Research Article | Vol. 31, No. 12 / 5 Jun 2023 / Optics Express 20489
Optics EXPRESS

Snapshot hyperspectral imaging based on equalization designed DOE

NAN XU,[1] HAO XU,[1] SHIQI CHEN,[1] HAIQUAN HU,[1] ZHIHAI XU,[1] HUAJUN FENG,[1] QI LI,[1] TINGTING JIANG,[2] AND YUETING CHEN[1,*]

[1]State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China
[2]Research Center for Intelligent Sensing Systems, Zhejiang Laboratory, Hangzhou 311100, China
[*]chenyt@zju.edu.cn

Abstract: Hyperspectral imaging attempts to determine distinctive information in spatial and spectral domain of a target. Over the past few years, hyperspectral imaging systems have developed towards lighter and faster. In phase-coded hyperspectral imaging systems, a better coding aperture design can improve the spectral accuracy relatively. Using wave optics, we post an equalization designed phase-coded aperture to achieve desired equalization point spread functions (PSFs) which provides richer features for subsequent image reconstruction. During the reconstruction of images, our raised hyperspectral reconstruction network, CAFormer, achieves better results than the state-of-the-art networks with less computation by substituting self-attention with channel-attention. Our work revolves around the equalization design of the phase-coded aperture and optimizes the imaging process from three aspects: hardware design, reconstruction algorithm, and PSF calibration. Our work is putting snapshot compact hyperspectral technology closer to a practical application.

1. Introduction

Hyperspectral imaging (HSI) technology is an imaging method based on several narrow spectral bands. Combining imaging technology with hyperspectral technology [1] allows detection of two-dimensional image information and one-dimensional spectral information of the target with high spectral resolution. Given this characteristic, HSI is gaining widespread use in biology [2,3], medicine [4,5], food detection [6,7], remote sensing [8,9] and other fields.

Traditional hyperspectral imaging solutions generate images by scanning in a single spectral dimension or a single spatial dimension [10]. This staring imaging method takes great time costs, precluding dynamic scenes or video capture. To address this, researchers have developed snapshot compressive imaging systems (SCI) to obtain HSI cubes. An excellent example is the coded aperture snapshot spectral imaging (CASSI) system [11,12]. However, it suffers from large system volume and weight due to the complex structure involving dispersion elements. Consequently, there are attempts to reconstruct spectral information from RGB images captured by ordinary cameras directly [13–15], which maybe not accurate enough.

To solve the above issues, more and more research has been devoted into compact snapshot hyperspectral imaging systems [16,17]. Jeon et al. [17] designed a specific diffractive optical element (DOE) to generate spectrally-varying point-spread-functions (PSFs). They realized hyperspectral imaging by adding a novel DOE in front of an exposed sensor. PSFs rotate in a variety of directions to encode the target's reflectance at different wavelengths. And finally, the reconstruction algorithm reconstructs the three-dimensional (3D) HSI cube from the detected encoded two-dimensional (2D) image. Furthermore, to enhance the sufficiency of information coding, Arguello et al. [18] introduced an optimized design of color-coded aperture (CCA) based on DOE imaging. In recent years, an end-to-end experimental method [19–23] has been proposed in the field of computational imaging. The DOE element and reconstruction algorithm are

#493498 https://doi.org/10.1364/OE.493498
Journal © 2023 Received 18 Apr 2023; accepted 22 May 2023; published 1 Jun 2023

Optics EXPRESS

simultaneously optimized by deep learning, and the best reconstruction performance is achieved for a specific scene.

Using designed diffractive elements for hyperspectral imaging is an excellent solution. Compared to traditional scanning spectrometers and CASSI hyperspectral systems with complex imaging optical paths, this compact imaging system is more versatile and portable. In addition, there are other solutions for a compact hyperspectral imaging system. The compact system based on wavelength coding [16,24,25] achieves hyperspectral imaging by plating filters of different wavelengths on the pixels. However, the imaging channels of this system are limited, since more imaging channels mean lower spatial resolution. The compact imaging system based on diffusers [16,26] encodes with off-the-shelf and non-designed phase encoder, which often leads to large PSFs, which makes the spatial resolution of the reconstructed image unable to be guaranteed. Designed DOE-based imaging system is free of the above shortcomings.

The study of hyperspectral reconstruction algorithms is also an important aspect of hyperspectral imaging. In general, the traditional iterative approach used by Jeon et al. [17] has the disadvantage of being time-consuming and having limited effectiveness in restoring complexly degraded images. The reconstruction algorithm based on convolutional neural network, which has developed apace in recent years, can achieve better recovery effect. Cai et al. [27,28] introduced transformer structures into hyperspectral reconstruction. Transformer [29–32] is a multiplication-based neural network with a unique overall structure, which is mainly composed of the self-attention module. Even though it may yield better results than convolutional neural network, the computational complexity is much greater. According to Yu [33], the excellent ability of transformer to perform visual tasks may not be due to complex self-attention operations, but rather to its overall structure. The results are also confirmed in a later work [34], although its consideration is different from Yu's [33]. Based on this consideration, we propose a CAFormer network structure that can reduce the computational burden by replacing self-attention with channel-attention. It achieves better results than the existing state-of-the-art neural network with lower computational load and less computing time.

Our work is an incremental improvement on Jeon's [17]. There are some areas in Jeon's work that could be improved, including the DOE design and the reconstruction algorithm. Hu [35] has made some improvements to the design of DOE in this system, changing the three-lobed PSFs into two-lobed ones to enhance imaging quality [36]. On the basis of previous works, we further improved the imaging effect, summarized and implemented the imaging process as shown in Fig. 1, and made the following contributions:

- We propose an equalization design of phase-coded aperture making PSFs at different wavelengths approximately equal in size, which provides richer features for subsequent image reconstruction. In Section 2 we theoretically analyze the effectiveness of our design. Our design improves the information content of DOE at spectral dimension, which results in a reduction of information loss and a more accurate reconstruction.

- We describe a CAFormer to reduce the computational complexity effectively by substituting self-attention with channel-attention and improve the performance of image reconstruction greatly, including GateFusion module and CAFM. Finally, more hyperspectral information is reconstructed, and the hyperspectral imaging wavelength bands are expanded from 460-700nm to 460-900nm.

In Section 2. Section 3 and Section 4 we present the specific conceptual foundations of our proposed method. Section 5 outlines the results, verifying the effectiveness and feasibility of our method. The major lessons learned thus far are summarized in the concluding Section 6.

Optics EXPRESS

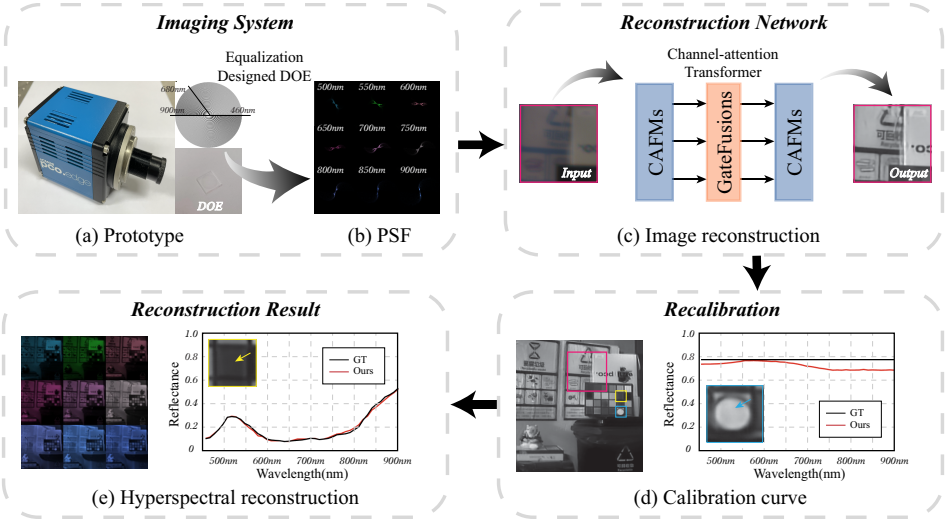

Fig. 1. (a-b) The equalization designed DOE generates a set of spectrally-varying PSFs with an anisotropic shape. Light emitted by the target is modulated by DOE and imaged on the detector. (c) Coded image is subsequently input into the CAFormer network to output the HSI cube. (d-e) After adding the recalibration information of the light source, the true reflectivity cube of the target can be recovered. GT is the ground truth.

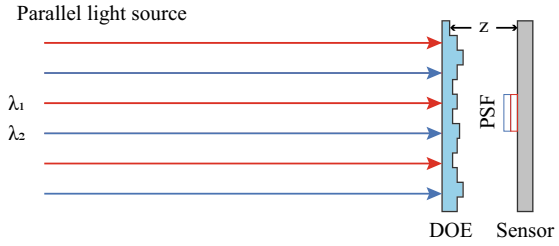

Fig. 2. Light field of a wide wavelength-band point light source propagating from infinity to detector surface. Different wavelengths of light are imaged simultaneously on the detector surface, and their images are linearly superimposed.

2. Design and analysis of diffraction elements

The main purpose of this section is to describe the principle of diffraction element design. We evaluated the effectiveness of designed DOE by using Fisher information as an evaluation index.

2.1. Mathematical model of diffraction imaging

Image coding relies on a DOE to modulate the phase of incident light. DOE transmits light with different wavelengths, which are converted into a phase changes, resulting in wavelength-specific encoded images. To calculate the PSF of the system, it's necessary to assume that point light is incident. As shown in Fig. 2, considering the point light source located at infinity, the light source is approximately directional. In the object space whose coordinates are expressed as (x, y), a plane incident light field u_0 with an initial phase ϕ_0 and an initial amplitude intensity A_0 can be expressed as:

$$u_0(x, y) = A_0(x, y)e^{i\phi_0(x,y)}. \tag{1}$$

Optics EXPRESS

DOE changes the phase but not the amplitude:

$$u_{DOE}(x,y) = A_0(x,y)e^{i(\phi_0(x,y)+\phi_{DOE}(x,y))}, \tag{2}$$

where u_{DOE} is the light field of the back surface of DOE. ϕ_{DOE} represents the phase change caused by DOE, which can be expressed as:

$$\phi_{DOE}(x,y) = \frac{2\pi}{\lambda}\Delta n_\lambda h_{DOE}(x,y), \tag{3}$$

where λ is the wavelength of incident light; Δn_λ is the refractive index contrast between DOE and air; h_{DOE} is the height function of DOE. To design DOE is to design its height function.

We calculate the propagation process of light from the DOE rear surface to the sensor surface using the Fresnel diffraction formula. Thus, the light field u_{sensor} at the sensor surface is given by:

$$u_{sensor}(x',y') = \frac{e^{ikz}}{i\lambda z}\iint u_{DOE}(x,y)e^{\frac{ik}{2z}\{(x'-x)^2+(y'-y)^2\}}dxdy, \tag{4}$$

where (x', y') represents the coordinates of image space, k is the wave number of the incident light field, and z is the distance from the rear surface of DOE to the sensor surface.

Image at a specific wavelength, which is PSF, can be calculated by the following formula:

$$p_\lambda(x',y') = |\frac{e^{ikz}}{i\lambda z}\iint u_{DOE}(x,y)e^{\frac{ik}{2z}\{(x'-x)^2+(y'-y)^2\}}dxdy|^2. \tag{5}$$

For convenience, we can calculate PSF with the help of Fourier transform $\mathscr{F}(\cdot)$:

$$p_\lambda(x',y') \propto |\mathscr{F}[e^{i\phi_{DOE}(x,y)}e^{i\frac{\pi}{\lambda z}(x^2+y^2)}]|^2. \tag{6}$$

To conclude, the image of a target consists of the sum of images at all wavelength bands:

$$I_{sensor}(x',y') = \int \Omega_{camera}(\lambda)([I_\lambda(x',y') * p_\lambda(x',y')]d\lambda, \tag{7}$$

where I_λ is the monochromatic image at wavelength λ, Ω_{camera} is spectral response function of the camera. In Eq. (7), we see that if we design a DOE that produces different PSFs for different wavelengths of incident light, then we can encode the information of different wavelengths using this imaging model. The encoded 2D image can be decoded into the hyperspectral cube of the target through a specific reconstruction algorithm.

2.2. Design method of the diffraction element

Based on the method described by Jeon et al. [17], the DOE plane is viewed as a polar coordinate plane (r, θ) in the first step. In terms of a Fresnel len with characteristic wavelength λ and focal length f, the height function h_{DOE} can be expressed as follows:

$$\begin{aligned}\Delta h(r;\lambda) &= h_0 - h_{DOE}\\ &= \frac{\lambda\Delta\phi_{DOE}}{2\pi\Delta n_\lambda}\\ &= \frac{2\pi m - \frac{2\pi}{\lambda}(\sqrt{r^2+f^2}-f)}{2\pi\Delta n_\lambda}\\ &= \frac{m\lambda - (\sqrt{r^2+f^2}-f)}{\Delta n_\lambda},\end{aligned} \tag{8}$$

where m is a positive integer, usually 1, h_0 is the height of the flat substrate on which DOE is etched, h_{DOE} is the DOE height after etching, and Δh is the etching depth of the substrate.

Optics EXPRESS

To modulate light with wavelengths from 460nm to 900nm rather than only a single characteristic wavelength, we can make radius at different angles θ on DOE plane correspond to different characteristic wavelengths. The characteristic wavelength increases or decreases as the angle in the polar coordinate system changes:

$$\lambda(\theta) = \begin{cases} \lambda_{min} + (\lambda_{max} - \lambda_{min})\frac{(N\theta)^{\alpha}}{(2\pi)^{\alpha}} & 0 \le \theta < \frac{2\pi}{N}, \\ \lambda(\theta - \frac{2\pi}{N}) & \theta \ge \frac{2\pi}{N}. \end{cases} \tag{9}$$

where N is a positive integer, α is the distribution scaling factor. When $N = 2$ and $\alpha = 1.0$, substituting Eq. (9) into Eq. (8), the height map of DOE obtained is shown in Fig. 3.

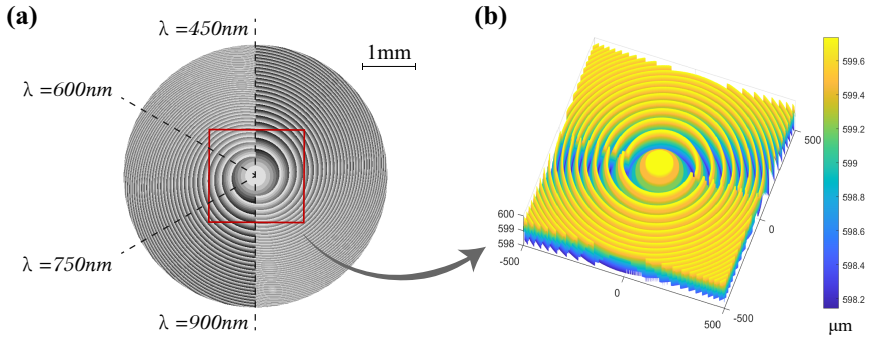

Fig. 3. Height pattern of designed DOE. (a) Plane view of DOE with diameter of 5mm and characteristic wavelength of 460 – 900nm. (b) Local 3D view of DOE etched on the substrate of 600μm.

Figure 4(b) illustrates PSFs generated by DOE in Fig. 3 as the characteristic wavelengths distribution rule. In some reconstructed hyperspectral images based on simulation, we find that the spectral accuracy of the long-wavelength band was lower than that of the short-wavelength band (as shown in Fig. 9). After analysis, we find that the PSF size decreases significantly with increasing wavelength. There is a possibility that the small size of the long-wavelength PSF will cause it to be obscured by the PSFs of other wavelengths, resulting in poor coding of the long-wavelength information, subsequently reducing the accuracy of long-wavelength spectral reconstruction. In view of this fact, we hope that the PSF size of all wavelengths will be more equal.

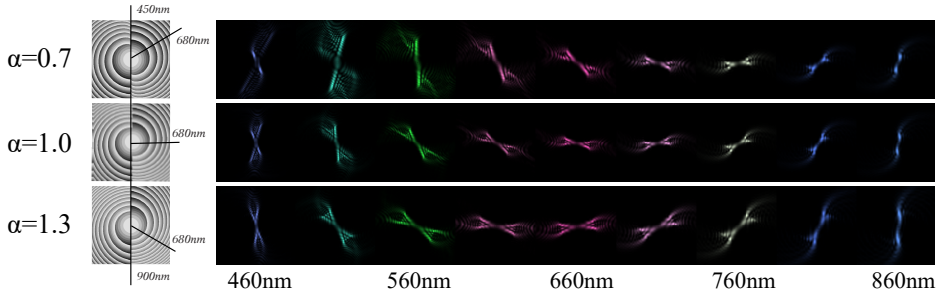

Fig. 4. When the scaling factor α we propose is 0.7, 1.0 and 1.3, different designed DOEs produce different PSF sets.

Optics EXPRESS

Height map of a DOE can be approximately differentiated into many rectangular holes, and the light intensity solution $I(x, y)$ of Fraunhofer diffraction of rectangular holes is:

$$I(x,y) = I_0 sinc^2(\frac{ax}{\lambda f}) sinc^2(\frac{by}{\lambda f}), \tag{10}$$

where I_0 is the intensity of the center point, a and b is the size of the rectangular hole, λ is the wavelength of the incident light, and f is the imaging distance. This fact suggests that on the DOE we designed, when the characteristic wavelength distribution is more compact, the size of the rectangle holes will become smaller, resulting in a larger diffraction spot. Therefore, we think of introducing a scaling factor, α, into Eq. (9) to adjust the distribution of characteristic wavelengths.

As shown in Fig. 4, When $\alpha = 1.0$, the designed DOE is the same as that proposed by Jeon. By increasing α value, the PSF size at long wavelength gradually increases, while that of short wavelength decreases. On the contrary, the PSF size of long wavelength reduces with an increase in α.

To calculate the PSF size, we binarize the PSF, calculate the outer rectangle of the PSF, and define the diagonal length of the rectangle as the size. Figure 5(a) shows the PSF sizes under different scaling factors. In our trial and analysis, we found that PSFs have approximate sizes at different wavelengths when $\alpha = 1.3$, achieving the desired equalization design.

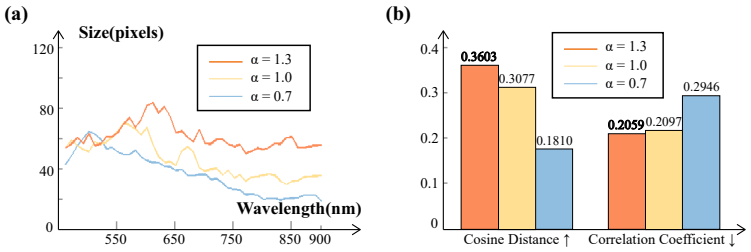

Fig. 5. (a) PSF sizes under different scaling factors. (b) Similarity between PSFs at adjacent wavelengths under different scaling factors. The higher the correlation coefficient and the lower the cosine distance, the more similar the PSFs are.

An effective DOE design can enhance the distinction between information at different wavelengths by increasing the difference between PSFs. This means that the lower the correlation among PSFs, the more efficient the DOE will be. As shown in Fig. 5(b), we calculate the correlation coefficient and cosine distance between PSFs of these DOEs. Our proposed equalization designed DOE (α) has the lowest correlation coefficient and the largest cosine distance, so the correlation between different PSFs is the lowest. This proves that the DOE we designed is more effective.

2.3. Evaluation indicator of the diffraction element

Fisher information is a way to measure the amount of information about unknown parameters $\boldsymbol{\theta}$ (λ in our case) carried by observable random variable X (PSF in our case). Fisher information is defined as:

$$\begin{aligned} FI(\boldsymbol{\theta}) &= E(\frac{\partial^2 lnX(s;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}) \\ &= \frac{1}{X(s;\boldsymbol{\theta}) + \beta}(\frac{\partial X(s;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2, \end{aligned} \tag{11}$$

where E represents the expectation of the function, s is the independent variable of X, and β represents system noise.

Optics EXPRESS

In microscopic 3D imaging, Fisher information is often used to evaluate depth sensitivity of PSF. Based on these studies, we propose to calculate the Fisher information related to wavelength of DOE to determine which DOE is best suited to encoding in hyperspectral imaging. According to Eq. (11), the Fisher information $FI(\lambda)$ is as follows:

$$FI(\lambda) = \sum_{k=1}^{N_p} \frac{1}{p_\lambda(k) + n} \left(\frac{\partial p_\lambda(k)}{\partial \lambda}\right)^2 \tag{12}$$

where $p_\lambda(k)$ is the intensity of PSF on pixel k at wavelength λ, and n represents the image noise.

Fisher information is an array for a 3D PSF. The total amount of information in DOE is the sum of all the numbers in the array. In Table 1, we list the Fisher information for different scaling factors. We can see from the data that the equalization designed DOE proposed ($\alpha = 1.3$) has the maximum Fisher information. This indicates that this DOE is capable of carrying more wavelength information. In addition to calculating the size, similarity, and the total amount of information, we conducted reconstruction of simulation and real images, as shown in Sec. 5.2, to demonstrate the efficacy of our proposed design.

Table 1. Fisher information of DOEs with different scaling factors.

Scaling factor	$\alpha = 0.7$	$\alpha = 1.0$	$\alpha = 1.3$
Fisher information	0.9603	1.2991	**1.6848**

Table 2. Reconstruction experiment results of equalization designed DOE.

Pattern	MPSNR↑	MSSIM↑	SAM↓
α=0.7	29.27	0.973	7.139
α=1.0	29.83	0.977	4.614
α=1.3	**31.69**	**0.987**	**3.585**

3. Reconstruction network

The two most significant research subjects in snapshot hyperspectral imaging are the coding mode and the reconstruction method. An efficient reconstruction algorithm can be applied to a variety of different snapshot hyperspectral imaging systems (such as CASSI, our system, and other phase-coded based hyperspectral imaging systems). In contrast to Jeon et al., we propose a reconstruction algorithm that is faster, requires less computation, but provides a better result. This section provides a detailed description of the structure of our reconstruction network.

3.1. Architecture of reconstruction network

The overall architecture of CAFormer is shown in Fig. 6. We adopt a U-shaped structure that consists of an encoder, a bottleneck, and a decoder. CAFormer is built up by channel-attention former module (CAFM) and GateFusion module. **Firstly**, we feed the coded image I into the model. Then a 3×3 *conv* (convolution with *kernel size* $= 3$) layer is used to map I into feature $I_0 \in \mathbb{R}^{H \times W \times C}$. **Secondly**, CAFormer exploits N_1 CAFMs, N_2 CAFMs, and N_3 CAFMs to generate hierarchical features. The downsample module is a 3×3 *conv* layer that down samples half scale the feature maps and doubles the channels. **Thirdly**, there is a bottleneck that consists of N_4 CAFMs. **Finally**, a structure opposite to the encoder is employed to decode these features. Output of our network is an HSI cube $\hat{I} \in \mathbb{R}^{H \times W \times C}$.

Optics EXPRESS

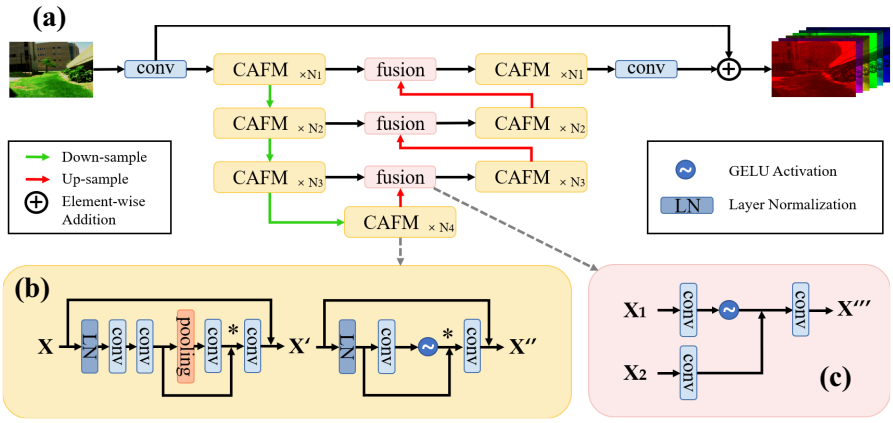

Fig. 6. (a) The overall architecture of CAFormer. (b) Structure of Channel-attention transformer module. (c) Structure of GateFusion module.

3.2. *Details of reconstruction network*

Convolutional neural networks have been shown to provide a faster and more accurate hyperspectral reconstruction than iterative optimization due to their time efficiency. With the development of neural networks, another form of neural network, transformer [37], was proposed. Transformer enhances the performance of recovery primarily by replacing the convolution with the more computationally intensive matrix multiplication within the convolutional neural network. Meanwhile, its internal connection structure is designed to make multiplication operations more efficient. Using transformer for CASSI hyperspectral reconstruction, Cai et al. [27,28] achieved the SOTA effect. We present a CAFormer network to reduce the computation of transformer and maintain the reconstruction performance at the same time.

Channel-attention Former Module. Recently researchers have come up with an idea that goes against the grain of human experience that the effectiveness of Transformer does not depend on the complexity of its computation, but rather on its overall structure. In this case, we suggest substituting a simple convolution operation for the complicated attention operation. It preserves the overall structure while reducing the calculation. In contrast, in the work of hyperspectral reconstruction, the difference from ordinary low-level tasks lies in that the spatial information of HSI cube is similar in different channels, and the spectral information between different channels is of greater significance. This property is in accordance with the channel attention operation. As shown in Fig. 6(b), we propose a CAFM to promote information recovery between spectral bands by replacing self-attention with channel-attention which greatly reduce the amount of computation. We developed the CAFormer network by stacking CAFMs with different parameters.

GateFusion. To strengthen the fusion of the same-scale features of the encoder and decoder, skip connections were used in the past to permit the decoder to borrow relevant information from the encoder. The importance of these pieces of information is different, and there will be some redundancy among them. A simple skip connection is not capable of sorting out relevant information from redundant information. So, we introduce a gate module as shown in Fig. 6(c). GateFusion can extract learnable characteristics from encoders and filter out unnecessary information, which enables computing resources to be used to reconstruct high-precision results.

Optics EXPRESS

4. Dataset and PSF calibration

We have fabricated the DOE designed by Jeon and the DOE designed in this article. Figure 1 illustrates the hyperspectral imaging process in the experiment. The first step is to install the processed DOE in front of the imaging detector and the target scene with spectral information is imaged by DOE. PSF encodes different wavelength information with different rotation directions. The coded information of all wavelengths is superimposed on the detector to obtain the hyperspectral coded image. Input the encoded image into a trained CAFormer to obtain the decoded hyperspectral information cube. This hyperspectral cube contains both the reflectivity of the target and the spectral information of the light. It is necessary to calibrate the spectrum of light for accurate reflectivity measurements of the target.

For image reconstruction, neural networks require a large amount of paired data. Nevertheless, when it comes to real scene, different systems have different pixel sizes, image resolution, and field of view, which makes it difficult to obtain a one-to-one correspondence between the coded image and the HSI cube. Thus, we propose to establish the training dataset utilizing real PSFs and simulated imaging process.

Figure 7 shows our device for calibrating PSFs. The point light source is simulated by passing a broad-spectrum light source through a small hole. The directional light tube helps to simulate a point light at infinity. Monochromatic light is simulated by placing a narrow band filter in front of the camera. We can calibrate the PSFs by imaging monochromatic point lights with different wavelengths at infinity with our imaging system. The encoded image simulation is carried out using the following formula [38]:

$$I(x,y) = \sum_{m=1}^{M} p_{\lambda_m}(x,y) * I_{\lambda_m}(x,y) + n, \tag{13}$$

where M is the number of spectral channels, I is the encoded image, p_λ and I_λ are the PSF and image at wavelength λ respectively, and n is the simulated image noise.

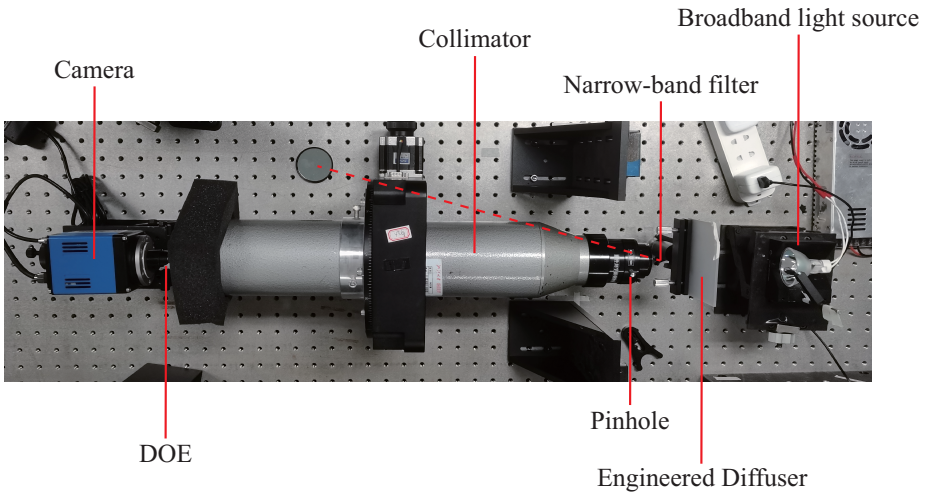

Fig. 7. Experimental device for calibrating PSF.

Simulation image I and its corresponding HSI cube $\boldsymbol{I_\lambda}$ constitute a pair of data. With enough training data, CAFormer is capable of decoding real images. A one-to-one correspondence exists between the PSF set, dataset, and network. It is necessary to rebuild the dataset and retrain the network when the PSF changes due to a change in DOE.

Optics EXPRESS

5. Experiments and results

This section describes detailed experiments to verify the effectiveness of our method. In Sec. 5.2, we implement simulation and real scene experiments to show that our designed equalization coded aperture can provide a distinct improvement. Sec. 5.3 demonstrates that our reconstruction algorithm is accurate and fast.

5.1. Implementation

Hardware Preparation. The RGB camera we use is pco.edge 5.5 with CIS2521 image sensor. Its pixel size is $6.5\mu m$, and the image resolution is 2560×2160. In the actual experiments of Sec. 5.2, we use the Gaia Field V10 hyper spectrometer to obtain reflection curves of targets and compare them with the reconstruction results of real encoded images to verify the performance of our system. We have processed Jeon's designed DOE ($\alpha = 1.0$) and our improved equalization designed DOE ($\alpha = 1.3$). The DOE pattern is etched on JGS2 quartz glass with a diameter of 5mm and a thickness of 0.6mm.

Dataset. Our hyperspectral imaging systems image at wavelengths between 460nm and 900nm with a spectral sampling rate of 10nm. Our simulation dataset is derived from the open source ICVL [39] dataset and the PSFs of the real system. ICVL dataset consists of 140 hyperspectral images with a spatial resolution around 1400×1300. We select 20 complete images as the test set, and the remaining 120 images are cut into 256×256 pixels as the training set. Our training set contains a total of 2950 data pairs. As described in Sec. 4 different PSF set need to be used to build different simulation datasets in different experiments. Sec. 5.2 uses simulation PSF sets (as shown in Fig. 4) for the simulation experiments and calibrated PSF sets (as shown in Fig. 10) for the actual experiments. In Sec. 5.3, we generate the simulation training dataset based on the actual PSF and hyperspectral data.

Training Configurations. Following the settings of MST++, all the models are trained with Adam optimizer (lr=0.0002, β_1= 0.9 and β_2 = 0.999) for 100 epochs and cosine schedule is used to decay the learning rate. Considering that PSF is directional in nature, we did not apply any data enhancement such as rotation to training data. Training is completed on an RTXA6000 GPU.

Evaluation Metrics. Generally, researchers evaluate the accuracy of RGB image reconstruction by comparing peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [40]. While in hyperspectral image reconstruction tasks, PSNR and SSIM are calculated separately for each channel and then averaged to determine the spatial resolution accuracy of the reconstructed images. In addition to spatial resolution, spectral accuracy is also an imperative evaluation criterion, which is usually gauged by spectral angle mapping (SAM) [41].

To assess the effectiveness of each method, we utilize the following three evaluation metrics to evaluate the reconstruction results: mean peak signal-to-noise ratio (MPSNR), mean structural similarity (MSSIM) and spectral angle mapping (SAM). Using the following formula, we can calculate the SAM of the spectral cube:

$$SAM = cos^{-1}\left(\frac{y^T\hat{y}}{\|y\|\ \|\hat{y}\|}\right) = cos^{-1}\left(\frac{y^T\hat{y}}{\sqrt{y^Ty}\sqrt{\hat{y}^T\hat{y}}}\right). \tag{14}$$

5.2. Validation of designed DOE

Here, we provide an innovative equalization designed phase-coding method. We design detailed simulation and actual experiments to test the effectiveness of this solution in the next subsection.

Simulation experiment. We simulate three DOEs with MATLAB, adding a processing error of within 10 nm to the DOE height function. Figure 4 illustrates the generated three sets of PSFs. Separately, we create three sets of simulation datasets for testing and training. Table 2 summarizes the reconstruction results. Gradually improving the long-band features will result in

Optics EXPRESS

incremental improvements in the quantitative image reconstruction index. According to Fig. 8, within 700nm wavelength, the quality of the long wavelength features hardly affect the image quality. But at 850nm and 900nm, only under the condition of sufficient features (α=1.3), the spectral accuracy and image quality can be maximized. When α=1.0, PSFs lead to a wrong spectrum but the image quality is not bad. When α=0.7, PSFs have worse results in both spectral accuracy and image quality. Figure 9 illustrates the difference in spectral accuracy clearly. The spectral accuracy at short wavelengths tends to be consistent, while the spectral accuracy at long wavelengths has a big difference. The equalization design we propose achieves the highest spectral accuracy.

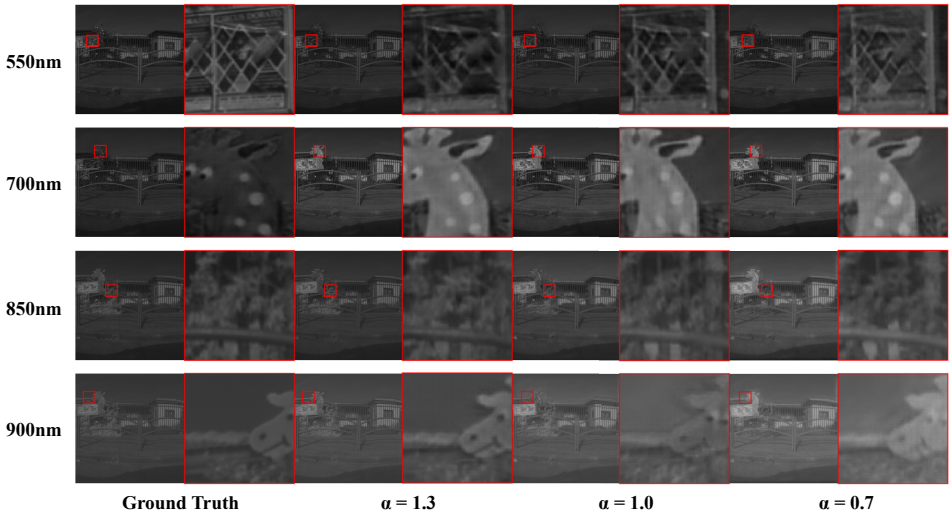

Fig. 8. Reconstructed simulation hyperspectral images of Scene 3 with 4 out of 45 spectral channels. 3 sets of PSFs lead to different results.

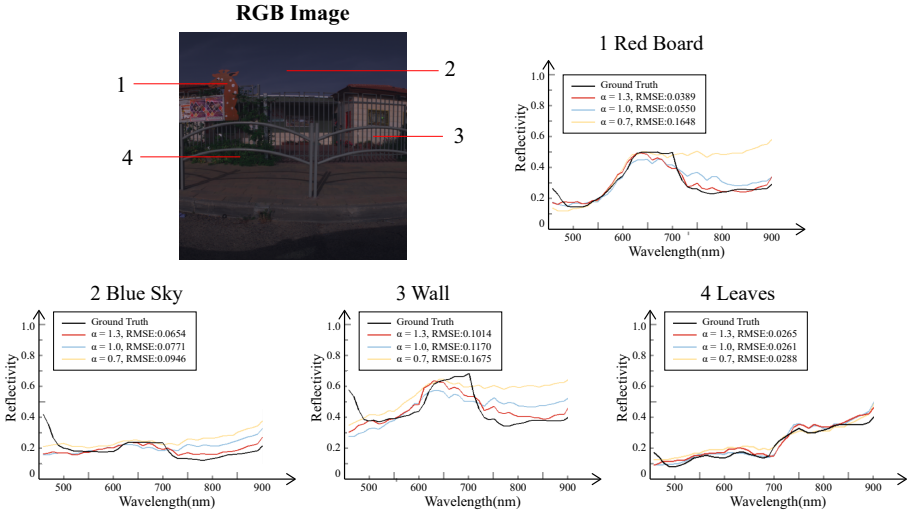

Fig. 9. Spectral curve of reconstructed simulation hyperspectral images of Scene 3. 3 sets of PSFs lead to different results.

Optics EXPRESS

Real scene experiment. We further apply our method to the reconstruction of real images. We etched the DOE designed by Jeon et al. (α = 1.0) and the DOE described in this paper (α = 1.3) on JGS2 quartz glass with a diameter of 5mm and a thickness of 0.6mm. In Fig. 10, real PSFs generated from these two DOEs are calibrated. Figure 11 shows the reconstruction results on real images. As for image quality, the DOE we designed is superior in terms of clarity and detail. In comparison to a commercial spectrometer, our system yields a more accurate spectral curve.

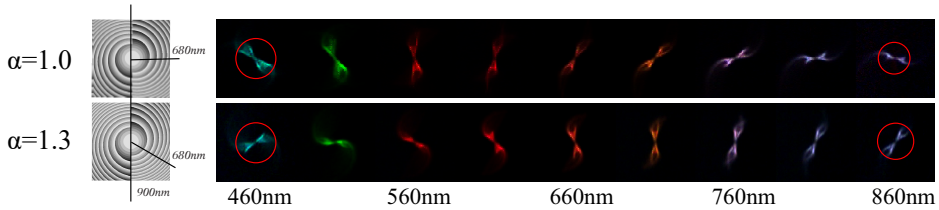

Fig. 10. Calibrated real PSFs. The red circle is the manually marked circumcircle of PSF.

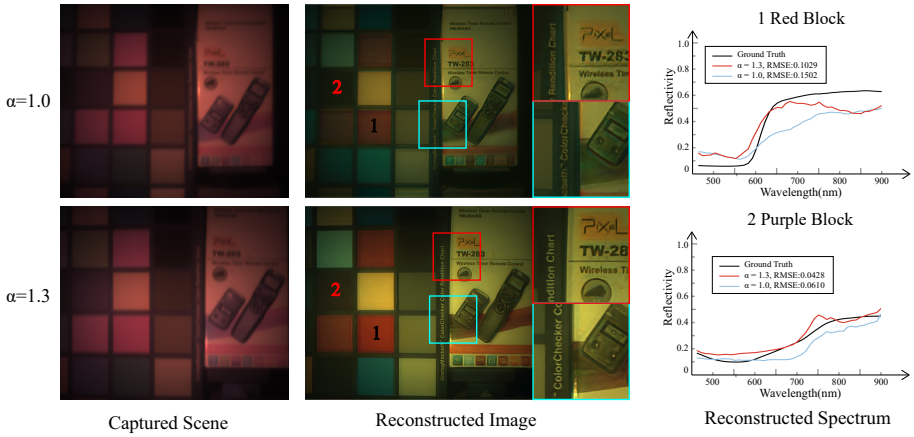

Fig. 11. Reconstruction results on real scenes.

5.3. *Validation of designed network*

Our proposed algorithm for coded hyperspectral reconstruction is superior to both the reconstruction algorithm proposed by Jeon and the existing coded hyperspectral reconstruction solution in terms of computational complexity and reconstruction accuracy. In this subsection, we use real PSFs (as shown in Fig. 10) to generate simulation data sets. We demonstrate the advantages and effectiveness of our algorithm by reconstructing simulation images.

GateFusion Module. We suggest using GateFusion module instead of simple concatenation for the jump connections in a u-shaped network. By filtering the pre-features and subsequently absorbing the post-features for reference, the gate module can facilitate mutual learning among features. Table 3 shows the ablation results of the GateFusion module. As can be seen from the data, the GateFusion module improves image quality and spectral accuracy to a certain extent. Hence, GateFusion can provide better assistance in screening spectral information, confirming our hypothesis. We consider it a good alternative to concatenation fusion, given its low overhead, which provides a reference for other spectral reconstruction tasks.

Optics EXPRESS

Table 3. Ablation experiment results of GateFusion module.

Network	GateFusion	MPSNR↑	MSSIM↑	SAM↓
CAFormer	-	29.48	0.961	6.474
CAFormer	✓	**29.54**	0.961	**6.196**

Quantitative Comparison. We quantitatively compare the performance of CAFormer and other works (HINet [29], MIRNet [42], MST++ [28]) and the results are shown in Table 4 and Fig. 12. Here we put the best results in bold and bold. Among these works, MST++ used to be the most efficient network, won a number of hyperspectral reconstruction competitions [28]. Overall, our proposed CAFormer outperformed these works with lower computation. As shown in Table 4, ours FLOPs is less than the 3 SOTA networks, which means our algorithm is three times faster than MST++. In contrast to other algorithms, the PSNR of our reconstructed image is higher with the minimum computational resources. And our reconstructed spectral is more accurate. All algorithms share the same dataset, which demonstrates the benefits of our network.

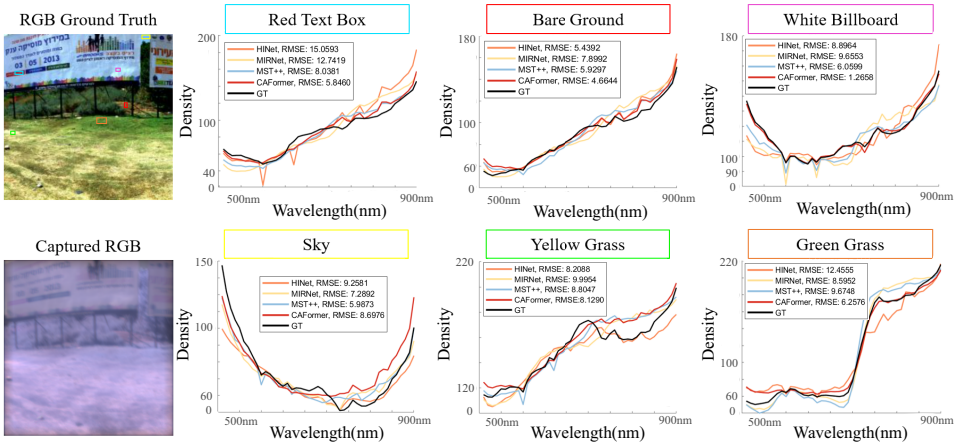

Fig. 12. We select some areas to output the spectral curve reconstructed by 3 SOTA algorithms, HINet and our CAFormer. Please zoom in for better visualization performance.

Table 4. Quantitative experiment results of 3 SOTA works and ours CAFormer.

	MPSNR↑	MSSIM↑	SAM↓	Params↓	FLOPs↓
HINet	22.92	0.903	9.445	11M	70.84G
MIRNet	26.25	0.933	7.862	**3.17*M***	60.81G
MST++	29.33	0.955	6.265	3.37M	48.23G
Ours	**29.54**	**0.961**	**6.196**	5.76M	**17.42*G***

Qualitative Comparison. We also select some samples of the results to analyze the performance of each method qualitatively. Fig. 13 illustrate qualitative comparisons of our CAFormer with some representative learning-based methods. We select one sample with more colors and multiple image information taken from the test data set to evaluate the networks' performance. Our network perform better than HINet and MIRNet visibly. In terms of the image as a whole, we're about as good as MST++. But by zooming in on the details, our image information is much more accurate. In the case of insufficient features at the long wavelength bands, MST++

Optics EXPRESS

will show wrong images (orange boxes) in images at 850nm and 900nm, but our image recovery is correct.

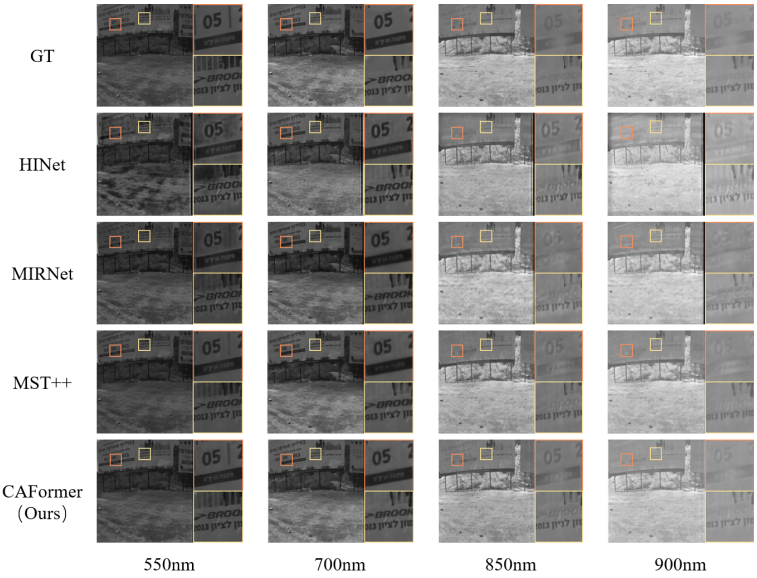

Fig. 13. Reconstructed simulation hyperspectral information comparisons of Scene 1 with 45 spectral channels from 460nm to 900nm.

We also compared our method with the iterative reconstruction method proposed by Jeon using a hyperspectral dataset captured by Xu [43]. The reconstruction results are shown in Fig. 14. Our method has advantages in both image quality and spectral accuracy.

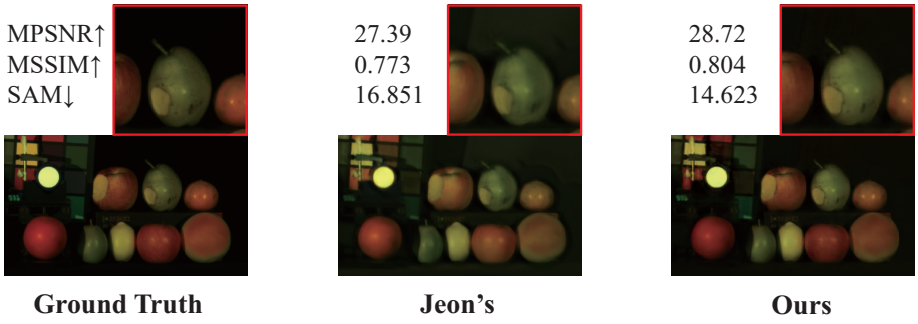

Fig. 14. Reconstructed simulation hyperspectral information comparisons between ours and Jeon's method.

6. Conclusion

This project was undertaken to design an overall imaging framework superior to existing SCI technology. We proposed a better reconstruction algorithm and implemented a complete hyperspectral imaging process. Using our designed DOE, spectral information can be better encoded. The most relevant aspect of the paper is that we propose a state-of-the-art (SOTA) algorithm for reconstructing hyperspectral images from RGB images. Eventually, to calibrate our system, we propose a method based on the comparison of simulation data and real data. We

Optics EXPRESS

have proposed a method which is superior to the existing technology in terms of both imaging system and algorithm, putting snapshot compact hyperspectral technology closer to a practical application. Additionally, there are some degradation processes in the real shooting data that are different from those in the simulation data. This causes the real shooting performance to be worse than the simulation performance. This problem affects all RGB images and HSI cube technologies. We hope that in the further research, real-shot data sets or more realistic simulation data sets will be established in order to achieve better real shooting performance.

Funding. Civil Aerospace Pre-Research Project (No. D040104); National Natural Science Foundation of China (No. 62275229); Key Research Project of Zhejiang Lab (No.2021 MH0AC01).

Acknowledgments. We thank Meijuan Bian from the facility platform of optical engineering of Zhejiang University for instrument support.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. S. A. Macenka and M. P. Chrisp, "Airborne visible/infrared imaging spectrometer (aviris) spectrometer design and performance," Int. Soc. for Opt. Photonics. **834**, 32–43 (1987).
2. P. Vermeulen, P. Flémal, O. Pigeon, P. Dardenne, J. Fernández Pierna, and V. Baeten, "Assessment of pesticide coating on cereal seeds by near infrared hyperspectral imaging," J. Spectr. Imaging. **6**, a1 (2017).
3. P. Hatfield and P. Pinter, "Remote sensing for crop protection," Crop Protection. **12**(6), 403–413 (1993).
4. E. Ciurczak and B. Igne, *Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy* (2014).
5. K. Masood and N. Rajpoot, "Texture based classification of hyperspectral colon biopsy samples using clbp," Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1011–1014 (2009).
6. A. Gowen, C. O'Donnell, P. Cullen, G. Downey, and J. Frías, "Hyperspectral imaging-an emerging process analytical tool for food quality and safety control," Trends Food Sci. & Technol. **18**(12), 590–598 (2007).
7. S. Lohumi, S. Lee, H. Lee, and B.-K. Cho, "A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration," Trends Food Sci. & Technol. **46**(1), 85–98 (2015).
8. S. L. Ustin and J. A. Gamon, "Remote sensing of plant functional types," New Phytol. **186**(4), 795–816 (2010).
9. R. Lucke, M. Corson, N. McGlothlin, S. Butcher, and D. Wood, "The hyperspectral imager for the coastal ocean (hico): fast build for the iss," Remote. Sens. Syst. Eng. **7813**, 78130D (2010).
10. L. Huang, R. Luo, X. Liu, and X. Hao, "Spectral imaging with deep learning," Light: Sci. Appl. **11**(1), 61 (2022).
11. D. Brady and M. Gehm, "Compressive imaging spectrometers using coded apertures," Vis. Inf. Process. XV. **6246**, 62460A (2006).
12. M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," Opt. Express **15**(21), 14013–14027 (2007).
13. Z. Zhu, H. Liu, J. Hou, H. Zeng, and Q. Zhang, "Semantic-embedded unsupervised spectral reconstruction from single rgb images in the wild," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 2279–2288 (2021).
14. Y. Zhao, L.-M. Po, Q. Yan, W. Liu, and T. Lin, "Hierarchical regression network for spectral reconstruction from rgb images," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. pp. 1695–1704 (2020).
15. J. Li, C. Wu, R. Song, Y. Li, and F. Liu, "Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* pp. 1894–1903 (2020).
16. K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, "Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array," Optica. **7**(10), 1298–1307 (2020).
17. 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. **38**(4), 1–13 (2019).
18. H. Arguello, S. Pinilla, Y. Peng, H. Ikoma, J. Bacca, and G. Wetzstein, "Shift-variant color-coded diffractive spectral imaging system," Optica. **8**(11), 1424–1434 (2021).
19. Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, "Phasecam3d — learning phase masks for passive single view," *2019 IEEE International Conference on Computational Photography (ICCP)*. (2019).
20. S.-H. Baek, H. Ikoma, D. S. Jeon, Y. Li, W. Heidrich, G. Wetzstein, and M. H. Kim, "Single-shot hyperspecal depth imaging with learned diffractive optics," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 2651–2660 (2021).
21. X. Dun, H. Ikoma, G. Wetzstein, Z. Wang, X. Cheng, and Y. Peng, "Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging," Optica. **7**(8), 913–922 (2020).

Optics EXPRESS

22. J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 10192–10201 (2019).
23. S. Chen, T. Lin, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Computational optics for mobile terminals in mass production," IEEE Trans. on Pattern Anal. Mach. Intell. **45**(4), 4245–4259 (2023).
24. C. V. Correa, H. Arguello, and G. R. Arce, "Snapshot colored compressive spectral imager," J. Opt. Soc. Am. A **32**(10), 1754–1763 (2015).
25. H. Rueda, H. Arguello, and G. R. Arce, "Compressive spectral testbed imaging system based on thin-film color-patterned filter arrays," Appl. Opt. **55**(33), 9584–9593 (2016).
26. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," Optica. **5**(1), 1–9 (2018).
27. Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. Van Gool, "Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction," arXiv, arXiv.2111.07910 (2021).
28. Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, and L. Van Gool, "Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," arXiv, arXiv.2204.07908 (2022).
29. H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 12294–12305 (2021).
30. J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. pp. 1833–1844 (2021).
31. M. Kumar, D. Weissenborn, and N. Kalchbrenner, "Colorization transformer," *International Conference on Learning Representations*. (2021).
32. F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 5790–5799 (2020).
33. W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, and S. Yan, "Metaformer is actually what you need for vision," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 10819–10829 (2022).
34. L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," arXiv, arXiv.2204.04676 (2022).
35. H. Hu, H. Zhou, Z. Xu, Q. Li, H. Feng, Y. Chen, T. Jiang, and W. Xu, "Practical snapshot hyperspectral imaging with doe," Opt. Lasers Eng. **156**, 107098 (2022).
36. S. Chen, H. Feng, D. Pan, Z. Xu, Q. Li, and Y. Chen, "Optical aberrations correction in postprocessing using imaging simulation," ACM Trans. Graph. **40**(5), 1–15 (2021).
37. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16×16 words: Transformers for image recognition at scale," arXiv, arXiv:2010.11929 (2020).
38. S. Chen, H. Feng, K. Gao, Z. Xu, and Y. Chen, "Extreme-quality computational imaging via degradation framework," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, (2021), pp. 2612–2621.
39. B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," Computer Vision - ECCV 2016. pp. 19–34 (2016).
40. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Process. **13**(4), 600–612 (2004).
41. F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (sips)-interactive visualization and analysis of imaging spectrometer data," Remote. Sens. Environ. **44**(2-3), 145–163 (1993).
42. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," *European Conference on Computer Vision*. pp. 492–511 (2020).
43. H. Xu, H. Hu, S. Chen, Z. Xu, Q. Li, T. Jiang, and Y. Chen, "Hyperspectral image reconstruction based on the fusion of diffracted rotation blurred and clear images," Opt. Lasers Eng. **160**, 107274 (2023).