R. John, R. Willett, *et al.*, "Single disperser design for coded aperture snapshot spectral imaging," Appl. Opt. **47**(10), B44–B51 (2008).
25. D. Kittle, K. Choi, A. Wagadarikar, *et al.*, "Multiframe image estimation for coded aperture snapshot spectral imagers," Appl. Opt. **49**(36), 6824–6833 (2010).
26. Y. Wu, I. O. Mirza, G. R. Arce, *et al.*, "Development of a digital-micromirror-device-based multishot snapshot spectral imaging system," Opt. Lett. **36**(14), 2692–2694 (2011).
27. L. Wang, Z. Xiong, D. Gao, *et al.*, "Dual-camera design for coded aperture snapshot spectral imaging," Appl. Opt. **54**(4), 848–858 (2015).
28. Y. Qiu, S. Zhao, X. Ma, *et al.*, "Hyperspectral image reconstruction via patch attention driven network," Opt. Express **31**(12), 20221–20236 (2023).
29. Z. Xiong, Z. Shi, H. Li, *et al.*, "Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*(2017), pp. 518–525.
30. L. Wang, T. Zhang, Y. Fu, *et al.*, "Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging," IEEE Trans. on Image Process. **28**(5), 2257–2270 (2019).
31. X. Miao, X. Yuan, Y. Pu, *et al.*, "l-net: Reconstruct hyperspectral images from a snapshot measurement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*(2019), pp. 4059–4069.
32. H. Xie, Z. Zhao, J. Han, *et al.*, "Dual camera snapshot hyperspectral imaging system via physics-informed learning," Optics and Lasers in Engineering **154**, 107023 (2022).
33. H. Xie, Z. Zhao, J. Han, *et al.*, "Dual camera snapshot high-resolution-hyperspectral imaging system with parallel joint optimization via physics-informed learning," Opt. Express **31**(9), 14617–14639 (2023).

Optics EXPRESS

34. X. Wang, L. Xie, C. Dong, *et al.*, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *Proceedings of the IEEE/CVF international conference on computer vision*(2021), pp. 1905–1914.
35. C. Sun, H. Liu, M. Jia, *et al.*, "Review of calibration methods for Scheimpflug camera," J. Sens. **2018** (2018).
36. R. H. Shepard, C. Fernandez-Cull, R. Raskar, *et al.*, "Optical design and characterization of an advanced computational imaging system," in *Optics and Photonics for Information Processing VIII* (SPIE 2014), pp. 73–87.
37. H. Rueda-Chacon, F. Rojas, and H. Arguello, "Compressive spectral image fusion via a single aperture high throughput imaging system," Sci. Rep. **11**(1), 10311 (2021).
38. N. Parikh and S. Boyd, "Proximal algorithms," FNT in Optimization **1**(3), 127–239 (2014).
39. J. Zhang, B. Chen, R. Xiong, *et al.*, "Physics-inspired compressive sensing: Beyond deep unrolling," IEEE Signal Process. Mag. **40**(1), 58–72 (2023).
40. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE 2009), pp. 693–696.
41. X. Zhang, Y. Zhang, R. Xiong, *et al.*, "Herosnet: Hyperspectral explicable reconstruction and optimal sampling deep network for snapshot compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*(2022), pp. 17532–17541.
42. W. Jahr, B. Schmid, C. Schmied, *et al.*, "Hyperspectral light sheet microscopy," Nat. Commun. **6**(1), 7990 (2015).
43. X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *2016 IEEE International conference on image processing (ICIP)* (IEEE 2016), pp. 2539–2543.
44. S. Liu, W. Zou, H. Sha, *et al.*, "Source code of SHCNet and a partial dataset for Deep learning-enhanced snapshot hyperspectral confocal microscopy imaging system," GitHub (2024), https://github.com/ClementZou/SHCNet.