Fast Deep Learning Reconstruction Algorithm for On-chip Snapshot Hyperspectral Imaging

1st Shijie Rao
Department of Electronic Engineering
Tsinghua University
Beijing, China
raosj20@mails.tsinghua.edu.cn

2nd Kaiyu Cui
Department of Electronic Engineering
Tsinghua University
Beijing, China
kaiyucui@tsinghua.edu.cn

3rd Yidong Huang
Department of Electronic Engineering
Tsinghua University
Beijing, China
yidonghuang@tsinghua.edu.cn

4th Jiawei Yang
Department of Electronic Engineering
Tsinghua University
Beijing, China
2782291135@qq.com

5th Sheng Xu
Department of Electronic Engineering
Tsinghua University
Beijing, China
xu-s20@mails.tsinghua.edu.cn

6th Yali Li
Department of Electronic Engineering
Tsinghua University
Beijing, China
liyali13@mail.tsinghua.edu.cn

***Abstract*—We proposed a reconstruction algorithm powered by deep learning technics for snapshot hyperspectral imaging based on compressive sensing. The fast computational speed and high reconstruction fidelity make it practical in real-time on-chip hyperspectral imaging systems.**

***Index Terms*—Hyperspectral reconstruction, compressive sensing, deep learning, hyperspectral imaging**

I. Introduction

As an important tool for the analysis of material composition, spectroscopy is widely used in many fields, such as astronomy [1], earth science [2] and medical diagnosis [3]. However, the traditional hyperspectral imaging devices are usually bulky and expensive, and it usually takes tens of seconds to scan a hyperspectral image (HSI), which makes it inconvenient to perform real-time hyperspectral sensing. Recently, on-chip snapshot spectral imaging devices have been proposed based on compressive sensing (CS) theory [4]–[6]. The performance of these devices is highly related to the CS reconstruction method. Usually, the reconstruction is realized by solving the sparse optimization algorithms such as absolute shrinkage and selection operator (LASSO) [7] and gradient projection (GP) [8]. These optimization methods are generally supported by reliable mathematical derivations and converges to predictable and good results by iterative computation, which is rather time consuming. Another problem is that these optimization algorithms based on sparsity prior are usually sensitive to noise. The reconstruction fidelity drops dramatically when the signal-to-noise ratio (SNR) of the compressively sensed measurements is low. These limitations slow down the practical usage of the on-chip snapshot hyperspectral imaging systems for real-time applications.

In our work, we proposed an algorithm based on deep learning to solve the CS problem for hyperspectral reconstruction. Deep neural networks (DNNs) like Transformer [9] have shown great performance in machine translation tasks. From our perspective, the compressively sensed measuring process projects the spectrum from natural domain to a certain compressed domain, and what the reconstruction process does is actually translating the spectrum from compressed domain back to natural domain. Therefore, we designed a DNN algorithm based on Transformer to perform reconstruction. Our algorithm is non-iterative, fast and very effective. Furthermore, after anti-noise training, our algorithm can keep a relatively high fidelity when the SNR of measurements is low, showing great potential in real-world on-chip snapshot hyperspectral sensing applications.

II. Algorithm and Data Generation

The procedure of the on-chip snapshot hyperspectral imaging system is shown in Fig. 1. The hyperspectral imaging sensor can be regarded as a grid of spectrometers. Therefore, we take one snapshot spectrometer as an example to explain the workflow of compressive spectrum sensing. The on-chip miniature spectrometer usually utilizes metasurfaces to modulate the incident light in broadband. Then the energy of the modulated light is measured by the CMOS image sensor (CIS) that attached under the metasurfaces. Assume that the spectrum of the incident light is $\boldsymbol{x}$, and the transmission response matrix of the metasurface array is $\boldsymbol{\Phi}$. Then, the electrical signal outputted by CIS can be mathematically modelled as:

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{x} \in R^{N\times 1}, \boldsymbol{y} \in R^{M\times 1}, \boldsymbol{n} \in R^{M\times 1}, \boldsymbol{\Phi} \in R^{M\times N}$, and $\boldsymbol{n}$ is the measurement noise. The aim of the reconstruction algorithm is to find $\boldsymbol{x}$ through the known parameters $\boldsymbol{y}$ and $\boldsymbol{\Phi}$. However, as $M < N$, the equation is under-determined. Conventional optimization algorithms tend to solve the problem by finding the optimal solution under the sparsity constraint of $\boldsymbol{x}$. However, we solve it by DNN. The DNN can be regarded as a black box $F(\cdot)$. It takes $\boldsymbol{y}$ and $\boldsymbol{\Phi}$ as inputs and outputs

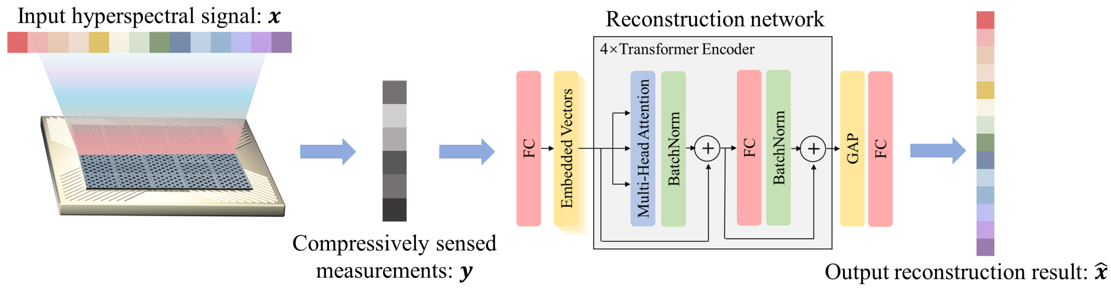
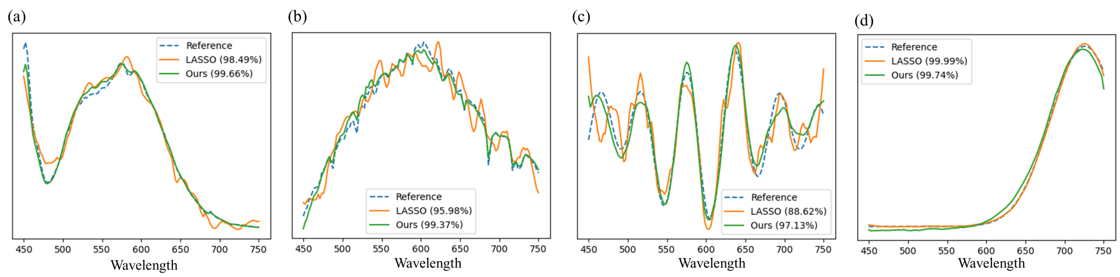
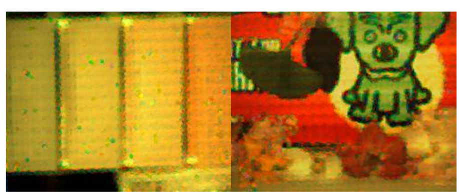


Fig. 1. The procedure of the on-chip snapshot hyperspectral imaging system



(e)



(f)

	MSE ($\times 10^{-3}$)	Similarity (%)	Running Time (ms)
Ours	0.74	96.88	5(CPU)/0.1(GPU)
LASSO	3.52	92.11	120

Fig. 2. Reconstruction results on testing set. (a) Reconstruction of the white LED spectrum. (b) Reconstruction of sunlight spectrum. (c) Reconstruction of the spectrum generated by Fourier series. (d) Reconstruction of the spectrum generated by Gaussian distribution functions. (e) The reconstructed HSIs of two real-world scenes. (f) Quantitative comparison between our network and the LASSO reconstruction algorithm on the two metrics mean squared error (MSE) and cosine similarity.

the reconstruction result $\hat{\boldsymbol{x}} = F(\boldsymbol{\phi}, \boldsymbol{y})$. In our algorithm, $\boldsymbol{\Phi}$ is firstly divided by $\boldsymbol{y}$ in each row to simplify the inputs.

In our network, as is shown in Fig. 1, each row of $\boldsymbol{\Phi}'$ is firstly embedded by the same fully connected (FC) layers. Finally, the output of the Transformer block is averaged by columns and mapped by other FC layers to get the reconstruction result $\hat{\boldsymbol{x}}$. During training, we optimized our network to minimize the mean squared error and maximize the cosine similarity between $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}$. Adadelta [10] optimizer was adopted to train the network.

To train the network, we need to generate a dataset containing all of the $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{n}$ and $\boldsymbol{\Phi}$. As metasurface is the mostly used modulator in on-chip spectrometers, we generate $\boldsymbol{\Phi}$ by a set of metasurfaces. However, our algorithm is not limited to metasurface-based spectrometers. We simulate the transmission responses of more than 100 thousand different metasurfaces, and for each input, transmission responses of M metasurfaces is randomly selected and form the matrix $\boldsymbol{\Phi}$. To generate $\boldsymbol{x}$, we deployed a commercial hyperspectral camera to capture 20 scenes in the real world and combined our own HSIs with the public ICVL HSI dataset [11]. The HSIs and the transmission responses are in 450 750nm bands. Finally, $\boldsymbol{y}$ is calculated by 1 and $\boldsymbol{n}$ is simulated by Gaussian noise with a SNR of 25dB.

III. Results and Discussion

In our experiments, we adopted $M = 25$ and $N = 128$. SNR is set to 25dB. After training, the reconstruction results of our network on the testing set are displayed in Fig. 2. The spectrum in Fig. 2(a) and Fig. 2(b) are the real-world measurement of white LED light and sunlight. The spectrum in Fig. 2(c) and Fig. 2(d) are synthetic data generated by Fourier series and Gaussian distribution functions. LASSO represents the reconstruction result performed by convex optimization algorithm absolute shrinkage and selection operator, which is widely used in on-chip spectrometers. Results indicates that our DNN algorithm performs well on the reconstruction task. Compared with LASSO algorithm, the reconstruction results of our network show less distortion and the reconstruction speed is much faster. We also reconstruct some HSIs of the real-world scenes, which is shown in Fig. 2(e). And Fig.

2(f) shows the quantitative comparison between our network and LASSO. Our network can outperform the conventional iterative optimization algorithm. When running parallel on GPU, our network needs 1000 times shorter running time than LASSO algorithm. It provides a much more efficient reconstruction strategy for real-world on-chip hyperspectral sensing applications.

IV. Conclusions

Our DNN-based hyperspectral reconstruction algorithm has the advantages of high reconstruction fidelity and low noise sensitivity. It is well suited for on-chip spectrometers and hyperspectral imaging devices based on compressive sensing. Experimental results prove that our algorithm can speed up the reconstruction process by 1000 times while keeping high reconstruction precision.

Acknowledgment

This work is supported by China State key development program in 14th Five-Year (2021YFF0602103); National Natural Science Foundation of China (Grant No. U22A6004); The National Key Research and Development Program of China (2022YFF1501600).

References

[1] K. Sharma, A. Kembhavi, A. Kembhavi, T. Sivarani, S. Abraham and K. Vaghmare, "Application of convolutional neural networks for stellar spectral classification," *Monthly Notices of the Royal Astronomical Society*, vol. 491, (2), pp. 2280-2300, 2020.

[2] D. Ramakrishnan and R. Bharti, "Hyperspectral remote sensing and geological applications," *Curr. Sci.*, pp. 879-891, 2015.

[3] N. Kumar, et al, "Hyperspectral tissue image segmentation using semi-supervised NMF and hierarchical clustering," *IEEE Trans. Med. Imaging*, vol. 38, (5), pp. 1304-1313, 2018.

[4] J. Xiong, et al, "Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces," *Optica*, vol. 9, (5), pp. 461-468, 2022.

[5] J. Yang, et al, "Ultraspectral Imaging Based on Metasurfaces with Freeform Shaped Meta-Atoms," *Laser & Photonics Reviews*, pp. 2100663, 2022.

[6] S. Rao, Y. Huang, K. Cui and Y. Li, "Anti-spoofing face recognition using a metasurface-based snapshot hyperspectral image sensor," *Optica*, vol. 9, (11), pp. 1253-1259, 2022.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, (1), pp. 267-288, 1996.

[8] M. A. Figueiredo, R. D. Nowak and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, (4), pp. 586-597, 2007.

[9] A. Vaswani, et al, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[10] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv Preprint arXiv:1212.5701*, 2012.

[11] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *European Conference on Computer Vision*, 2016, pp. 19-34.