Efficient Snapshot Spectral Imaging: Calibration-Free Parallel Structure with Aperture Diffraction Fusion

Tao Lv[1,3], Lihao Hu[1,3], Shiqiao Li[1,3], Chenglong Huang[1,3], and Xun Cao[1,2,⋆]

[1] School of Electronic Sciene and Engineering, Nanjing University, Nanjing, China
[2] Key Laboratory of Optoelectronic Devices and Systems with Extreme Performances of MOE, Nanjing University, Nanjing, China
[3] NJU-ZPTECH University-Enterprise Joint Laboratory for Computational Spectral Imaging
{lvtao, lihaohu, lishiqiao, Chenglong-Huang}@smail.nju.edu.cn, caoxun@nju.edu.cn

Abstract. Aiming to address the repetitive need for complex calibration in existing snapshot spectral imaging methods, while better trading off system complexity and spectral reconstruction accuracy, we demonstrate a novel Parallel Coded Calibration-free Aperture Diffraction Imaging Spectrometer (PCCADIS) with simplest parallel architecture and enhanced light throughput. The system integrates monochromatic acquisition of diffraction-induced blurred spatial-spectral projections with uncalibrated chromatic filtering for guidance, enabling portable and lightweight implementation. In the inverse process of PCCADIS, aperture diffraction produces blurred patterns through global replication and band-by-band integration, posing challenges in registration with clearly structured RGB information by feature matching. Therefore, the Self-aware Spectral Fusion Cascaded Transformer (SSFCT) is proposed to realize the fusion and reconstruction of uncalibrated inputs, which demonstrates the potential to substantially improve the accuracy of snapshot spectral imaging while concurrently reducing the associated costs. Our methodology is rigorously evaluated through comprehensive simulation experiments and real reconstruction experiments with a prototype, confirming the high accuracy and user-friendliness of PCCADIS.

Keywords: Calibration-Free · Aperture Diffraction Fusion · Snapshot Spectral Imaging

1 Introduction

Spectrum represents the intensity distribution across various light frequencies, encapsulating the fundamental characteristics of materials. Human vision perceives the world through the differential responses of various cones to distinct spectral bands [39], with RGB cameras imitating this trichromatic vision for color imaging [28]. However, more refined spectral details and broader spectral ranges exceed both human visual capabilities and those of existing RGB-based machine vision systems.

Snapshot Spectral Imaging (SSI) aims to obtain the 2D distribution of spectral information in a single exposure [11, 15], facilitating an enhanced multidimensional understanding, helping to elucidate underlying mechanisms in domains such as combustion

⋆ Corresponding author: caoxun@nju.edu.cn

dynamics [23], galactic astronomy [1], and tissue and cellular dynamics [40]. However, existing SSI systems are constrained by their **substantial footprint** [12, 13, 16, 18, 42], **compromised spectral imaging fidelity** [3, 24, 35] and **extremely complex optical calibration process** [3, 12, 13, 16, 18, 24, 35, 42], posing challenges for their integration as spectral solutions in machine vision applications [49]. It's important to recognize that despite the inherent limitations of the first two points, SSI methods still have wide application. However, one of the major challenges lies in its susceptibility to external disturbances like physical vibrations or temperature variations. Even slight perturbations necessitate recalibration, demanding expensive laboratory equipment such as super-continuum lasers and involving complex calibration procedures.

Meanwhile, improving the quality of SSI reconstructions fundamentally hinges on optimizing the coding method, with parallel architecture showing great promise in this endeavor. For instance, integrating RGB cameras, standard for capturing scene color information on mobile devices, enhances color clarity, edge definition, and textural detail, thereby boosting reconstruction accuracy. However, adopting a parallel structure escalates system complexity and calibration challenges. In practical applications, calibration limitations manifest in various aspects, as illustrated in Fig. 1(c).

• **Calibration in focal plane**: It is necessary to calibrate the absolute position of spatial coding devices or spatial confinement devices, such as mask or field stops. These are typically situated on the focal plane between relay systems. They include random masks for CASSI-based methods [17, 18, 42, 44, 45], uniform masks for PMVIS-based methods [12, 13], and field stops used in the CTIS-based methods [16, 50]. Even slight movements or out-of-focusing of these devices will significantly alter the optical encoding process, leading to disastrous reconstruction results.

• **Calibration in aperture plane**: Devices like gratings, prisms, and diffractive optical elements (DOEs) play a key role in spectral encoding for SSI methods. Precise calibration to their dispersion laws is crucial for accurate forward processing. In DOE-based methods, despite carefully designed phase encoding, achieving perfect fabrication is challenging. As a result, each DOE-based system [2, 19, 24, 48] must be paired with complex pre-calibrations and a unique, well-trained reconstruction network.

• **Calibration between branches**: Using parallel encoding to obtain multi-paradigm 2D measurements of a scene is an effective approach to enhance SSI reconstruction accuracy. Typically, this involves integrating RGB or monochrome camera as a prior branch [13, 17, 44, 50]. However, it introduces a significant challenge: SSI utilizes dispersers to create a dispersive version of HISs, which is substantially different from the well-constructed, dispersion-free measurements from the prior branch. This modality differences make direct registration difficult. Hence, **pixel-accurate calibration** of the correspondence between the two optical branch images is necessary, or alternatively, **training a cross-modal images registration model**. **Unfortunately, recalibration or retraining becomes necessary following system perturbations** (Supp. Mat. Sec. 5).

So, can we develop a calibration-free SSI framework and achieves an efficient trade-off between reconstruction accuracy and system complexity?

Here, we start with the Aperture Diffraction Imaging Spectrometer (ADIS) [31], which features a compact structure and reduced reliance on calibration. However, it employs joint serial coding of spatial-spectral projection generate by orthogonal masks and

mosaic filtering, leading to a great challenge to solve the inverse problem and compromised imaging quality. Therefore, here we innovatively adapt the serial coding of ADIS to parallel coding, as shown in Fig. 1(a), and rely on the physical parameters of the system to calculate the diffraction pattern without calibration by deeply exploring the properties of aperture diffraction. And for the cross-modal image registration/calibration challenges brought by parallel optical paths, as shown in Fig. 1(b), we innovatively propose the spectral self-aware fusion cascade transformer (SSFCT), which aims to directly realize high-precision HSI recovery from data collected from uncalibrated camera branches. Overall, we introduced a calibration-free system called the **P**arallel **C**oded **C**alibration-free **A**perture **D**iffraction **I**maging **S**pectrometer (PCCADIS). Leveraging parallel optical coding and algorithm optimization, PCCADIS unlocking the potential of aperture diffraction as a spectral coding scheme for deployment in mobile terminals and improving the light throughput compared to serial structures.

In summary, the main contributions of this paper include:

• A novel calibration-free parallel-structured SSI framework for fusing RGB and monochromatic aperture diffraction acquisition is proposed, which is the most compact in parallel structure with enhanced light throughput.

• A novel algorithm is designed for the efficacious fusion of uncalibrated acquisitions, which autonomously recovers high-precision HSI from different encoded measurements. The feasibility of the algorithm was also verified on a DOE-based approach.

• A prototype is developed to verify the PCCADIS's excellent performance in terms of fusion accuracy of uncalibrated data and the user-friendliness of the system.

2 Related Works

Snapshot Spectral Imaging. In comparison to traditional scanning imaging spectrometer, which trades scanning time for spectral and spatial resolution [5], snapshot spectral imaging offers a notable enhancement in capturing the dynamics of scenes. Over recent decades, there has been a notable emergence of SSI techniques [12, 13, 16, 18, 42], ushering in a transition from static to dynamic spectral imaging technology. However, these imaging systems typically incorporate a variety of optical encoding elements, such as prisms, coded apertures, and relay lenses. This complexity contributes to the bulky volumes and intricate structures of these imaging systems [49].

Numerous methodologies prioritize system compactness. This includes the generation of a band-specific rotating Point Spread Functions (PSFs) for snapshot spectroscopic imaging through the utilization of skillfully designed Diffractive Optical Elements (DOEs) [2,19,24]. Moreover, the exploitation of encoded modulation capabilities within the scattering medium [26, 35] and metasurfaces [20] offers additional avenues for enhancing compactness. Nonetheless, these techniques do introduce a trade-off, as they impact the precision of the inverse reconstruction process. Therefore, this paper endeavors to establish a paradigm that simultaneously achieves high-precision SSI while maintaining a portable and user-friendly architecture. It accomplishes SSI by switching from the serial encoding of ADIS to a parallel camera architecture, resulting in the most compact parallel structure and a substantial enhancement in reconstruction accuracy.

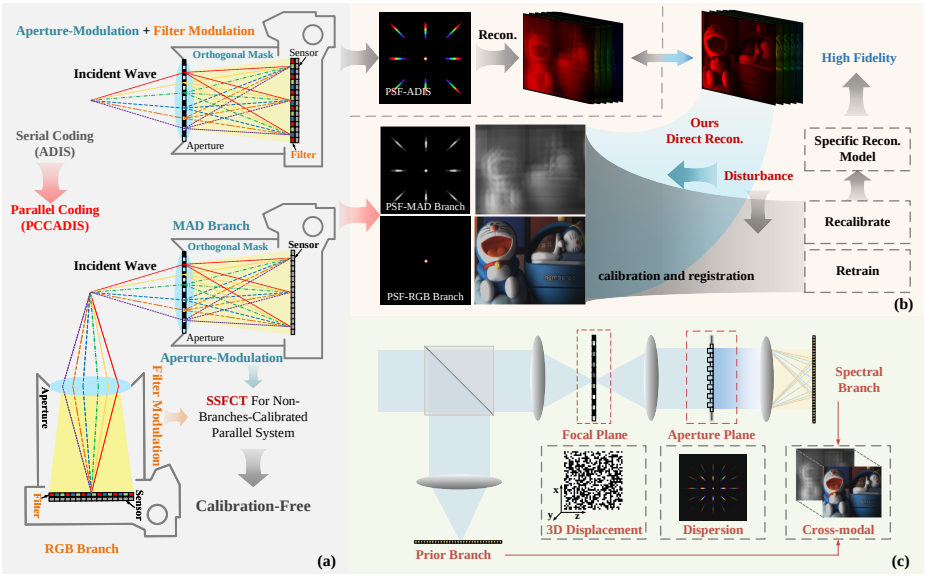


Fig. 1: (a) PCCADIS parallelizes the serial aperture modulation and filter modulation in ADIS [31], achieving system calibration-free characteristics and oprecision enhancement through algorithm design; (b) demonstrates PCCADIS's ability to directly reconstruct high-fidelity spectra from uncalibrated input and robustness to external interference; (c) shows three calibration requirements (focal plane, aperture plane, different branches) for various current SSI methods.

Parallel cameras. Parallel cameras offer a viable solution to the inherent trade-offs in imaging system design. By employing front-end parallel image acquisition and back-end parallel image processing, these systems effectively fuse multiple parallel images, enhancing the overall information throughput [7]. Beyond improving performance in individual dimensions such as space [6,32,44,48,50], time [45,46], and depth of field [4], parallel cameras can also capture higher-dimensional plenoptic light field information, including hyperspectral light field imaging [47].

Despite these advantages, a significant challenge arises from the neglect of registration issues in existing methods. Typically, image registration is achieved through intricate optical calibration processes, compromising system stability against external disturbances. This paper addresses this crucial concern by focusing on achieving high-precision reconstruction of uncalibrated data. The proposed approach involves optimizing the fusion framework and reconstruction algorithm, aiming to enhance accuracy and mitigate cross-model image registration challenges.

Fusion and Reconstruction Algorithm. The core challenges in addressing the inverse process of SSI rooted in parallel architecture lie in fusion and reconstruction. With the advancement of deep learning, it becomes possible to enhance imaging quality from cross-modal or cross-view images [41, 53] utilizing parallel architecture. Additionally, source image registration errors can be addressed by introducing a preprocessing registration module before fusion, as shown in prior studies [22, 43]. Liang et al. [29] proposed a self-supervised learning framework for image fusion with unpaired images,

while Jiang et al. [25] pioneered a comprehensive scene-aware fusion method, incorporating multi-view and multi-modal data integration. Meanwhile, Transformer-based methods [9, 52] have surfaced as superior alternatives, surpassing previous approaches and significantly enhancing reconstruction efficiency. Furthermore, some investigations have amalgamated the merits of iterative and deep-learning methodologies by employing deep unfolding networks for HSI reconstruction [10, 31].

3 System Overview

In this section, we introduce the proposed parallel SSI architecture, PCCADIS, covering its basic configuration, aperture diffraction analysis of two complementary Orthogonal mask, branch calibration analysis for different parallel architectures, and overall system complexity and calibration complexity analysis.

3.1 System Configuration

The system comprises two imaging branches designed for monochrome acquisition of spatial-spectral projections generated by an orthogonal mask (OM) and color acquisition of RGB-Bayer filtering. The former consists of a monochrome camera and a specialized diffractive lens featuring an OM for creating aperture diffraction. The latter is a standard commercial RGB camera. To address the challenge of fusing uncalibrated data, we establish a guided architecture with a Monochromatic Aperture Diffraction (MAD) as the core, complemented by any RGB prior with known filtering parameters to capture a similar scene. Furthermore, to facilitate SSI for mobile applications.

The mask placed on the principal plane of the lens can be understood as a stack of N parallel lines orthogonal to each other and can be described in the following form:

$$M_m = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} rect[\frac{x-(nd-\tau)}{a}] rect[\frac{y-(md-\tau)}{b}] \tag{1}$$

where d denotes the interval between neighboring parallel lines; $(d-a)$, $(d-b)$ denote the linewidths of the parallel lines in the orthogonal direction, respectively; and $\tau = \frac{N-1}{2}d$ denotes a fixed offset. Let D be the size of the mask, the complementary form of this mask can be expressed as $M_c = rect(\frac{x}{D})rect(\frac{y}{D}) - M_m$.

Since the masks have a uniform grayscale distribution and are placed on the aperture plane, there's no need to correct mask positions and calibrate their spatial projections, as required in CASSI-based and PMVIS-based methods. (Supp. Mat. Sec. 1)

3.2 PSFs Analysis

The intensity and position relationship of the diffraction pattern can be represented as:

$$I_m(x,y) = \frac{I_0}{\lambda^2 {f_2}^2} a^2 b^2 E_m^2 \tag{2}$$

$$E_m = \frac{\sin\beta_1}{\beta_1} \frac{\sin(N\gamma_1)}{\sin\gamma_1} \frac{\sin\beta_2}{\beta_2} \frac{\sin(N\gamma_2)}{\sin\gamma_2} \tag{3}$$

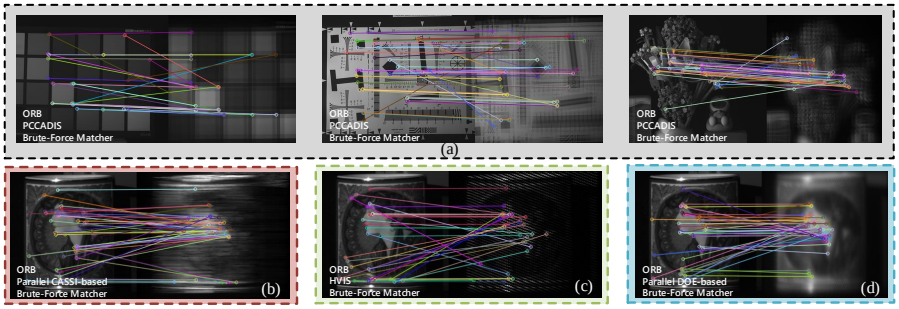


Fig. 2: (a) Presents the results of image registration using ORB for three real scenes captured by PCCADIS. (b), (c), and (d) show the results of image registration using ORB for simulated measurements based on parallel optical paths using CASSI, PMVIS, and DOE, respectively. The registration performance in all cases is poor, indicating the need for precise calibration.

where f_2 denotes the focal length of lens, θ_1 and θ_2 are the diffraction angles in x- and y-directions respectively, $\beta_1 = \frac{1}{2}kb\sin\theta_1$, $\beta_2 = \frac{1}{2}ka\sin\theta_2$, $\gamma_1 = \frac{1}{2}kd\sin\theta_1$, $\gamma_2 = \frac{1}{2}kd\sin\theta_2$. Further, by utilizing the paraxial approximation in far-field imaging, the angular relationship can be transformed into a position relationship ($\sin\theta_1 \approx \tan\theta_1 = \frac{x}{f_2}$, $\sin\theta_2 \approx \tan\theta_2 = \frac{y}{f_2}$). On this basis, the diffraction pattern defined by M_c is:

$$I_c(x,y) = \frac{I_0}{\lambda^2 {f_2}^2}\left[D^2\,\mathrm{sinc}(\alpha_1)\,\mathrm{sinc}(\alpha_2) - abE_m\right]^2 \tag{4}$$

where $\alpha_1 = \frac{1}{2}kD\sin\theta_1$, $\alpha_2 = \frac{1}{2}kD\sin\theta_2$. According to the expression of $I_m(x,y)$ and $I_c(x,y)$, if the distribution of parallel lines in the orthogonal direction is uniform ($a = b$), the intensity value at the diffraction order of (o_x, o_y) can be expressed as:

$$I\left(o_x, o_y\right) = \frac{I_0}{\lambda^2 {f_2}^2} a^2 b^2 N^4 A\left(o_x, o_y\right) \tag{5}$$

where $A\left(o_x, o_y\right)$ denotes the intensity coefficient of diffraction at all levels of diffraction in different bands:

$$A\left(o_x, o_y\right) = \begin{cases} \left(\varepsilon^2 - 1\right)^2 & (0,0), M_c \\ 1 & (0,0), M_m \\ \left(\frac{\varepsilon \sin(\frac{\pi}{\varepsilon} o_x)}{\pi o_x}\right)^2 & o_x \neq 0, o_y = 0 \\ \left(\frac{\varepsilon \sin(\frac{\pi}{\varepsilon} o_y)}{\pi o_y}\right)^2 & o_x = 0, o_y \neq 0 \\ \left(\frac{\varepsilon^2 \sin(\frac{\pi}{\varepsilon} o_x)\sin(\frac{\pi}{\varepsilon} o_y)}{\pi^2 o_x o_y}\right)^2 & o_x \neq 0 \& o_y \neq 0 \end{cases} \tag{6}$$

where $\varepsilon = \frac{a}{d} = \frac{b}{d}$ denotes the duty ratio of the mask in the orthogonal direction. Let s_{pixel} be the sensor pixel size, then the expression for the diffraction order position is:

$$\left(x_{o_x,o_y}, y_{o_x,o_y}\right) = \left(o_x \frac{\lambda f_2}{d \cdot s_{pixel}}, o_y \frac{\lambda f_2}{d \cdot s_{pixel}}\right) \tag{7}$$

Table 1: Complexity analysis of parallel computational spectral imaging systems. OPS: Optical Path Structure; $Cali_{spa}$: Spatial Calibration of different branches; $Cali_{Disper}$: Dispersion Calibration; $Cali_{PSFs/M}$: PSFs/Mask Calibration; N_{BP}: Necessity of Beam Splitter.

Imaging Method	OPS	$Cali_{spa}$	$Cali_{Disper}$	$Cali_{PSFs/M}$	N_{BP}
PCCADIS+SSFCT	2f+2f	✗	✗	✗	✗
prior branch+PMVIS [13]	2f+4f	✓	✓	✓	✓
prior branch+CASSI [17, 44]	2f+4f	✓	✓	✓	✓
prior branch+CTIS [50]	2f+4f	✓	✓	✓	✓
prior branch+DOE-coded [48]	2f+2f	✓	✗	✓	✓

Utilizing Equation 6 and Equation 7, it is possible to calculate the PSFs of the spatial-spectral projection for the current hardware parameters by calculating the position and amplitude of each diffraction order.

3.3 Optical Branches Registration and System Complexity Analysis

In parallel systems, ensuring accurate spatial registration and calibration between the scenes captured by two optical branches is crucial for maintaining spectral imaging quality. To visually assess the challenges posed by different coding schemes used in parallel SSI systems for image registration and optical path calibration, we employ the classical image registration algorithm: Oriented FAST and Rotated BRIEF (ORB) [38] to match feature points in images captured by the different branches. However, all methods yield unsatisfactory feature point matching results as shown in Fig. 2.

4 Fusion and Reconstruction

The key contribution highlighted is development of the **calibration-free** system with most compact **parallel structure**, which relies on PCCADIS and SSFCT for reconstructing uncalibrated inputs. And It's simplicity is evident from its low hardware complexity, calibration needs, and the necessity for consistent input, as detailed in Table 1. Considering the optical branches of PCCADIS are uncalibrated, the forward process is:

$$\begin{aligned} Y_M(x_1, y_1) &= \sum_{\lambda \in S_\lambda} \{X_\lambda(x_1, y_1) \otimes H_\lambda(x, y)\} + N_1 \\ Y_P(x_2, y_2) &= \sum_{\lambda \in S_\lambda} \{X_\lambda(x_2, y_2) \odot \xi_\lambda(x, y)\} + N_2 \end{aligned} \tag{8}$$

where $\lambda \in S_\lambda$ denotes the discrete domain of the spectral dimension, N_1, N_2 denote the imaging noise of the two optical branches, respectively, $H_\lambda(x, y)$ denotes the PSFs defined by the orthogonal mask diffraction, and $\xi_\lambda(x, y)$ denotes the filter function. $X_\lambda(x_1, y_1)$, $X_\lambda(x_2, y_2)$ denote the spectral truth of the information of different scenes, and Y_M and Y_P are diffraction projection coding and filtering coding for the corresponding spectral information, respectively. The uncalibrated inputs increase the ill-posedness of the reconstruction process.

To mitigate the constraint, we have integrated the multi-modality image fusion [53] and spectral reconstruction algorithms to develop a novel pipeline, termed SSFCT,

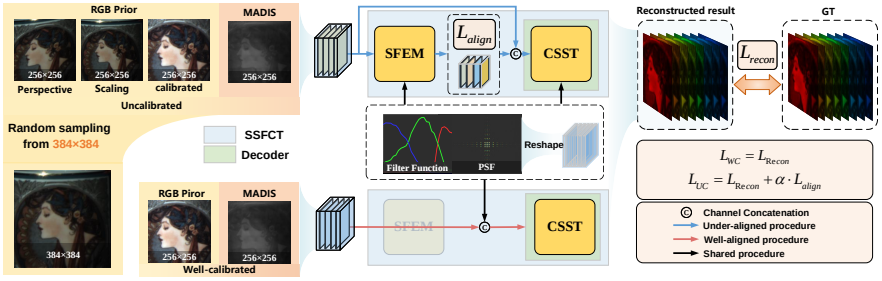


Fig. 3: Illustration of the fusion and reconstruction pipeline of PCCADIS. in the uncalibrated (UC) mode, SSFCT extracts and fuses the features through SFEM and guides them through L_{align}, and in the well-calibrated (WC) mode, SSFCT will degenerate to the decoder CSST.

for reconstructing uncalibrated input. As shown in Fig. 3, SSFCT advances the CSST framework by incorporating a self-aware feature encoding module (SFEM) to the Cascaded Shift-Shuffle Spectral Transformer (CSST) [31]. The algorithm operates with two distinct input modes: well-calibrated (WC) and uncalibrated (UC). In the WC mode, SSFCT simplifies into the CSST decoder, where the activation of the SFEM yields negligible benefits. Conversely, for UC inputs, the SFEM's output and input are concatenated along the channel dimension before being processed by CSST, significantly enhancing the reconstruction quality of uncalibrated inputs.

4.1 Self-aware Feature Encoding Module (SFEM)

As depicted in Figure 4, the SFEM has three components: the Restormer block-based encoder [52], the Lite Transformer block-based base transformer encoder and the Invertible Neural networks (INN) block-based detail CNN encoder. Their infrastructure is consistent with the design of encoder in CDDFuse [53]. If we denote SFEM by $\mathbf{E}(\cdot)$:

$$\begin{aligned}(I_A^D, I_A^B) &= \mathbf{E}(\mathcal{CAT}(Y_M, H_\lambda)) \\ (I_P^D, I_P^B) &= \mathbf{E}(\mathcal{CAT}(Y_P, \xi_\lambda))\end{aligned} \tag{9}$$

where Y_M denotes the scene information captured by MAD, Y_P represents the a priori RGB image, $\mathcal{CAT}(\cdot)$ denotes the concatenation operation along the channel dimension. I_A^D, I_A^B correspond to the detailed and basic features obtained from the Y_M, Y_P after processing by the Encoder, while I_P^D, I_P^B denote the detailed and basic features obtained from the Y_M and Y_P.Then, the output of SPEM can be represented as follows:

$$I_{out} = \mathcal{CAT}(\mathbf{F_D}(I_A^D, I_P^D), \mathbf{F_B}(I_A^B, I_P^B)) \tag{10}$$

where $\mathbf{F_B}$ and $\mathbf{F_D}$ denote the base/detail feature fusion which is similar to base(light Transformer block)/detail(INN block) feature extraction in the encoder. In order to align the features during feature extraction, we use L_{align} to guide the extraction [53]:

$$L_{align} = \frac{(\mathcal{CC}(I_A^D, I_P^D))^2}{\mathcal{CC}(I_A^B, I_P^B) + \kappa} \tag{11}$$

where $\mathcal{CC}(\cdot,\cdot)$ is the correlation coefficient operator, and here κ is set to 1.01 to ensure L_{align} always being positive.

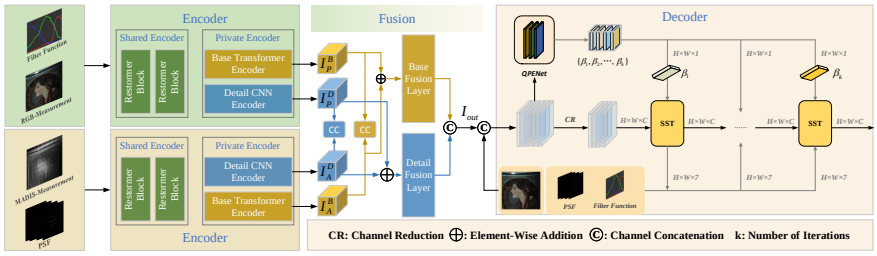


Fig. 4: The overall architecture of SSFCT. The SSFCT framework is structured to optimize feature extraction and alignment, processing the data garnered from each branch of the optical pathway. The Encoder output, integrated with RGB prior data, is concatenated along the channel dimension as the input of the Decoder.

4.2 Algorithm Framework

SSFCT consists of SFEM and decoder, here, we use the CSST algorithm as the decoder of the network to complete the self-aware fusion and reconstruction algorithm:

$$\hat{x} = \mathcal{D}(\mathcal{CAT}(I_{out}, Y_p, H_\lambda, \xi_\lambda)) \tag{12}$$

where $\mathcal{D}(\cdot)$ denotes the decoder CSST, $\hat{x}$ denotes the estimate of $X_\lambda(x_1, y_1)$. Building on this, the comprehensive loss function L_{UC} for SSFCT is formulated:

$$L_{UC} = L_{\text{Recon}} + \alpha \cdot L_{align} \tag{13}$$

where $L_{\text{Recon}} = RMSE(\hat{x}, GT)$ and α is set to 200 to balance the loss.

5 Experimental Analysis

Dataset. Similar to [8, 10, 21, 31, 33, 51], 28 wavelengths are selected from 450nm to 650nm and derived by spectral interpolation manipulation for the HSIs. We adopt two datasets, i.e., CAVE-1024 [36] and KAIST [14] for simulation. CAVE-1024 contains 205 HSIs of spatial size 1024×1024, while the KAIST contains 30 HSIs of spatial size 2704×3376. (Supp. Mat. Sec. 2, 4 for input preprocessing and testset content).

5.1 Simulation with Uncalibrated Inputs

Implementation Details. 10 scenes from the KAIST are selected for testing, while the CAVE-1024 and another 20 scenes from the KAIST are selected for training. In the simulation experiment of uncalibrated inputs, we set a maximum offset of 40 pixels for the 256×256 input, that is, the RGB prior is randomly sampled from the true value of $296 \times 296 \times 28$ and obtained by band-by-band integration (including 256×256 RGB prior). All models are trained with Adam [27] optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) using Cosine Annealing scheme [30] for 150 epochs on an RTX 3090 GPU.

Importantly, the test set's RGB guidance is derived by systematically selecting, filtering, and integrating information from $296 \times 296 \times 28$ HSIs. Each scene incorporates three types of RGB guidance: perspective variation, scaling, and calibrated.

Table 2: Quantitative Comparison of reconstruction results of different algorithms with uncalibrated inputs, Params, FLOPS, PSNR (dB) and SSIM are reported. The PSNR and SSIM of each scene are the mean results of the three different RGB guided images.

Algorithm	Params	GFLOPS	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	**Avg**
UNet [37]	23.29M	4.90	38.60	39.12	36.15	35.38	34.49	34.02	38.32	37.23	35.74	34.82	**36.46**
			0.970	0.974	0.957	0.910	0.942	0.935	0.968	0.953	0.954	0.947	**0.951**
Lambda-Net [34]	32.73M	23.29	36.49	37.63	34.46	31.65	33.23	32.77	37.25	36.48	35.59	34.85	**35.04**
			0.942	0.956	0.929	0.771	0.907	0.898	0.957	0.944	0.940	0.934	**0.918**
TSA-Net [33]	44.25M	91.47	40.38	40.59	37.42	36.34	**37.27**	35.68	40.15	38.01	37.31	38.00	**38.28**
			0.972	0.975	0.969	0.927	**0.963**	0.957	0.979	0.954	0.958	0.972	**0.962**
MST++ [9]	1.34M	17.57	40.81	40.26	37.87	35.83	35.79	36.55	**40.94**	39.76	38.56	37.92	**38.34**
			0.984	0.982	0.975	0.927	0.947	0.962	0.981	0.971	0.969	0.974	**0.966**
Restormer [52]	15.11M	87.54	**42.00**	**42.53**	**39.51**	**36.68**	36.93	**36.78**	40.75	**40.22**	**39.31**	**40.56**	**39.48**
			0.987	**0.987**	**0.980**	**0.936**	0.962	**0.968**	**0.982**	**0.977**	**0.978**	**0.985**	**0.974**
SSFCT-9stg (Ours)	6.62M	73.50	**42.67**	**43.56**	**40.80**	**40.89**	**39.24**	**40.55**	**43.99**	**39.99**	**40.69**	**40.93**	**41.33**
			0.988	**0.989**	**0.985**	**0.982**	**0.978**	**0.981**	**0.990**	**0.976**	**0.979**	**0.975**	**0.982**

Table 3: Comparisons of ADIS and PCCADIS (PSNR/SSIM)

Methods	**lambda-Net**	**TSA-Net**	**MST++**	**Restormer**	**CSST**	**SSFCT**	**Epoches**
ADIS	29.62	30.71	32.23	33.29	34.08	–	300
	0.860	0.915	0.942	0.948	0.958		
PCCADIS(UC)	35.04	38.28	**38.34**	39.48	–	**41.33**	150
	0.918	0.962	**0.966**	0.974		**0.982**	
PCCADIS(WC)	36.99	40.01	**41.43**	41.86	**41.97**	–	150
	0.935	0.958	**0.984**	0.984	**0.985**		

Quantitative Analysis. **(1) Different Algorithms.** Table 2 presents a comparative analysis between SSFCT and 6 methods including Unet [37], lambda-Net [34], TSA-Net [33] and MST++ [9] and Restormer [52] on 10 simulation scenes (each scene contains three forms of uncalibrated RGB bootstrapping). SSFCT shows the best experimental results on the PCCADIS reconstruction task with uncalibrated inputs, i.e., 41.33dB in PSNR and 0.982 in SSIM, markedly outperforms two recent SOTA methods Restormer and MST++ by 1.85dB and 2.99dB, which demonstrates the superiority, effectiveness and acceptability of the system and algorithm design.

(2) ADIS Vs. PCCADIS. As shown in the Table 3, SSFCT enables PCCADIS reconstruction without requiring calibration between different optical paths, yielding results similar to well-calibrated experimental setups in simulated experiments (achieving perfect calibration is virtually impossible in real experiments). It achieves a significant improvement of 7.25 dB in PSNR compared to serially encoded ADIS.

(3) Vs. DOE-based methods. We further compare ADIS, PCCADIS, and DOE-based SSI methods as shown in Fig. 5. PSFs generated in different bands for the DOE for simulations are depicted on the right side. It's important to note that SSFCT can also be applied to parallel DOE-based methods for fusion and reconstruction of uncalibrated inputs, yielding results similar to PCCADIS. However, the parallel DOE-based method still requires accurate calibration of each manufactured DOE. **This simultaneously demonstrates the migrability and effectiveness of SSFCT across different parallel systems, as well as the fact that the calibration-free framework, PCCADIS, is not inferior to parallel DOE-based methods in terms of reconstruction accuracy.**

	ADIS	PCCADIS	DOE-based	Parallel DOE-based
SSFCT / PSNR (dB)	34.08	41.33	36.58	41.79
SSFCT / SSIM	0.958	0.982	0.964	0.984

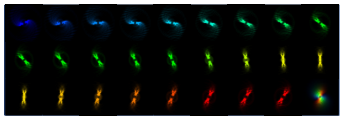


Fig. 5: Simulation experiments of SSFCT for DOE-based vs. SSFCT for ADIS and PCCADIS.

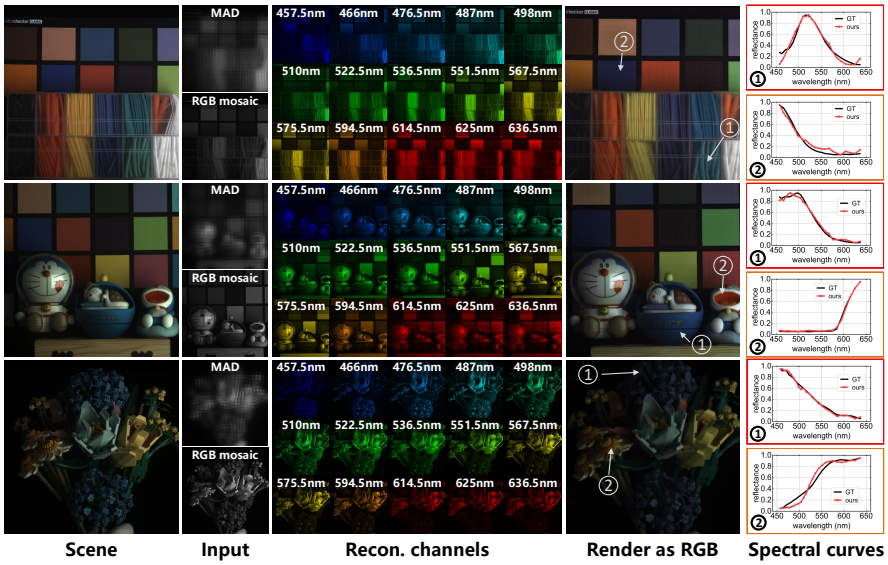


Fig. 6: Three scenes captured by our PCCADIS prototype camera, with spectral plots comparing our reconstruction to ground truth data from a point spectrometer.

5.2 Real Experiments

Implementation Details. Consistent with the methodologies outlined in Section 5.1, our experimental setup maintained identical training parameters and datasets.

For uncalibrated mode, we expand the input offsets to a maximum of 300 pixels horizontally and vertically for scenes with a resolution of 512×512 (additional 192 pixels for padding). This increase allowed for a more rigorous evaluation of the algorithm's reconstruction capabilities under misregistration, with RGB prior randomly sampled from GT with a resolution of $384 \times 384 \times 28$ and integrated into 256×256 resolution as inputs by perspective transformation and wave-by-wave integration. The constructed prototype incorporates MAD (comprising a monochrome camera and diffractive lens) and a standard RGB sensor paired with a conventional lens for prior acquisition, employing $450nm - 650nm$ bandpass filters in both optical pathways.

Result Analysis (WC mode). Fig. 6 exemplifies the PCCADIS's reconstruction capabilities with well-calibrated inputs, attaining recovery of high-resolution HSI data ($1824 \times 1824 \times 28$) from a single exposure. The system precisely delineates scene information, preserving edge clarity and high-frequency details.

Result Analysis (UC mode). When faced with uncalibrated inputs, as demonstrated in Fig. 7, the reconstruction leverages a priori segments derived from RGB data to guide

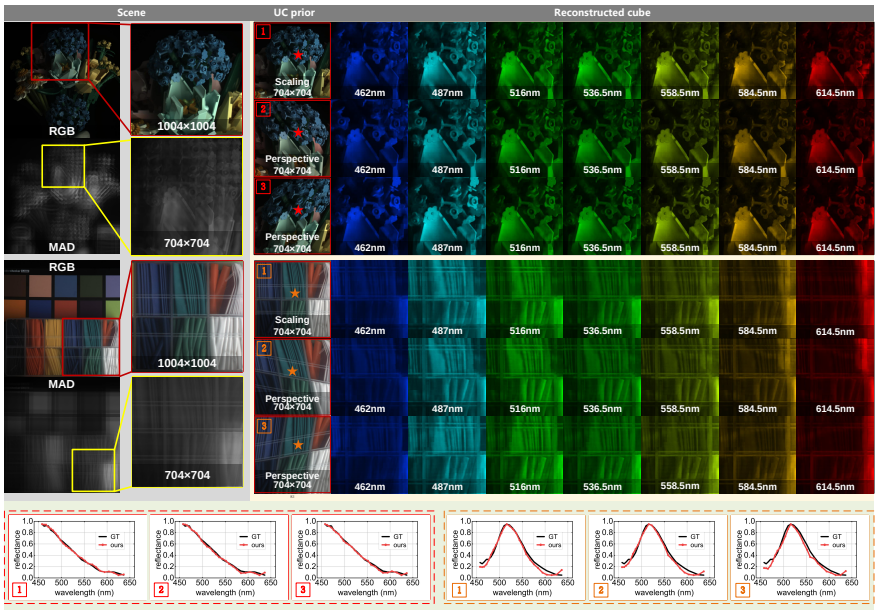


Fig. 7: We captured two scenes for the UC experiments. Scaling and perspective transformations were applied to intensify the scenes' uncalibrated features. The spectral plots compare our reconstruction results with the ground truth measured by a point spectrometer.

the process. This methodology ensures the clarity of scene information and the fidelity of spectral details, even when the guiding input is significantly distorted.

5.3 Simulation with Well-calibrated Inputs

10 scenes from the KAIST dataset are selected for testing, while the CAVE-1024 dataset is selected for training. All SSFCT models are trained with Adam [27] optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) using Cosine Annealing scheme [30] for 150 epochs on an RTX 3090 GPU. The initial learning rate is 4×10^{-4}.

Comparison of dispersion parameters. According to the dispersion law of the orthogonal mask, the color dispersion length between $\lambda_{\min}$ and $\lambda_{\max}$ can be expressed as: $\Delta x_m = \frac{f_2}{d}(\lambda_{\max} - \lambda_{\min})$, where d is the period of the orthogonal mask and f_2 is the imaging focal length, which means that the pattern of aperture diffraction spatial-spectral projection is directly determined by d and f_2. Here we fix f_2 and adjust the parameter d to simulate the reconstruction of PCCADIS under different dispersion conditions. Comparing the reconstruction results in Fig. 8, PCCADIS clearly shows stable reconstruction performance under different dispersion parameters.

5.4 Ablation study

The SSFCT-5stg model was trained using the CAVE dataset over 100 epochs, during which an ablation study was conducted to evaluate the functions of SFEM and L_{UC}.

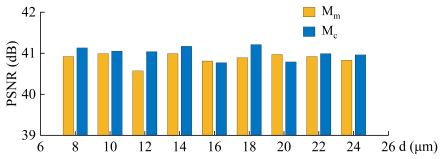


Fig. 8: Influence on PSNR of hole period d.

Training mode	PSNR(dB)	SSIM
SSFCT+L_{UC}(MAD+RGB)	37.38	0.959
SSFCT+L_{Recon}(MAD+RGB)	36.76	0.951
Decoder+L_{Recon}(MAD+RGB)	36.33	0.950
Decoder+L_{Recon}(only RGB)	35.53	0.945

Table 4: Result of ablation on SFEM, L_{UC} and imaging mode.

Table 5: Comparison of light throughput between PCCADIS and ADIS. (LT: light throughput)

Methods	ADIS(M_m)	ADIS(M_c)	Ours(M_m)	Ours(M_c)
LT	$\frac{1}{4}\times\frac{1}{3}$	$\frac{3}{4}\times\frac{1}{3}$	$\frac{1}{2}(\frac{1}{4}+\frac{1}{3})$	$\frac{1}{2}(\frac{3}{4}+\frac{1}{3})$
LT	8.3%	25.0%	29.2%	54.2%

Comparative analysis was also performed on the system's reconstruction results derived from RGB super-resolution.

Table 4 illustrates that the concurrent acquisition of MAD and RGB data significantly enhances the spectral reconstruction accuracy. Additionally, the experimental findings underscore the efficacy of L_{UC} and SFEM within the algorithm's architecture.

5.5 Light Throughput Analysis

The PCCADIS we proposed exhibits an augmented light throughput relative to the ADIS with serial encoding. For fully considering the number of sensors in parallel system, we add a **factor of 0.5** to the light Throughput calculation of each branch. As shown in Table 5, the light throughput of the PCCADIS system is demonstrated to be superior, delivering a light throughput that is 3.5-fold greater than that of the ADIS system when employing M_m, and 2.2-fold when utilizing M_c.

5.6 Evaluation of Real system

The real experimental setup is illustrated in Fig. 9(a). The beam splitter used in the prototype is unnecessary for far-field imaging. Please refer to Supp. Mat. Sec. 3 for the raw data captured in the real experiment.

Spectral resolution. In the actual experimental setup, the spectral sampling interval Δx_m is calculated as $\Delta x_m = \frac{f_2}{d}(\lambda_{\max} - \lambda_{\min}) = 25$ pixels, with a total of 28 reconstructed spectral bands. Consequently, the system boasts a spectral sampling interval of $8nm$ and achieves a reconstruction resolution of $7.14nm$.

Spatial resolution. To demonstrate the superiority of imaging spatial resolution, as exemplified in Figure 9(c), we conducted an experimental comparison by imaging a resolution chart (International Organization for Standardization standard ISO 12233). The reconstructed results indicate that PCCADIS yields sharper and more distinctly defined edges, leveraging both RGB- and MAD-derived information.

6 Further Discussions and Limitations

PSFs. The PSFs are approximately computed based on camera parameters according to Equation 6 and Equation 7, significantly simplifying the calibration process. However,

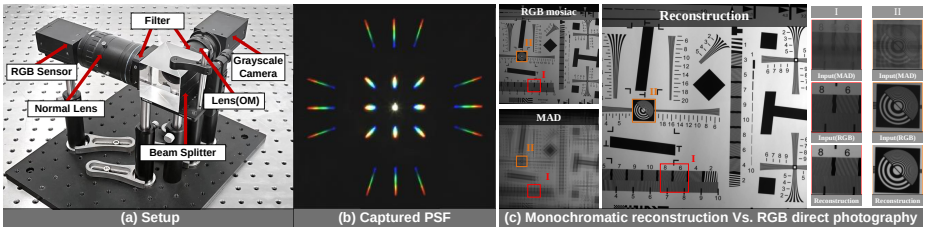


Fig. 9: (a) Hardware setup of the PCCADIS and (b) the chromatic PSF of the aperture diffraction. (RGB sensor: FLIR GS3-U3-51S5C, 2448×2048; Grayscale sensor: FLIR GS3-U3-41S4M, 2016×2016; interval d of the OM: $10\mu m$); (c) compare spatial resolution before and after reconstruction Monochromatic scene (the GT of resolution chart closely match the RGB prior).

it is important to note that there exists a discernible disparity between this approximation and the actual PSFs. Particularly, when dealing with longer wavelengths, a comprehensive forward modeling approach becomes necessary for accurately determining PSFs with the OM's interval d increasing. It's important to highlight a key distinction of PCCADIS from DOE-based methods: **the simplicity of the PSFs defined by the OM**.

Real Experiments. In our practical experimentation, we acknowledge that we did not fully account for the influence of the filter boundary effects at 450nm and 650nm on the sensor response. Consequently, this oversight led to deviations in the recorded spectra at these specific wavelengths during the real experiment.

System Deployment. The deployment of the reconstruction algorithm on edge and mobile terminals presents heightened demands on computational resources. To genuinely enable the fusion of aperture diffraction and arbitrary RGB camera data for the attainment of SSI, it becomes imperative to prioritize the development of a lightweight algorithm design. But there's no denying that SSFCT's design allows PCCADIS to utilize any RGB camera in a mobile terminal as a prior branch for SSI, enhancing hardware resource utilization at the edge.

7 Conclusions

In this paper, we propose a parallel-structured calibration-free SSI framework, PCCADIS, that achieves minimal parallel complexity with enhanced light throughput. To facilitate its deployment on mobile platforms, we develop the SSFCT method, which combines uncalibrated inputs fusion and spectral reconstruction, establishing a highly effective a priori guidance paradigm for mobile deployment. The efficacy and superior performance of SSFCT and PCCADIS are substantiated through a comprehensive suite of real-world experiments using a prototype device.

8 Acknowledgments

This research was supported by National Science Fund for Distinguished Young Scholars (62025108).

References

1. Akiyama, K., Alberdi, A., Alef, W., Asada, K., Azulay, R., Baczko, A.K., Ball, D., Baloković, M., Barrett, J., Bintley, D., et al.: First m87 event horizon telescope results. iv. imaging the central supermassive black hole. The Astrophysical Journal Letters **875**(1), L4 (2019)
2. Arguello, H., Pinilla, S., Peng, Y., Ikoma, H., Bacca, J., Wetzstein, G.: Shift-variant color-coded diffractive spectral imaging system. Optica **8**, 1424–1434 (2021)
3. Baek, S.H., Kim, I., Gutierrez, D., Kim, M.H.: Compact single-shot hyperspectral imaging using a prism. ACM Trans. Graph. **36**(6), 1–12 (2017)
4. Boominathan, V., Mitra, K., Veeraraghavan, A.: Improving resolution and depth-of-field of light field cameras using a hybrid imaging system. pp. 1–10. IEEE (2014)
5. Brady, D.J.: Optical imaging and spectroscopy. John Wiley & Sons (2009)
6. Brady, D.J., Gehm, M.E., Stack, R.A., Marks, D.L., Kittle, D.S., Golish, D.R., Vera, E., Feller, S.D.: Multiscale gigapixel photography. Nature **486**(7403), 386–389 (2012)
7. Brady, D.J., Pang, W., Li, H., Ma, Z., Tao, Y., Cao, X.: Parallel cameras. Optica **5**(2), 127–137 (2018)
8. Cai, Y., Lin, J., Hu, X., Wang, H., Yuan, X., Zhang, Y., Timofte, R., Van Gool, L.: Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 17502–17511 (2022)
9. Cai, Y., Lin, J., Lin, Z., Wang, H., Zhang, Y., Pfister, H., Timofte, R., Van Gool, L.: Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 745–755 (2022)
10. Cai, Y., Lin, J., Wang, H., Yuan, X., Ding, H., Zhang, Y., Timofte, R., Van Gool, L.: Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. arXiv preprint arXiv:2205.10102 (2022)
11. Cao, X.: Hyperspectral/multispectral imaging. In: Computer Vision: A Reference Guide, pp. 592–598. Springer (2021)
12. Cao, X., Du, H., Tong, X., Dai, Q., Lin, S.: A prism-mask system for multispectral video acquisition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2423–2435 (2011)
13. Cao, X., Tong, X., Dai, Q., Lin, S.: High resolution multispectral video capture with a hybrid camera system. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 297–304. IEEE (2011)
14. Choi, I., Kim, M., Gutierrez, D., Jeon, D., Nam, G.: High-quality hyperspectral reconstruction using a spectral prior. Tech. rep. (2017)
15. Dai, Q., Ma, C., Suo, J., Cao, X.: Computational hyperspectral imaging. In: JSAP Annual Meetings Extended Abstracts The 75th JSAP Autumn Meeting 2014. pp. 3821–3821. The Japan Society of Applied Physics (2014)
16. Descour, M., Dereniak, E.: Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. Applied optics **34**(22), 4817–4826 (1995)
17. Galvis, L., Lau, D., Ma, X., Arguello, H., Arce, G.R.: Coded aperture design in compressive spectral imaging based on side information. Applied optics **56**(22), 6332–6340 (2017)
18. Gehm, M.E., John, R., Brady, D.J., Willett, R.M., Schulz, T.J.: Single-shot compressive spectral imaging with a dual-disperser architecture. Optics express **15**(21), 14013–14027 (2007)
19. Hu, H., Zhou, H., Xu, Z., Li, Q., Feng, H., Chen, Y., Jiang, T., Xu, W.: Practical snapshot hyperspectral imaging with doe. Optics and Lasers in Engineering **156**, 107098 (2022)
20. Hua, X., Wang, Y., Wang, S., Zou, X., Zhou, Y., Li, L., Yan, F., Cao, X., Xiao, S., Tsai, D.P., et al.: Ultra-compact snapshot spectral light-field imaging. Nature communications **13**(1), 2732 (2022)
21. Huang, T., Dong, W., Yuan, X., Wu, J., Shi, G.: Deep gaussian scale mixture prior for spectral compressive imaging. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 16216–16225 (2021)

22. Huang, Z., Liu, J., Fan, X., Liu, R., Zhong, W., Luo, Z.: Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In: Eur. Conf. Comput. Vis. pp. 539–555. Springer (2022)
23. Hunicz, J., Piernikarski, D.: Investigation of combustion in a gasoline engine using spectrophotometric methods. In: Optoelectronic and Electronic Sensors IV. vol. 4516, pp. 307–314. SPIE (2001)
24. Jeon, D.S., Baek, S.H., Yi, S., Fu, Q., Dun, X., Heidrich, W., Kim, M.H.: Compact snapshot hyperspectral imaging with diffracted rotation (2019)
25. Jiang, Z., Zhang, Z., Fan, X., Liu, R.: Towards all weather and unobstructed multi-spectral image stitching: Algorithm and benchmark. pp. 3783–3791 (2022)
26. Kim, T., Lee, K.C., Baek, N., Chae, H., Lee, S.A.: Aperture-encoded snapshot hyperspectral imaging with a lensless camera. APL Photonics **8** (2023)
27. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
28. Lee, H.C.: Introduction to color imaging science. Cambridge University Press (2005)
29. Liang, P., Jiang, J., Liu, X., Ma, J.: Fusion from decomposition: A self-supervised decomposition approach for image fusion. In: Eur. Conf. Comput. Vis. pp. 719–735. Springer (2022)
30. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
31. Lv, T., Ye, H., Yuan, Q., Shi, Z., Wang, Y., Wang, S., Cao, X.: Aperture diffraction for compact snapshot spectral imaging. In: Int. Conf. Comput. Vis. pp. 10574–10584 (2023)
32. Ma, C., Cao, X., Tong, X., Dai, Q., Lin, S.: Acquisition of high spatial and spectral resolution video with a hybrid camera system. Int. J. Comput. Vis. **110**, 141–155 (2014)
33. Meng, Z., Ma, J., Yuan, X.: End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In: Eur. Conf. Comput. Vis. pp. 187–204. Springer (2020)
34. Miao, X., Yuan, X., Pu, Y., Athitsos, V.: l-net: Reconstruct hyperspectral images from a snapshot measurement. In: Int. Conf. Comput. Vis. pp. 4059–4069 (2019)
35. Monakhova, K., Yanny, K., Aggarwal, N., Waller, L.: Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array. Optica **7**(10), 1298–1307 (2020)
36. Park, J.I., Lee, M.H., Grossberg, M.D., Nayar, S.K.: Multispectral imaging using multiplexed illumination. In: Int. Conf. Comput. Vis. pp. 1–8. IEEE (2007)
37. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
38. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: Int. Conf. Comput. Vis. pp. 2564–2571. Ieee (2011)
39. Schnapf, J., Kraft, T., Baylor, D.: Spectral sensitivity of human cone photoreceptors. Nature **325**(6103), 439–441 (1987)
40. Taruttis, A., Ntziachristos, V.: Advances in real-time multispectral optoacoustic imaging and its applications. Nature photonics **9**(4), 219–227 (2015)
41. Trinidad, M.C., Brualla, R.M., Kainz, F., Kontkanen, J.: Multi-view image fusion. In: Int. Conf. Comput. Vis. pp. 4101–4110 (2019)
42. Wagadarikar, A., John, R., Willett, R., Brady, D.: Single disperser design for coded aperture snapshot spectral imaging. Applied optics **47**(10), B44–B51 (2008)
43. Wang, D., Liu, J., Fan, X., Liu, R.: Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration. arXiv preprint arXiv:2205.11876 (2022)
44. Wang, L., Xiong, Z., Gao, D., Shi, G., Wu, F.: Dual-camera design for coded aperture snapshot spectral imaging. Applied optics **54**(4), 848–858 (2015)

45. Wang, L., Xiong, Z., Gao, D., Shi, G., Zeng, W., Wu, F.: High-speed hyperspectral video acquisition with a dual-camera architecture. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 4942–4950 (2015)
46. Wang, T.C., Zhu, J.Y., Kalantari, N.K., Efros, A.A., Ramamoorthi, R.: Light field video capture using a learning-based hybrid imaging system. ACM Trans. Graph. **36**(4), 1–13 (2017)
47. Xiong, Z., Wang, L., Li, H., Liu, D., Wu, F.: Snapshot hyperspectral light field imaging. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 3270–3278 (2017)
48. Xu, H., Hu, H., Chen, S., Xu, Z., Li, Q., Jiang, T., Chen, Y.: Hyperspectral image reconstruction based on the fusion of diffracted rotation blurred and clear images. Optics and Lasers in Engineering **160**, 107274 (2023)
49. Yang, Z., Albrow-Owen, T., Cai, W., Hasan, T.: Miniaturization of optical spectrometers. Science **371**(6528), eabe0722 (2021)
50. Yuan, L., Song, Q., Liu, H., Heggarty, K., Cai, W.: Super-resolution computed tomography imaging spectrometry. Photonics research **11**(2), 212–224 (2023)
51. Yuan, X., Brady, D.J., Katsaggelos, A.K.: Snapshot compressive imaging: Theory, algorithms, and applications. IEEE Signal Processing Magazine **38**(2), 65–88 (2021)
52. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 5728–5739 (2022)
53. Zhao, Z., Bai, H., Zhang, J., Zhang, Y., Xu, S., Lin, Z., Timofte, R., Van Gool, L.: Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 5906–5916 (2023)