Chip

Research Article

ARTIFICIAL INTELLIGENCE & BIG DATA

DOI: 10.1016/j.chip.2023.100045

Deep-learning based on-chip rapid spectral imaging with high spatial resolution

Jiawei Yang[1,2], Kaiyu Cui[1,2,*], Yidong Huang[1,2,3,*], Wei Zhang[1,2,3], Xue Feng[1,2] & Fang Liu[1,2]

[1]*Department of Electronic Engineering, Tsinghua University, Beijing 100084, China* [2]*Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China* [3]*Bejing Academy of Quantum Information Science, Beijing 100084, China*

E-mails: kaiyucui@tsinghua.edu.cn (Kaiyu Cui), yidonghuang@tsinghua.edu.cn (Yidong Huang)

Cite as: Yang, J. et al. Deep-learning based on-chip rapid spectral imaging with high spatial resolution. *Chip* **2**, 100045 (2023).
https://doi.org/10.1016/j.chip.2023.100045

Received: 23 November 2022

Accepted: 24 March 2023

Published online: 7 April 2023

Spectral imaging extends the concept of traditional color cameras to capture images across multiple spectral channels and has broad application prospects. Conventional spectral cameras based on scanning methods suffer from the drawbacks of low acquisition speed and large volume. On-chip computational spectral imaging based on metasurface filters provides a promising scheme for portable applications, but endures long computation time due to point-by-point iterative spectral reconstruction and mosaic effect in the reconstructed spectral images. In this study, on-chip rapid spectral imaging was demonstrated, which eliminated the mosaic effect in the spectral image by deep-learning-based spectral data cube reconstruction. The experimental results show that 4 orders of magnitude faster than the iterative spectral reconstruction were achieved, and the fidelity of the spectral reconstruction for the standard color plate was over 99% for a standard color board. In particular, video-rate spectral imaging was demonstrated for moving objects and outdoor driving scenes with good performance for recognizing metamerism, where the concolorous sky and white cars can be distinguished via their spectra, showing great potential for autonomous driving and other practical applications in the field of intelligent perception.

Keywords: Spectral imaging, Deep learning, Metasurface

INTRODUCTION

Spectral imaging technology aims to capture spectral information for each two-dimensional spatial point to form a spectral data cube. It has been applied in a broad range of fields such as remote sensing[1,2], precision agriculture[3], medical diagnostics[4,5], food inspection[6], environmental monitoring[7], art conservation[8,9] and astronomy[10]. Traditional spectral imagers rely on either spatial scanning, such as whiskbroom scanning[11] and push-broom scanning[12], or spectral scanning, such as filter wheels[13] and tunable filters[14,15]. However, scanning methods suffer from low acquisition speed, which is not applicable for dynamic recording of moving targets. To overcome this limitation, snapshot spectral imaging methods[16] are explored. Early snapshot techniques, such as integral field spectrometry[17-19], multi-spectral beam splitting[20], and image-replicating imaging spectrometer[21], still rely on light splitting, also the optical systems of which are bulky. With the development of compressive sensing (CS)[22,23], growing research interests have been attracted by the computational snapshot spectral imaging technique[24], which can be categorized into three groups: coded aperture, speckle-based and spectral filter array methods. For coded aperture methods, the classical system is coded aperture snapshot spectral imager (CASSI)[25-32], which uses fixed masks and dispersive elements to implement band-wise modulation. CASSI is capable of capturing and reconstructing the hyperspectral images rapidly with deep-learning techniques. However, the complicated optical components lead to large system volume, which cannot meet the growing demand for portable applications. Speckle-based methods[33-38] utilize the wavelength dependence of speckle from a scattering media or diffractive optical element to achieve spectral imaging. Although the systems can be compact, the spectral resolution is limited by the speckle correlation through wavelengths. The spectral filter array methods can be viewed as an extension of Bayer filters, which adopt a super-pixel containing a group of spectral filters for spectral recovery. Even though the methods of this class are endowed with the advantages of compact device size and high spectral accuracy, there exist mosaic effect in the reconstructed spectral images, where the recovered spectra for the edge points are inaccurate. Recently, our group demonstrated a snapshot spectral imaging chip based on metasurface-based spectral filter arrays with the ultra-high center-wavelength accuracy of 0.04 nm and the spectral resolution of 0.8 nm[39]. Furthermore, the spectral resolution was improved to 0.5 nm when adopting metasurfaces with freeform shaped meta-atoms in our latest work[40]. However, in addition to the Mosaic effect mentioned above, the classical iterative CS algorithm adopted in the current work makes the computation time of data cube reconstruction still remain very long, which limits its application in the mobile systems with speed requirements, such as pilotless automobile.

In the current work, on-chip mosaic-free rapid spectral imaging was demonstrated by employing advanced deep-learning-empowered algorithms developed for CASSI to the metasurface spectral imager reported in our previous work[40]. The metasurface spectral imager produces different amplitude modulation patterns in different spectral bands, which plays the role of a fixed mask plus a disperser in CASSI. Specifically, the spectral imager was designed by integrating a metasurface layer com-

DOI: 10.1016/j.chip.2023.100045

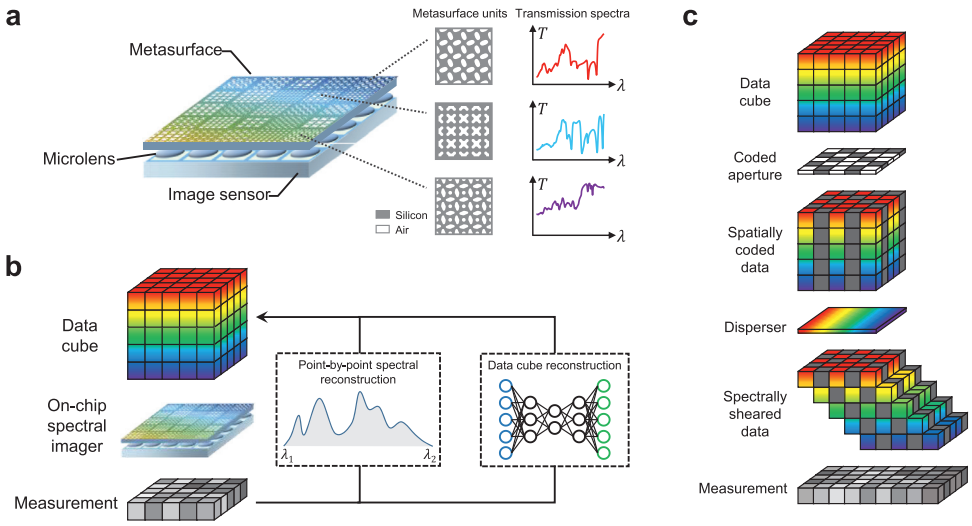

Fig. 1 | a, Schematic representation of the metasurface-based spectral imager, which consists of a metasurface layer, a microlens layer and an image sensor layer. The silicon metasurface contains 360 × 440 metasurface units with freeform shaped meta-atoms. There are 400 types of metasurface units with distinctive transmission spectra. Typically, 5 × 5 metasurface units are combined to form a micro-spectrometer (spectral pixel). **b,** Two methods of data cube reconstruction from the measurement of the spectral imager in a including the point-by-point spectral reconstruction using iterative optimization algorithms, and fast reconstruction of the whole data cube via deep learning algorithms. **c,** Schematic diagram of coded aperture snapshot spectral imaging. The spectral data cube is first modulated by a fixed coded aperture (mask), then sheared by a disperser, and finally measured by a detector.

posed of 360 × 440 metasurface units with freeform shaped meta-atoms onto a CMOS image sensor (CIS). There are totally 400 kinds of metasurface units, each of which is of a distinctive spectral response function. As a proof of principle, 256 × 256 metasurface units and 26 wavelength bands were selected from 450 to 700 nm with a step of 10 nm. A deep unfolding network based on the alternating direction method of multipliers (ADMM) algorithm, which is called ADMM-net[41], was adopted for the fast reconstruction of spectral images. Here, the network was trained on a synthetic dataset containing 750,000 spectral data cubes with the size of 256 × 256 × 26, which are generated from the CAVE[41] and KAIST[42] datasets. Besides, additive white Gaussian noise was imposed on the measurements to mimic real-test cases. On-chip rapid spectral imaging eliminating the mosaic effect was realized by applying the ADMM-net to reconstruct the spectral data cube directly. Compared with the point-by-point iterative spectral reconstruction, four orders of magnitude speed improvement was achieved by the ADMM-net, which enables a spectral data cube reconstruction rate of 55 per second, and the average spectral reconstruction fidelity exceeds 99% for a 24-patch Macbeth color checker. In practice, video-rate spectral image reconstruction was demonstrated for moving objects and outdoor driving scenes. It is found that the concolorous sky and moving white cars can be effectively distinguished by their spectra, while the existing driverless vehicles can easily mistake a white truck for the sky and cause a crash. The approach adopted in the current work is capable of solving the huge safety problem caused by the defect in metamerism recognition for not only autonomous driving[43] but also other fields of intelligent perception, and shows great potential for various applications.

HARDWARE STRUCTURE

The schematic of the metasurface-based spectral imager is shown in Fig. 1a, as reported in ref.[40]. A silicon metasurface layer was integrated on the image sensor with a microlens layer. The metasurface layer was composed of 360 × 440 metasurface units, which were obtained by arranging 20 × 20 kinds of metasurface units repeatedly 18 × 22 times. Different metasurface units exhibit distinctive transmission spectra. Each metasurface unit is a periodic array with freeform shaped meta-atoms. The periods and shapes of the corresponding 400 kinds of meta-atoms were optimized to maximize the mutual differences of the transmission spectra (see ref.[40] for details). Each metasurface unit represents a spatial point, and light will be detected after being modulated by each unit. For a certain point (i, j), the spectrum of incident light can be reconstructed from the $N = (2n + 1)^2$ detected signals at the surrounding N points (that is, $(i - n, j - n)$, $(i - n,\ j - n + 1)$, …, $(i + n, j + n)$, typically, we set $n = 2$). The spectral reconstruction was implemented by solving such a system of linear equations:

$$\begin{bmatrix} y^{[i-n,j-n]} \\ \vdots \\ y^{[i+n,j+n]} \end{bmatrix} = \begin{bmatrix} M_1^{[i-n,j-n]} & M_2^{[i-n,j-n]} & \cdots & M_{N_\lambda}^{[i-n,j-n]} \\ \vdots & \vdots & \ddots & \vdots \\ M_1^{[i+n,j+n]} & M_2^{[i+n,j+n]} & \cdots & M_{N_\lambda}^{[i+n,j+n]} \end{bmatrix} \begin{bmatrix} x_1^{[i,j]} \\ x_2^{[i,j]} \\ \vdots \\ x_{N_\lambda}^{[i,j]} \end{bmatrix} \quad (1)$$

Where, $y^{[i,j]}$ denotes the measured signal at the (i, j) point, $M_k^{[i,j]}$ represents the modulation intensity (i.e., transmittance of the metasurface unit) at the (i, j) point for the k-th spectral channel, $x_k^{[i,j]}$ is the k-th element of the target spectrum vector at the (i, j) point, N_λ is the total number of spectral channels. Since there are a total of 360 × 440 points, 158,400 groups of equations like Eq. (1) are needed to be solved to reconstruct the spectra of all points to recover the whole spectral data cube, which is time-consuming. For the iterative CS algorithm, it is assumed that the spectra of incident light at the N points around (i, j) are the same with those in Eq. (1), which results in mosaic effect in the reconstructed spectral images. In order to address the above problems, the deep-learning algorithms[44] were proposed to be exploited for directly reconstructing the data cube inspired by CASSI (see Supplement 1, S1 for details), as indicated in Fig. 1b.

The basic principle of CASSI is presented in Fig. 1c, where the spectral data cube is spatially coded by a fixed physical mask (coded aperture), and then spectrally sheared by a dispersive element, and finally measured

DOI: 10.1016/j.chip.2023.100045

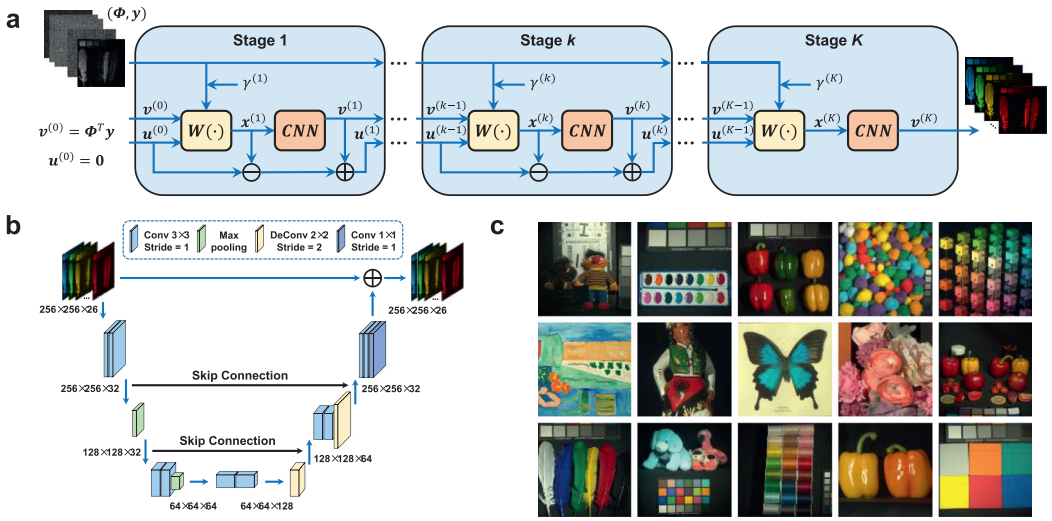

Fig. 2 | a, Data flow graph of the ADMM-net with K stages, where each stage contains a linear projection $W(\cdot)$ denoting the computation in Eq. (4) and a *CNN* denoiser. **b,** Architecture of the U-net used in a for denoising. **c,** RGB images of 15 samples in the basic spectral image dataset.

by a detector. Therefore, for CASSI, a fixed mask and a disperser were used to achieve different masks at different spectral channels. The fixed mask can be traditional blocking-unblocking coded aperture[25,26] or colored coded aperture[27-30]. In the current work, the metasurface layer was adopted to achieve different masks at different wavelengths, and its mathematical model can be written as:

$$\begin{bmatrix} y^{[1,1]} \\ y^{[2,1]} \\ \vdots \\ y^{[N_x,N_y]} \end{bmatrix} = \left[\begin{array}{cccc|c|cc} M_1^{[1,1]} & 0 & \cdots & 0 & \cdots & M_{N_\lambda}^{[1,1]} & 0 \\ 0 & M_1^{[2,1]} & \cdots & 0 & \cdots & 0 & M_{N_\lambda}^{[2,1]} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & M_1^{[N_x,N_y]} & \cdots & 0 & 0 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_\lambda} \end{bmatrix} \tag{2}$$

Where, $x_k = [\ x_k^{[1,1]} \quad x_k^{[2,1]} \quad \cdots \quad x_k^{[N_x,N_y]}\ ]^T$ denotes the k-th frame of the spectral data cube, N_x, N_y represent the number of points along the horizontal and vertical dimensions, respectively. When taking the measurement noise into consideration, the Eq. (2) can be expressed in a vectorized formulation as follows:

$$y = \Phi x + e \tag{3}$$

Where, $y \in \mathbb{R}^{N_x N_y}$ is the compressed measurement, $\Phi \in \mathbb{R}^{N_x N_y \times N_x N_y N_\lambda}$ is the sensing matrix, $x \in \mathbb{R}^{N_x N_y N_\lambda}$ is the target signal and $e \in \mathbb{R}^{N_x N_y}$ is the measurement noise. As a proof of principle, 256 × 256 metasurface units were selected for spectral imaging with 26 wavelengths from 450 nm to 700 nm (that is, $N_x = N_y = 256,\ N_\lambda = 26$).

RECONSTRUCTION NETWORK

In the current work, a deep unfolding network based on the ADMM algorithm dubbed ADMM-net was employed for data cube reconstruction. The framework of ADMM-net with K stages ($K = 12$) is depicted in Fig. 2a. As in ref.[41], let v denote an estimate of the desired signal, and by introducing two auxiliary variables x, u, the three steps for updating variables in each stage are listed as follows:

$$x^{(k)} = \left[\Phi^T\Phi + \gamma^{(k)} I\right]^{-1}\left[\Phi^T y + \left(v^{(k-1)} + u^{(k-1)}\right)\right] \tag{4}$$

$$v^{(k)} = \mathcal{D}_k\left(x^{(k)} - u^{(k-1)}\right) \tag{5}$$

$$u^{(k)} = u^{(k-1)} - \left(x^{(k)} - v^{(k)}\right) \tag{6}$$

Here, $\gamma^{(k)} > 0$ is another auxiliary trainable parameter, I is an identity matrix, D_k denotes a denoiser. Eq. (4) represents a linear projection. Since Φ is a concatenation of N_λ diagonal matrices, it can be seen that $\Phi\Phi^T$ is a diagonal matrix. Therefore, Eq. (4) can be solved efficiently via element-wise operation instead of calculating the huge matrix inversion as derived in ref.[45]. Eq. (5) is a denoising process performed by the *CNN*, as shown in Fig. 2a. A 15-layer U-net[46] was adopted as the denoiser, and its architecture is described in Fig. 2b, where the skip-connection can be regarded as residual learning, which is shown to be necessary for the denoiser[41].

A basic dataset containing 262 scenes with a size of 512 × 512 × 26 was constructed from the publicly available hyperspectral dataset CAVE and KAIST (see Supplement 1, S2 for details). The RGB images of the selected 15 scenes in the basic dataset, which are converted from the spectral images via the International Commission on Illumination color-matching function[48], are shown in Fig. 2c. For model training, 252 scenes were randomly selected from the basic dataset, implementing data augmentation to obtain 750,000 samples with the size of 256 × 256 × 26. The operations of data augmentation include random cropping, rotation and multiplying with the spectra of LED light or sunlight sources (see Supplement 1, S3 for details). The remaining 10 scenes were downsampled to the size of 256 × 256 × 26 for testing. In addition, the loss function is the root mean square error between the ground truth and the output result of the network. The network was trained by the Adam optimizer on Pytorch using NVIDIA GeForce RTX 3080 GPUs. The total number of epochs is 300, and the batch size was set to 4 due to the limitation of GPU memory. The initial learning rate was set to 0.001, and scaled to 90% of the previous one every 20 epochs.

Firstly, network training and test were conducted under noisy conditions. Since it takes a lot of time to acquire measured images for real objects to train the network, the method of simulated measurement

DOI: 10.1016/j.chip.2023.100045

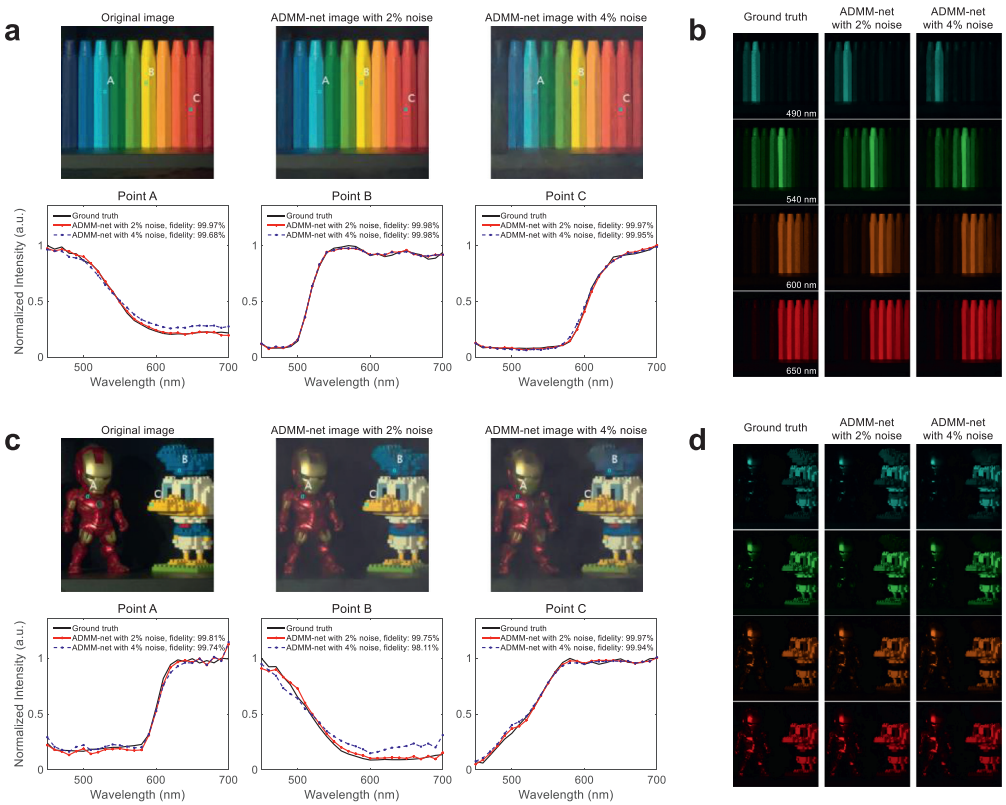

Fig. 3 | a-c, Reconstruction results of simulated *crayon* and *toy* hyperspectral data under 2% and 4% noise levels, respectively. The original and reconstructed RGB images of the scenes are shown on the top row with a size of 256 × 256 pixels. The spectra of three selected points are shown on the bottom row, where the fidelity of the reconstructed spectra is shown in the legends. **b-d,** Reconstructed frames of simulated *crayon* and *toy* hyperspectral data under 2% and 4% noise levels, respectively.

was adopted, that is, using the calculated measured images y through Eq. (3) with pre-calibrated Φ, spectral images x in the dataset and suitable noise distribution e. Both x and y were normalized to $[0, 1]$. It is assumed that each element of the measurement noise vector e in Eq. (3) follows an independent zero-mean Gaussian distribution[47], that is, $e_i \sim \mathcal{N}(0, \sigma_n^2)$. Here, the standard deviation σ_n represents the noise level, and it is randomly chosen between 0 and 0.05 to increase the robustness to noise of different levels. The reconstruction results of two test scenes under 2% and 4% noise levels are given in Fig. 3. For the *crayon* scene with 2% noise, it can be seen that the average fidelity of the recovered spectra for the selected three points exceeds 99.97%, as indicated in the legends of Fig. 3a. When the noise level is increased to 4%, it can be seen that the ADMM-net can still recover the spectra reliably with an average fidelity of 99.87%. Here the fidelity is defined as follows:

$$F(f_1, f_2) = \langle f_1, f_2 \rangle \tag{7}$$

where f_1, f_2 are the l_2-normalized ground truth and reconstructed result, respectively, and $\langle\rangle$ represents the inner product. The four reconstructed exemplar frames of the *crayon* hyperspectral data are shown in Fig. 3b, which are highly consistent with the ground truth under different noise levels. For the *toy* scene, the ADMM-net can also provide accurate reconstructed results, as shown in Figs. 3c and 3d. When the noise level is increased from 2% to 4%, the spectral reconstruction quality is degraded more seriously compared with the *crayon* scene since there are more fine spatial details in Fig. 3c, while the four frames in Fig. 3d are still recovered with high quality.

RAPID SPECTRAL IMAGING

Then, the ADMM-net trained with noise was applied to reconstruct the real data from snapshot measurements captured by the spectral imager, as shown in Fig. 4a. The metasurface layer with the size of 8 mm × 6.4 mm was integrated on top of a CIS (Thorlabs, CS235MU). The spectral imager is assembled with a 50 mm-length fixed focal lens (Thorlabs, MVL50M23) for imaging. The recovered results for a standard 24-patch Macbeth color checker are displayed in Fig. 4b. In order to better show the reconstruction performance, a commercial spectral camera (Dualix Instruments, GaiaField Pro-V10) based on line scanning was employed to capture the spectral image as a reference. Moreover, the results of ADMM-net were compared with those obtained by traditional iterative algorithm GAP-TV[45], end-to-end CNN λ-net[48] and point-by-point spectral reconstruction using the iterative CS algorithm implemented by CVX[48], an open-source package for convex optimization. For CVX method, 25 metasurface units ($N = 25$) in Eq. (1) were adopted for spectral reonstruction with 601 wavelength channels (from 450 to 750 nm with a step of 0.5 nm), and then the result was downsampled to 26 channels (from 450 to 700 nm with a step of 10 nm). The CVX method used l_1-norm regularization based on sparsity in the spectral transformation domain via dictionary learning. It can be clearly seen from Fig. 4b that higher spectral reconstruction accuracy is achieved by ADMM-net than other algorithms. For the selected four points, the average fidelities of the reconstructed spectra using ADMM-net, GAP-TV, λ-net and CVX are 99.53%, 97.18%, 97.72% and 97.32%, respectively. Besides, ADMM-net is also superior to the CVX in spatial details, which eliminates the mosaic effect. Additional reconstruction results for a Thorlabs box are provided in Supplement 1, S4.

DOI: 10.1016/j.chip.2023.100045

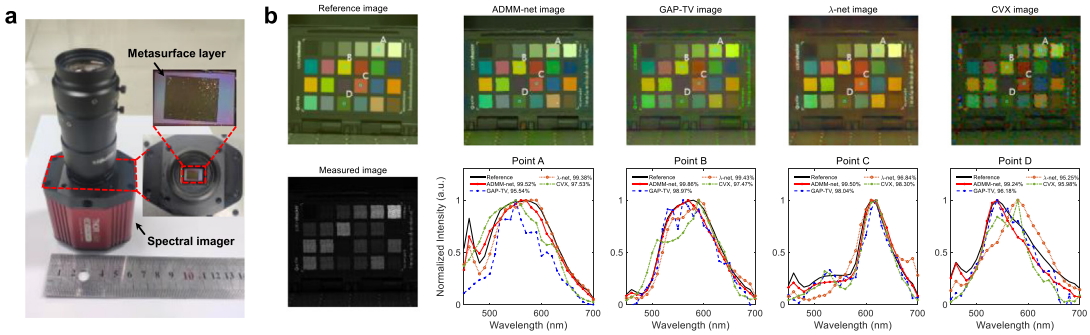

Fig. 4 | a, The spectral imager with a metasurface layer (8×6.4 mm^2) integrated on a CMOS image sensor (Thorlabs, CS235MU) and a lens with fixed focal length (Thorlabs, MVL50M23) are used for spectral imaging. **b,** Reconstruction results of real hyperspectral data of a 24-patch Macbeth color chart. The reference RGB image captured by a commercial spectral camera (Dualix Instruments, GaiaField Pro-V10), the reconstructed RGB images using ADMM-net, GAP-TV, λ-net and CVX, are shown on the top row from left to right with a size of 256 × 256 pixels. The snapshot measurement and spectra of four selected points are shown on the bottom row, where the fidelity of the reconstructed spectra is shown in the legends.

Table 1 | Comparison of different methods.

Methods	Line scanning	ADMM-net	GAP-TV	λ-net	CVX	
Data cube size	256 × 256 × 26				256 × 256 × 601	256 × 256 × 26
Running time (s)	~60	1.72 @CPU0.018 @GPU	110 @CPU	2.44 @CPU0.095 @GPU	7767 @CPU	4854 @CPU
Running time per channel (s)	~2.31	0.066 @CPU0.00069 @GPU	4.23 @CPU	0.094 @CPU0.0037 @GPU	12.9 @CPU	186.7 @CPU

Fig. 5 | Reconstructed results of real hyperspectral data of a moving Thorlabs box. Snapshot measurements captured by the on-chip spectral imager are shown on the first row. The reconstructed RGB images are shown on the second row. The reconstructed spectral images at different wavelengths are shown on the third to sixth rows.

In addition, the running time was compared using different methods so as to obtain a spectral data cube, as displayed in Table 1. For the commercial spectral camera based on line scanning, it takes about 1 minute to capture a data cube with a size of 256 × 256 × 26. For the computational methods, about 110 s are required for the GAP-TV to recover the spectral data cube of size 256 × 256 × 26, since it is an iterative method. The ADMM-net and λ-net spend It takes 1.72 and 2.44 s for the ADMM-net and λ-net on CPU (Intel Xeon Gold 6226R), or 0.018 and 0.095 s on GPU (NVIDIA GeForce RTX 3080), respectively. The CVX is the slowest, which takes 7767 and 4854 s for the spectral cube of size 256 × 256 × 601 and 256 × 256 × 26, respectively. It is clear that ADMM-net is the most efficient. The running time was divided by the number of wavelength channels, as indicated in the last row of Table 1. It can be seen that the reconstruction speed of ADMM-net is over four orders of magnitude faster than that of CVX, which enables 256 × 256 × 26 × 55 4D data cube reconstruction per second for rapid spectral imaging in real applications.

Video spectral imaging experiments were carried out to verify the capability of the proposed approach for high-speed spectral image reconstruction. The first example was a moving Thorlabs box indoors with an LED light source. The reconstructed hyperspectral video contains a total of 200 frames (19.84 s), from which 10 frames are extracted and shown in Fig. 5. From the results of recovered RGB images and spectral images at four wavelengths, it can be seen that the spatial, spectral and motion details are reconstructed with high quality. The full video is provided in the Supplement 2.

DOI: 10.1016/j.chip.2023.100045

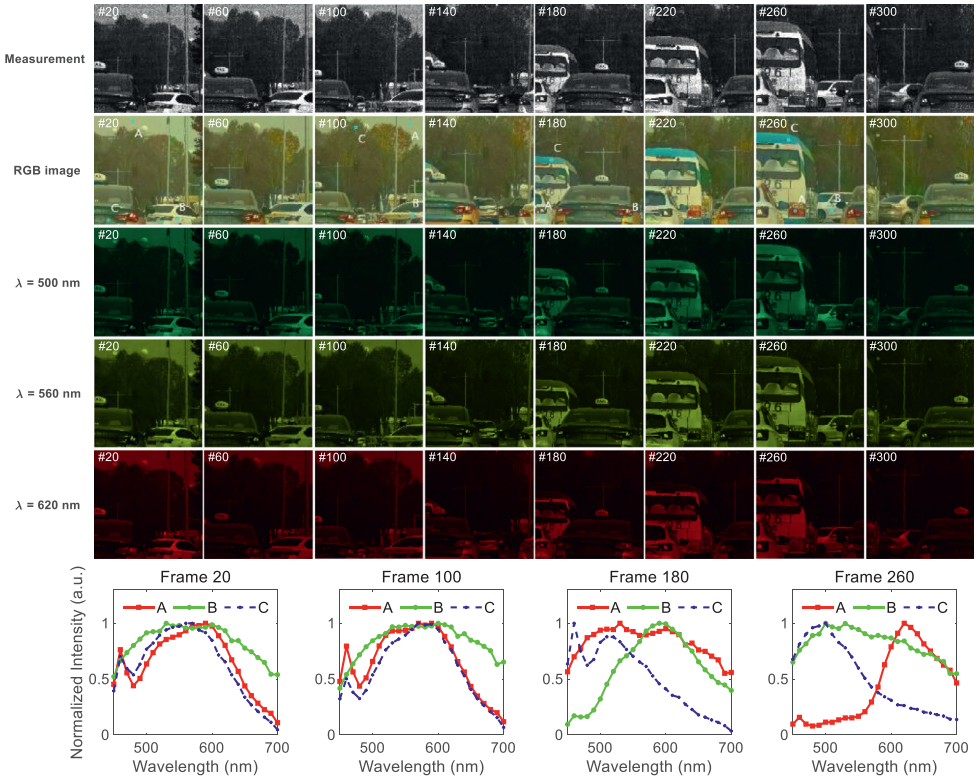

Fig. 6 | Reconstructed results of real hyperspectral data of an outdoor driving scene. Snapshot measurements captured by the on-chip spectral imager are shown on the first row. The reconstructed RGB images are shown on the second row. The reconstructed spectra of three selected points at different time frames are shown on the third row.

The second example is outdoor driving scene under the sunlight. The reconstructed video contains 300 frames (8.38 s) with a shorter exposure time, compared with the first example. That is, the approach employed in the current work can support a refresh rate of about 2 data cubes per meter for autonomous vehicles running at 60 km/h. Note that the refresh rate contains both the capture and reconstruction process, while the aforementioned 55 spectral data cube per second only takes the reconstruction process into account. Fig. 6 shows the recovered RGB images and spectral images at three wavelengths of the extracted 8 frames, as well as the recovered spectra for the selected three points. It can be seen that the driving cars with different colors can be reconstructed with fine spatial details, and the spectral characteristics are also recovered with high quality. In particular, from the spectra of points A and B in the frame 20 and 100, it can be observed that there exist obvious differences between the spectra of the sky and white car. Hence, the sky and white cars can be distinguished adopting the approach, which is significant for autonomous driving with the defect in metamerism recognition as a huge security concern[43]. The full video of this example is provided in the Supplement 3.

DISCUSSION

There are still some aspects remain to be improved in the work. Firstly, the wavelength sampling interval was set to 10 nm due to the limitation of GPU memory. Reconstruction of spectral images with more wavelength bands can be anticipated with improved computational power and GPU memory. Secondly, the capacity of the training dataset is still insufficient. Spectral image datasets with large scale, accuracy and diversity are required to improve the network performance. Thirdly, the structure of the ADMM-net can be further improved in several aspects such as increasing the number of stages, and using deeper and more advanced CNN denoisers. Besides, only Gaussian white noise was considered and employed to simulate the real situations in the network training. A more complete and accurate noise model which takes both Gaussian white noise and shot noise into account is required for better performance.

CONCLUSIONS

In summary, a deep unfolding network called ADMM-net was proposed in the manuscript and employed to fastly reconstruct the spectral image from the snapshot measurement of the metasurface-based spectral imager. The approach employed in the current work shows excellent performance in both simulated and real data reconstruction. It is worth noting that, compared with conventional point-by-point iterative spectral reconstruction, the reconstruction speed was improved by four orders of magnitudein the real experiment with high spectral accuracy and without mosaic effect. Moreover,video-rate spectral imaging was also demonstrated for moving objects and outdoor driving scenes with good performance of metamerism recognition, in which the concolorous sky and white cars can be effectively distinguished according to their spectra. The approach adopted in the current work could provide real-time capture and reconstruction of hyperspectral images, paving the way for autonomous driving and other various real-time applications in the field of intelligent perception.

METHODS

The metasurface-based spectral imager was fabricated on a Silicon-On-Insulator (SOI) wafer with a 220-nm silicon layer. Firstly, the metasurface

patterns were defined via electron beam lithography (EBL) using ZEP resist, and transferred onto the top silicon layer through reactive ion etching (RIE). Subsequently, the silicon dioxide under the patterned area was removed through immersion in buffered hydrofluoric (HF) acid in a water bath at the temperature of 40 °C, and maintained for approximately 3 min. Finally, the suspended silicon metasurface layer was peeled off and transferred onto a CIS chip via a polydimethylsiloxane (PDMS) adhesion layer.

REFERENCES

1 Shaw, G. A. & Burke, H.-H. K. Spectral imaging for remote sensing. *Linc. Lab. J.* **14**, 3–28 (2003). http://hawk.cfd.rit.edu/products/publications/Lincoln%20Lab/14_1remotesensing.pdf.

2 Williams, L. J. et al. Remote spectral detection of biodiversity effects on forest biomass. *Nat. Ecol. Evol.* **5**, 46–54 (2021). https://doi.org/10.1038/s41559-020-01329-4.

3 Lebourgeois, V. et al. Can commercial digital cameras be used as multispectral sensors? A crop monitoring test. *Sensors* **8**, 7300–7322 (2008). https://doi.org/10.3390/s8117300.

4 Lu, G. & Fei, B. Medical hyperspectral imaging: a review. *J. Biomed. Opt.* **19**, 010901 (2014). https://doi.org/10.1117/1.JBO.19.1.010901.

5 Yao, L. et al. Image enhancement based on in vivo hyperspectral gastroscopic images: a case study. *J. Biomed. Opt.* **21**, 101412 (2016). https://doi.org/10.1117/1.JBO.21.10.101412.

6 Feng, Y.-Z. & Sun, D.-W. Application of hyperspectral imaging in food safety inspection and control: a review. *Crit. Rev. Food Sci. Nutr.* **52**, 1039–1058 (2012). https://doi.org/10.1080/10408398.2011.651542.

7 Stuart, M. B., McGonigle, A. J. S. & Willmott, J. R. Hyperspectral imaging in environmental monitoring: a review of recent developments and technological advances in compact field deployable systems. *Sensors* **19**, 3071 (2019). https://doi.org/10.3390/s19143071.

8 Liang, H. Advances in multispectral and hyperspectral imaging for archaeology and art conservation. *Appl. Phys. A* **106**, 309–323 (2012). https://doi.org/10.1007/s00339-011-6689-1.

9 Gabrieli, F., Dooley, K. A., Facini, M. & Delaney, J. K. Near-UV to mid-IR reflectance imaging spectroscopy of paintings on the macroscale. *Sci. Adv.* **5**, eaaw7794 (2019). https://doi.org/10.1126/sciadv.aaw7794.

10 Bahauddin, S. M., Bradshaw, S. J. & Winebarger, A. R. The origin of reconnection-mediated transient brightenings in the solar transition region. *Nat. Astron.* **5**, 237–245 (2021). https://doi.org/10.1038/s41550-020-01263-2.

11 Green, R. O. et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.* **65**, 227–248 (1998). https://doi.org/10.1016/S0034-4257(98)00064-9.

12 Mouroulis, P., Green, R. O. & Chrien, T. G. Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information. *Appl. Opt.* **39**, 2210–2220 (2000). https://doi.org/10.1364/AO.39.002210.

13 Zhang, C., Rosenberger, M., Breitbarth, A. & Notni, G. A novel 3D multispectral vision system based on filter wheel cameras. In *Proceedings of the 2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, 267–272 (IEEE, 2016). https://doi.org/10.1109/IST.2016.7738235.

14 Gat, N. Imaging spectroscopy using tunable filters: a review. *Proc. SPIE* **4056**, 50–64 (2000). https://doi.org/10.1117/12.381686.

15 Antila, J. et al. Spectral imaging device based on a tuneable MEMS Fabry-Perot interferometer. In *Next-Generation Spectroscopic Technologies V* **8374**, 23–132 (SPIE, 2012).

16 Hagen, N. A. & Kudenov, M. W. Review of snapshot spectral imaging technologies. *Opt. Eng.* **52**, 090901 (2013). https://doi.org/10.1117/1.OE.52.9.090901.

17 Bowen, I. S. The image-slicer a device for reducing loss of light at slit of stellar spectrograph. *Astrophys. J.* **88**, 113 (1938). https://doi.org/10.1086/143964.

18 Gat, N., Scriven, G., Garman, J., Li, M. D. & Zhang, J. Development of four-dimensional imaging spectrometers (4D-IS). In *Conference on Imaging Spectrometry XI, 63020M* (2006). https://doi.org/10.1117/12.678082.

19 Bacon, R. et al. The integral field spectrograph TIGER. In *Proceedings of a ESO Conference on Very Large Telescopes and their Instrumentation*, 1185 (ESO, 1989). https://ui.adsabs.harvard.edu/abs/1988ESOC...30.1185B/abstractBacon.

20 Stoffels, J., Bluekens, A. A. J. & Jacobus, M. P. P. Color splitting prism assembly. United States patent US 084 4,084,180 (1978). https://patents.google.com/patent/US4084180A/en.

21 Harvey, A. R. & Fletcher-Holmes, D. W. High-throughput snapshot spectral imaging in two dimensions. *Proc. SPIE* **4959**, 46–54 (2003). https://doi.org/10.1117/12.485557.

22 Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006). https://doi.org/10.1109/TIT.2006.871582.

23 Candès, E. J., Romberg, J. & Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52**, 489–509 (2006). https://doi.org/10.1109/TIT.2005.862083.

24 Huang, L., Luo, R., Liu, X. & Hao, X. Spectral imaging with deep learning. *Light. Sci. Appl.* **11**, 1–19 (2022). https://doi.org/10.1038/s41377-022-00743-6.

25 Gehm, M. E., John, R., Brady, D. J., Willett, R. M. & Schulz, T. J. Single-shot compressive spectral imaging with a dual-disperser architecture. *Opt. Express* **15**, 14013–14027 (2007). https://doi.org/10.1364/OE.15.014013.

26 Wagadarikar, A., John, R., Willett, R. & Brady, D. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.* **47**, B44–B51 (2008). https://doi.org/10.1364/AO.47.000B44.

27 Correa, C. V., Arguello, H. & Arce, G. R. Snapshot colored compressive spectral imager. *J. Opt. Soc. Am. A* **32**, 1754–1763 (2015). https://doi.org/10.1364/JOSAA.32.001754.

28 Arguello, H. & Arce, G. R. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Trans. Image Process.* **23**, 1896–1908 (2014). https://doi.org/10.1109/TIP.2014.2310125.

29 Rueda, H., Arguello, H. & Arce, G. R. DMD-based implementation of patterned optical filter arrays for compressive spectral imaging. *J. Opt. Soc. Am. A* **31**, 80–89 (2015). https://doi.org/10.1364/JOSAA.32.000080.

30 Rueda, H., Arguello, H. & Arce, G. R. Compressive spectral testbed imaging system based on thin-film color-patterned filter arrays. *Appl. Opt.* **55**, 9584–9593 (2016). https://doi.org/10.1364/AO.55.009584.

31 Lin, X., Liu, Y., Wu, J. & Dai, Q. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. Graph.* **33**, 1–11 (2014). https://doi.org/10.1145/2661229.2661262.

32 Arce, G. R., Brady, D. J., Carin, L., Arguello, H. & Kittle, D. S. Compressive coded aperture spectral imaging: an introduction. *IEEE Signal Process. Mag.* **31**, 105–115 (2013). https://doi.org/10.1109/MSP.2013.2278763.

33 Sahoo, S. K., Tang, D. & Dang, C. Single-shot multispectral imaging with a monochromatic camera. *Optica* **4**, 1209–1213 (2017). https://doi.org/10.1364/OPTICA.4.001209.

34 French, R., Gigan, S. & Muskens, O. L. Speckle-based hyperspectral imaging combining multiple scattering and compressive sensing in nanowire mats. *Opt. Lett.* **42**, 1820–1823 (2017). https://doi.org/10.5258/SOTON/D0006.

35 Oktem, F. S., Kar, O. F., Bezek, C. D. & Kamalabadi, F. High-resolution multispectral imaging with diffractive lenses and learned reconstruction. *IEEE Trans. Comput. Imaging* **7**, 489–504 (2021). https://doi.org/10.1109/TCI.2021.3075349.

36 Monakhova, K., Yanny, K., Aggarwal, N. & Waller L. Spectral Diffuser Cam: lensless snapshot hyperspectral imaging with a spectral filter array. *Optica* **7**, 1298–1307 (2020). https://doi.org/10.1364/OPTICA.397214.

37 Ehira, K., Horisaki, R., Nishizaki, Y., Naruse, M. & Tanida, J. Spectral speckle-correlation imaging. *Appl. Opt.* **60**, 2388–2392 (2021). https://doi.org/10.1364/AO.418361.

38 Hu, H. et al. Practical snapshot hyperspectral imaging with DOE. *Opt. Lasers Eng.* **156**, 107098 (2022). https://doi.org/10.1016/j.optlaseng.2022.107098.

39 Xiong, J. et al. Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces. *Optica* **9**, 461–468 (2022). https://doi.org/10.1364/OPTICA.440013.

40 Yang, J. et al. Ultraspectral imaging based on metasurfaces with freeform shaped meta-atoms. *Laser Photonics Rev.* **16**, 2100663 (2022). https://doi.org/10.1002/lpor.202100663.

41 Yasuma, F., Mitsunaga, T., Iso, D. & Nayar, S. K. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Trans. Image Process.* **19**, 2241–2253 (2010). https://doi.org/10.1109/TIP.2010.2046811.

42 Choi, I., Jeon, D. S., Nam, G., Gutierrez, D. & Kim, M. H. High-quality hyperspectral reconstruction using a spectral prior. *ACM Trans. Graph.* **36**, 218 (2017). https://doi.org/10.1145/3130800.3130810.

43 Liang, J., Zhou, J., Tong, L., Bai, X. & Wang, B. Material based salient object detection from hyperspectral images. *Pattern Recognit.* **76**, 476–490 (2018). https://doi.org/10.1016/j.patcog.2017.11.024.

44 Yuan, X., Brady, D. J. & Katsaggelos, A. K. Snapshot compressive imaging: theory, algorithms, and applications. *IEEE Signal Process. Mag.* **38**, 65–88 (2021). https://doi.org/10.1109/MSP.2020.3023869.

45 Yuan, X. Generalized alternating projection based total variation minimization for compressive sensing. In *Proceedings of the 2016 IEEE International Conference on image processing (ICIP)*, 2539-2543 (2016). https://doi.org/10.1109/ICIP.2016.7532817.

46 Smith, T. & Guild, J. The C.I.E. colorimetric standards and their use. *Trans. Opt. Soc.* **33**, 3 (1931). https://doi.org/10.1088/1475-4878/33/3/301.

47 Liu, Y., Yuan, X., Suo, J., Brady, D. J. & Dai, Q. Rank minimization for snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 2990–3006 (2018). https://doi.org/10.1109/TPAMI.2018.2873587.

48 Grant, M. & Boyd, S. CVX: Matlab software for disciplined convex programming. *CVX* (January, 2020). http://cvxr.com/cvx.

MISCELLANEA

Authors' contributions J. Yang conceived the study and composed the manuscript. K. Cui and Y. Huang supervised the project, provided much support on experimentation, and reviewed the manuscript with contributions from all other co-authors. All authors read and approved the manuscript.

Availability of data and materials The data and materials that support the findings of this study and custom codes are available from the corresponding author upon reasonable request.

Supplementary materials Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chip.2023.100045.

Funding The National Natural Science Foundation of China (Grant No. U22A6004); The National Key Research and Development Program of China (2022YFF1501600).

Acknowledgments The authors would like to express their sincere gratitude to Beijing Seetrum Technology Co., Ltd. for their valuable discussions, and Tianjin H–Chip Technology Group Corporation, Innovation Center of Advanced Optoelectronic Chip and Institute for Electronics and Information Technology in Tianjin, Tsinghua University for their fabrication support with SEM and ICP etching.

Declaration of Competing Interests The authors declare no conflicts of interest.