

Thesis for Master's Degree

Development of Deep Learning Based
Computational Hyperspectral Camera Using
Microlens Array and Multilayer Thin Films

Jioh Lee

Artificial Intelligence Graduate School

Gwangju Institute of Science and Technology

2025

석사학위논문

마이크로 렌즈 배열과 다층 박막 필터 배열을
이용한 딥러닝 기반 초분광 카메라 개발

이지오

AI대학원

광주과학기술원

2025

Development of Deep Learning Based Computational Hyperspectral Camera Using Microlens Array and Multilayer Thin Films

Advisor: Heung-No Lee

by

Jioh Lee

Artificial Intelligence Graduate School

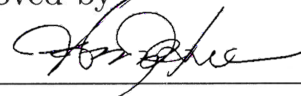
Gwangju Institute of Science and Technology

A thesis submitted to the faculty of the Gwangju Institute of Science and Technology in partial fulfillment of the requirements for the degree of Master of Science in the Artificial Intelligence Graduate School

Gwangju, Republic of Korea

Dec 2, 2024

Approved by



Professor Heung-No Lee

Committee Chair

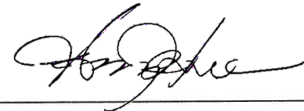
Development of Deep Learning Based
Computational Hyperspectral Camera Using
Microlens Array and Multilayer Thin Films

Jioh Lee

Accepted in partial fulfillment of the requirements for
the degree of Master of Science

Dec 2, 2024

Committee Chair



Prof. Heung-No Lee

Committee Member



Prof. Kyoobin Lee

Committee Member



Prof. Hoon Hahn Yoon

Dedicated to my family.

MS/AI Jioh Lee. Development of Deep Learning Based Computational Hyper-
20224022 spectral Camera Using Microlens Array and Multilayer Thin Films. Ar-
tificial Intelligence Graduate School. 2025. 39p. Advisor: Prof. Heung-No
Lee.

Abstract

Hyperspectral (HS) imaging systems have been extensively studied across various research fields, including remote sensing, biomedical imaging, and medical diagnostics. However, the bulkiness and time-consuming imaging process of conventional HS imaging systems have hindered their integration into mobile applications. In this study, we present a deep learning (DL)-based computational HS camera that integrates a microlens array (MLA) and a multilayer thin film filter array, enabling snapshot HS imaging with compact and low-cost configuration. Instead of implementing multiple filter arrays with multilayer structures across the complementary metal-oxide semiconductor (CMOS) image sensor which leads to transmittance inconsistencies and complicates the recovery process, we mount an MLA onto a single filter array with a CMOS image sensor, effectively reducing reconstruction complexity. We employed DL-based point-wise spectral reconstruction methods to recover an HS image. A synthetic spectral dataset was used to train the model. To train the model by reflecting the measurement noise in the real environment, we propose a data augmentation scheme using noise levels

sampled from a uniform distribution. Our experimental results by the best-performing model showed a 17-fold increase in reconstruction speed and a twofold reduction in reconstruction error with the proposed data augmentation scheme. By achieving snapshot structure and efficient reconstruction with compact and low-cost hardware configuration, the proposed system offers a promising solution for mobile applications such as on-site detection and consumer electronics.

©2025

Jioh Lee

ALL RIGHTS RESERVED

국 문 요 약

초분광 이미징 시스템은 원격 감지, 의료 영상 및 진단을 포함한 다양한 연구 분야에 서 광범위하게 연구되어 왔다. 그러나 기존 초분광 이미징 시스템의 큰 부피와 긴 시간이 소요되는 이미지 획득 과정으로 인해 초분광 카메라의 모바일 기기로의 응용이 제한되 어 왔다. 이 연구에서는 마이크로렌즈 배열 (MLA)과 다층 박막 필터 배열을 통합하여 저비용/초소형의 단노출 시간에 초분광 이미징이 가능한 딥 러닝 기반 계산 초분광 카메 라를 제안한다. 다층 구조의 필터 배열을 이미지 센서에 반복적으로 구현하여 투과율의 불일치를 야기하고 이미지 복원 과정의 복잡도를 높이는 대신, 이미지 센서에 단일 필터 배열과 MLA를 장착하여 초분광 이미지 복원 과정의 복잡도를 효과적으로 줄였다. 딥 러닝 기반 point-wise 스펙트럼 복원 방법을 사용하여 초분광 이미지를 복원했고, 학습을 위하여 합성 스펙트럼 데이터 세트를 구성하였다. 실제 환경에서 측정 노이즈를 반영하 여 모델을 학습하기 위해 균일한 분포에서 샘플링한 노이즈 레벨을 사용하는 데이터 증강 방식을 제안한다. 실험 결과, 가장 성능이 좋은 모델을 통해 이미지 복원 속도가 17배 증가하고 제안된 데이터 증강 방식을 통해 복원의 오류가 2배 감소하였다. 제안된 시스템은 저비용/초소형의 하드웨어 구성 및 단노출 시간에 초분광 이미징이 가능한 구조와 효율적인 초분광 이미지 복원을 통해 실시간 감지 및 소비자 가전과 같은 모바일

응용에 효과적인 해법을 제공한다.

©2025

이 지 오

ALL RIGHTS RESERVED

Contents

Abstract (English)	i
Abstract (Korean)	iii
List of Contents	v
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Motivation	1
1.2 Contribution	4
2 Related work	6
2.1 Multispectral Imaging with Microlens Array	6
2.2 Computational Spectral Imaging	7
2.3 Filter-based Computational Spectral Imaging	8
2.4 Summary of Related Works	10
3 Method	11
3.1 System Architecture	11
3.2 System Model	16
3.3 DL-based point-wise spectral reconstruction	18
3.3.1 Training Dataset	18
3.3.2 Data Augmentation with Uniform Noise Level	19
4 Experimentation	21
4.1 Experimental Setting	21
5 Results and Discussion	25
5.1 Results of Spectral Reconstruction	25
5.2 Results of Hyperspectral Image Reconstruction	28
6 Limitations	31
7 Conclusion	32

References	34
A Spectral Reconstruction Results of 18 Colors in the Checkerboard	40
Acknowledgements	44

List of Tables

4.1	Summary of experimental setup.	24
5.1	Average MSE values of reconstruction results in 18 colors on the checkerboard according to data augmentation configuration.	25
5.2	Time duration of HS imaging.	29

List of Figures

3.1	Illustration of hardware configuration. The MTFs and MLA are attached to the CMOS image sensor. 36 sub-images can be captured in a single exposure. The aperture stop is applied to prevent crosstalk between the sub-images.	12
3.2	Schematic of the imaging process of the proposed system. The sub-images are aligned based on the disparity and are cropped to obtain a measurement cube with 36 channels. The DL-based spectral reconstruction networks are used to recover the HS cube.	12
3.3	Image of the proposed module.	15
3.4	Photograph of the MLA and the MTFs.	15
3.5	The characteristics of MTFs. Enlarged image of MTFs (left), examples of transmission pattern of 6-th and 12-th filter (center), and the heatmap of transmission matrix (right).	15
4.1	Experimental setup for the proposed computational HS camera.	22
4.2	Raw image captured by the proposed camera module.	22
5.1	Image of the color checkerboard (left) and colorized image (rest) of reconstructed HS cube using 500, 550, and 650 nm band images.	27
5.2	Reconstructed spectra of four colors of the checkerboard (#1, #6, #10, and #15 shown in Fig. 5.1). The MSE between the reconstructed and reference spectra is written on the upper side of each graph.	27

5.3	Image of the clay doll	30
5.4	Reconstructed spectra at two points of the clay doll (#1 and #2 as shown in 5.3).	30
5.5	Colorized images using 500, 550, and 650 nm band images and grayscale images corresponding to different wavelengths (500, 550, 600, 650, 700, and 750 nm).	30
A.1	Color checkerboard with 18 colors.	40
A.2	Reconstruction results without data augmentation.	41
A.3	Reconstruction results with fixed noise level data augmentation.	42
A.4	Reconstruction results with uniform noise level data augmentation.	43

Chapter 1

Introduction

1.1 Motivation

A spectrum contains rich information that reflects the unique properties of materials, enabling their identification and analysis. Hyperspectral (HS) camera, which provides both spectral and spatial information, has been broadly studied in various research fields, such as remote sensing [1], precision agriculture [2, 3], medical imaging [4, 5, 6], and face recognition [7, 8]. HS cameras reveal the details of the target objects beyond the limits of human cognition by offering spectral-spatial information. However, conventional HS cameras that rely on scanning methods [9] face challenges, including a time-consuming imaging process, bulky size, high-cost configuration, and high calibration effort. These limitations have hindered the integration of HS cameras into mobile applications. Therefore, developing a snapshot spectral imaging system that is compact and low-cost is essential for the widespread adoption of HS cameras.

One approach to achieving snapshot spectral imaging is using an array of cameras and filters [10, 11, 12], inspired by the compound-eye imaging system. By arranging optical filters with cameras, these systems can capture multiple images with properties according to specific filters in a single exposure. Moreover, researchers have adopted microlens array (MLA) [13, 14, 15, 16] into their systems for further compactness.

Spectral imaging systems incorporating MLA and filter array are not only easy to fabricate but also have flexibility for integration with detector arrays such as single complementary metal-oxide semiconductor (CMOS) image sensors, thereby reducing the cost of mass production and the size of the imaging system. This advantage facilitates the practical deployment of spectral imaging systems across various applications. However, these filter arrays utilize band-pass filters which limit the spectral resolution and light sensitivity. To mitigate these issues, the systems proposed in [10, 11] adopt the computational approach for spectral reconstruction.

Extensive research in computational spectral imaging (CSI) systems has been conducted to achieve snapshot structures with high spectral resolution [17]. Under the theory of Compressed Sensing (CS) [18], the CSI systems encode input spectra into compressed measurements using optical elements. Then, reconstruction algorithms are employed to recover spectral images from measurements by solving optimization problems. These systems adopt iterative methods such as l_1 -norm minimization [19] and two-point iterative thresholding (TwIST) [20]. These approaches are widely used for spectral reconstruction in CSI systems. However, iterative algorithms often require significant computational complexity and can be sensitive to the noise in the measurements, which can lead to inaccurate reconstruction. As an alternative to iterative approaches, CSI systems with deep learning (DL)-based spectral reconstruction methods have been introduced to mitigate these issues [21, 22, 23]. DL-based methods can effectively replace traditional algorithms by providing fast, accurate, and robust spectral reconstruction. While numerical optimization methods rely on predefined models

and the assumption of sparsity, DL-based methods learn directly from data, allowing them to achieve better performance in practical applications.

In recent years, optical filter-based computational HS imaging systems have been developed, leveraging unique filter arrays and reconstruction algorithms to capture detailed spectral data [19, 24, 25, 26]. Zhang et al. [24] introduced the Broadband Encoding Stochastic (BEST) camera with a sliding 16-filter array and deep neural network (DNN)-based reconstruction, while Kim et al. [19] used a Multilayer Thin Film (MTF) filter array with a pinhole model and iterative optimization for spectral reconstruction. Though compact and low-cost configuration, both approaches require multiple exposures, increasing imaging time. Especially in [19], applying numerical optimization to each individual point of the image for spectral reconstruction increases computational demands, resulting in a reconstruction time of approximately 1.8 hours. In contrast, Yako et al [25] and Wen et al [26] realized spectral imaging with a single exposure by monolithically integrating filter arrays with a CMOS image sensor. Yako et al. [25] integrated Fabry-Pérot filter arrays onto a CMOS image sensor for single-shot imaging, reconstructing HS images using TwIST. Wen et al. [26] employed all-dielectric film filter arrays and designed a specialized DL-based model for HS image reconstruction, named the Massively Parallel Networks (MP-Net). However, these integrated designs require multiple implementations of the filter array with a multi-layer structure across the CMOS image sensor, causing discrepancies in transmittance among filter arrays. This inconsistency in fabrication complicates reconstruction algorithms for both DL-based and numerical optimization methods. Either of these systems

suffers from multiple exposures or transmittance discrepancy, with both factors leading to increased complexity of the imaging and reconstruction.

1.2 Contribution

To mitigate these issues identified above, we introduce a DL-based computational HS camera that integrates an MLA mounted onto an MTF filter array with a CMOS image sensor, enabling snapshot HS imaging with a low-cost and compact configuration. We utilize MLA to capture modulated images in a single exposure, thereby avoiding implementing multiple filter arrays with multilayer structures that cause transmittance discrepancy and increase reconstruction complexity. Instead of using time-consuming numerical optimization algorithms for the reconstruction of HS images from the captured measurements, we adopt DL-based spectral reconstruction networks, including CNN, ResCNN, U-Net, and ReSpecNN. To reflect measurement noise in real environments to our training dataset, we propose a data augmentation scheme with noise levels sampled from a uniform distribution. Our work presents a fully tested and verified solution based on cutting-edge DL technology and HS imaging hardware developed in our own research laboratory. We are confident in the practicality and feasibility of the proposed method for mobile applications such as electronic devices and on-site detection, addressing various challenges, including face anti-spoofing and food inspection.

The main contributions of our work are listed as follows:

- A new snapshot CSI system using an MLA and MTF filter array is proposed, which has a compact size and cost-effective hardware configuration.

- We adopt DL-based spectral reconstruction networks for spectrometers to recover HS images and reduce reconstruction time.
- To reflect measurement noise in the real environment and improve the accuracy of spectral reconstruction networks, we propose a data augmentation scheme with noise levels sampled from a uniform distribution.

Chapter 2

Related work

In this section, we introduce various approaches to snapshot spectral imaging aimed at achieving a compact, low-cost design while maintaining high spectral resolution. These approaches encompass the optical structures of spectral imaging systems and spectral reconstruction techniques.

2.1 Multispectral Imaging with Microlens Array

Optical systems using MLA have been introduced for the past decades to achieve the compact design of spectral imaging systems. R. Shogenji et al. [14] proposed a multispectral imaging system based on a thin observation module by bound optics (TOMBO), inspired by compound eye. By mounting the inference filter array onto MLA, each microlens can capture an image modulated by each filter. They utilized 7 bandpass filters to realize multispectral imaging. M. Hubold et al. [27] present multi-aperture system approaches to achieve snapshot multispectral imaging. They attached a slanted linear variable filter (LVF) onto 11×6 MLA. Instead of placing LVF and MLA in parallel, they tilted the LVF filter to obtain a linear spectral sampling between neighboring channels of LVF. Their demonstrations show that the proposed system can capture 66 spectral bands ranging from 450 nm to 850 nm with a compact module which has a size of $60 \times 60 \times 28$ mm³. X. Yu et al. [28] introduce a multispectral-panchromatic

imager (MSPI) that utilizes a multispectral-panchromatic filter array (MSPFA) for capturing both spectral images and panchromatic images. Unlike the multispectral filter array (MSFA), their configuration includes all-band pass filters to acquire high signal-to-noise ratio (SNR) images, enhancing the quality of spectral images. These systems achieved snapshot spectral imaging while maintaining compact and low-cost configurations. However, the filters adopted in these systems transmit a specific range of wavelength thereby the spectral resolution is limited to the number of filters.

2.2 Computational Spectral Imaging

The CSI system is one approach for the high spectral resolution in a snapshot spectral imaging system. Among CSI methods, the CASSI is the most popular approach [29, 30]. In CASSI, dispersive elements and coded aperture play crucial roles in capturing spectral and spatial information. The dispersive elements, such as prism, are responsible for separating the incoming light into its different wavelength components by exploiting the principle of dispersion. By spatially dispersing the spectral contents of the scene, the dispersive elements ensure spectral variations at each location. Following this, the coded aperture is applied for a unique spatial encoding to the incoming light, enabling CASSI to project 3D scenes with spatial and spectral information into 2D measurements. To retrieve a full HS image from these measurements, numerical optimization methods are used, enabling high-resolution spectral imaging in a single exposure. However, these optimization methods are computationally heavy and vulnerable to measurement noise of the system, reducing the effectiveness of spec-

tral reconstruction. Moreover, the long optical path and bulky structure constrain its potential for miniaturization. To mitigate these issues, Y. Zhao et al. [22] applied a random printed mask and convolutional neural network (CNN), thereby reducing the optical path of the system and increasing the reconstruction accuracy regardless of the measurement noise. Also, Z. Yu et al. [23] employed a reflective coded aperture to reduce the optical path and adopted a DL-based reconstruction method based on 3D CNN and U-Net structure, enhancing the reconstruction results. Despite these efforts in the development of CASSI, achieving compactness while maintaining efficient spectral reconstruction remains a challenge.

2.3 Filter-based Computational Spectral Imaging

Recently, the optical filter array-based HS camera has emerged as a distinct approach in the field of CSI. Unlike conventional systems that utilize bandpass filters, these systems adopt the random filter array [31] where each filter within the array has unique transmittance over the broad wavelength range. This feature allows reconstruction algorithms to recover high-resolution spectral information and to cover a wide working range. W. Zhang et al. [24] proposed a broadband encoding stochastic (BEST) camera. They placed 16 random filters in front of a CCD camera and slid the filter array to capture the images modulated by each filter. A deep neural network (DNN) was applied to reconstruct the input HS image. C. Kim et al. [19] utilized a pinhole camera model and MTF filter array with 36 filters to capture an HS image. They overlap the filters with a pinhole and slide them to capture modulated images. To recover

the HS image, a numerical optimization method was applied to each point of the captured image. These methods successfully achieved both compactness and high spectral resolution. However, the sliding procedure necessitates multiple exposures to capture the full HS image. For further compactness and snapshot HS imaging, systems incorporating the filter array monolithically integrated onto a CMOS image sensor have been proposed. M. Yako et al. [25] integrated a random array of Fabry-Pérot filters across all pixels of the CMOS image sensor. They applied the TwIST [20] algorithm to recover HS images, enabling fast and high spatial resolution HS imaging in a single exposure. J. Wen et al. [26] proposed an HS imager using all-dielectric films which is fully integrated with a CMOS image sensor. They also proposed a reconstruction network, massively parallel networks (MP-Net) to mitigate the issue of variability of transmittance. These fully integrated structures have advantages including compactness and spatial resolution. However, the structure in which the filter array with multilayer structure is fully integrated with the CMOS sensor requires multiple implementations of the filter array across the entire area of the CMOS image sensor. This process is prone to fabrication errors, causing discrepancies between the designed and actual transmittance of the filter array, thereby decreasing the accuracy of HS image reconstruction. Consequently, reconstruction algorithms must consider these variations, which can complicate reconstruction algorithms.

2.4 Summary of Related Works

Previous filter-based spectral imaging systems utilizing band-pass filters are compact and low-cost but limited in spectral resolution. Moreover, existing filter-based CSI systems suffer from multiple exposures or multiple implementations of filter arrays with multilayer structures, increasing imaging time and reconstruction complexity. To mitigate these issues, we utilize an MLA and an MTF filter array to realize the snapshot structure. Instead of implementing multiple filter arrays with multilayer structures onto a CMOS image sensor which complicates reconstruction, we mount an MLA onto an MTF filter array with a CMOS image sensor. We employ DL-based spectral reconstruction networks [32, 33, 34, 35] and apply them to each individual point of a measurement cube to recover the HS image. To reflect measurement noise in a real device and enhance reconstruction performance, we utilize a data augmentation technique. This design achieves snapshot structure and efficient spectral reconstruction while maintaining compact and low-cost hardware configuration, thereby highlighting the suitability of the proposed system for mobile applications such as on-site detection and consumer electronics.

Chapter 3

Method

In this section, we discuss the proposed method. First, we describe the details of the system concept and components including the MLA and the MTF filter array. Second, we describe the system model to represent the proposed HS camera. Finally, we describe the DL-based point-wise spectral reconstruction method for HS imaging.

3.1 System Architecture

Fig. 3.1 and 3.2 show the overall concept of the proposed system and HS imaging procedure. Fig. 3.1 illustrates the concept of the proposed HS camera. An MTF filter array and MLA are mounted onto a CMOS sensor, where each microlens corresponds to a single filter within the array comprising a unit. The array consists of 36 units, enabling the capture of 36 sub-images, each modulated by the corresponding filter's transmittance. An aperture stop is applied to prevent crosstalk between adjacent sub-images. Fig. 3.2 shows the imaging process. As shown in 6×6 sub-images in Fig. 3.2, disparities are observed between the sub-images. Therefore, alignment is performed using patch-matching algorithms as described in [10]. After alignment, the sub-images are stacked into a cube and cropped to a size of 250×250 , resulting in a measurement cube with a size of $250 \times 250 \times 36$. The DL-based spectral reconstruction network is applied to all points of the measurement cube to recover the HS cube in the range

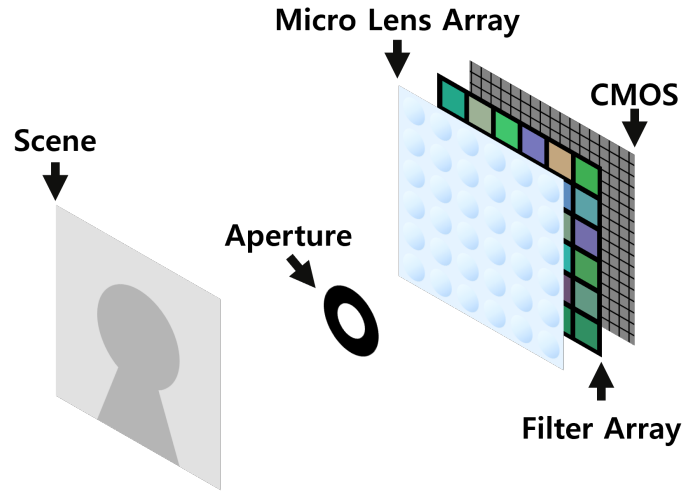


Figure 3.1: Illustration of hardware configuration. The MTFs and MLA are attached to the CMOS image sensor. 36 sub-images can be captured in a single exposure. The aperture stop is applied to prevent crosstalk between the sub-images.

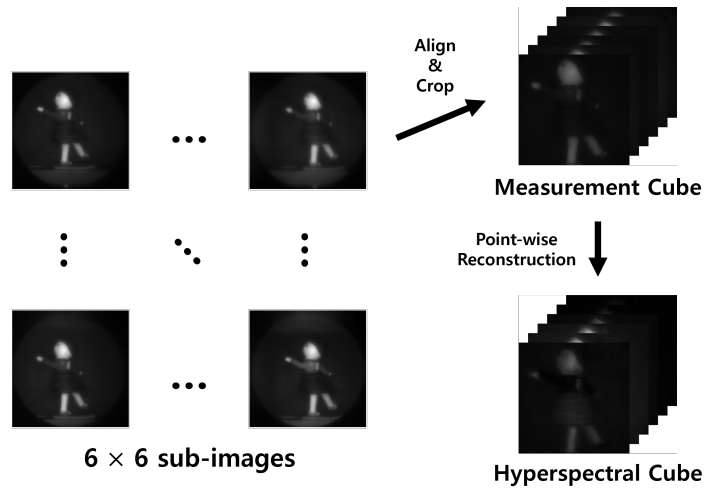


Figure 3.2: Schematic of the imaging process of the proposed system. The sub-images are aligned based on the disparity and are cropped to obtain a measurement cube with 36 channels. The DL-based spectral reconstruction networks are used to recover the HS cube.

from 500 nm to 750 nm with 1 nm spacing. Therefore, we can obtain an HS image with 251 spectral bands.

Fig. 3.3 is the photo of the proposed HS camera and Fig. 3.4 shows the scale of the fabricated MLA and MTF filter array. The MLA has a size of $5 \times 5 \text{ mm}^2$ with each microlens featuring a radius of $250 \text{ }\mu\text{m}$, a focal length of 2 mm, a lenslet pitch of $700 \text{ }\mu\text{m}$, a radius of curvature (ROC) of 1 mm, and an F-number of 4. The space between adjacent microlenses is $200 \text{ }\mu\text{m}$. We employed an MTF filter array in [19] as the filter array of the proposed HS camera. The scale of the MTF filter array is $4.5 \times 4.5 \text{ mm}^2$. Each MTF filter has a size of $400 \times 400 \text{ }\mu\text{m}^2$ and is spaced $300 \text{ }\mu\text{m}$ apart from the adjacent filter. Fig. 3.5 shows the enlarged image of the MTF filter array (left), the transmittance of the 6th and 12th filters (center), and the heatmap of the spectral sensitivity matrix (right). The transmittance of the filters was measured over the wavelength range of 500 nm to 750 nm, and results exhibit distinct patterns among the filters.

It is important to highlight that, in CS, it is well established that a random matrix, that is sensing matrix with the randomly sampled entries, can capture sufficient information about the input signal to recover it from the small number of measurements. Such a sensing matrix can be realized for filter-based computational spectrometers by designing the transmittance pattern of filters to produce random signal-like fluctuations [31]. The fabrication of MTF filter arrays was conducted using the stencil lithography process to follow this design principle [19], allowing each filter to have a unique pattern of transmittance as shown in Fig. 3.5. This process involves changing the number and

thickness of layers during the deposition process. To achieve the desired structure, 2 to 5 layers of different MTF filters were selectively removed using shadow masks. The materials with high refractive index (TiO_2) and low refractive index (SiO_2) were alternately deposited with each material deposited at different thicknesses. This deposition process was repeated 19 times to complete the MTF filter array.

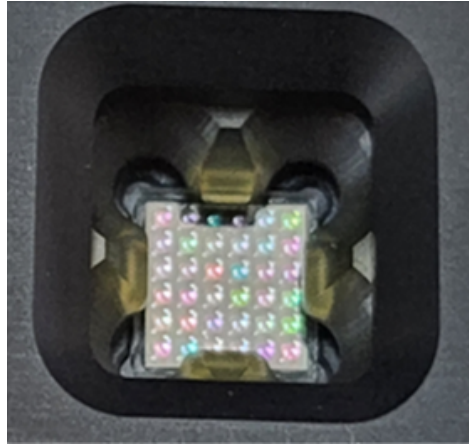


Figure 3.3: Image of the proposed module.

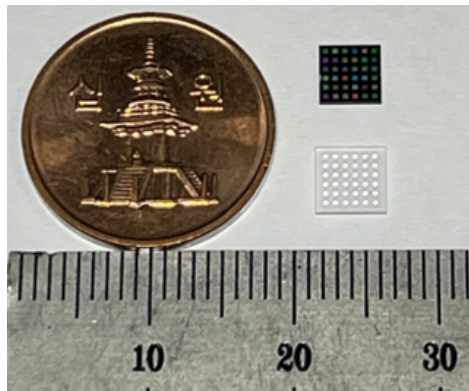


Figure 3.4: Photograph of the MLA and the MTFs.

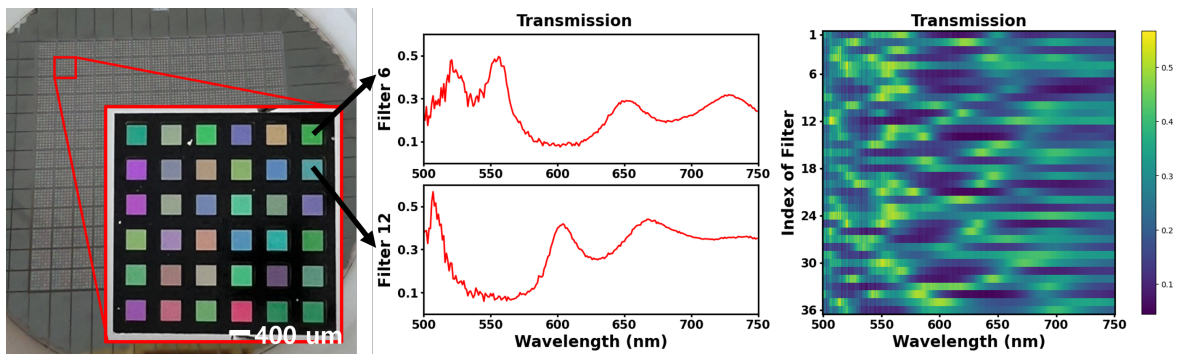


Figure 3.5: The characteristics of MTFs. Enlarged image of MTFs (left), examples of transmission pattern of 6-th and 12-th filter (center), and the heatmap of transmission matrix (right).

3.2 System Model

Let us denote that $X_i(r, c; \lambda)$ is an input spectrum at position (r, c) to be captured by the i -th unit. We denote λ as a continuous wavelength. The measurement $g_i(r, c)$ is represented as follows.

$$g_i(r, c) = \int R_i(\lambda) X_i(r, c; \lambda) d\lambda \quad (3.1)$$

Where spectral sensitivity $R_i(\lambda)$ can be represented as $T_i(\lambda)Q(\lambda)$. The $T_i(\lambda)$ is a transmission function of i -th MTF filter under i -th microlens of MLA and $Q(\lambda)$ is a quantum efficiency of the CMOS image sensor. Since the detector in the CMOS image sensor is pixelated with pixel size Δ , the measurement $g_i(r, c)$ is sampled to $y_i(n, m)$ as follows.

$$\begin{aligned} y_i(n, m) &= \iint g_i(r, c) \text{rect}\left(\frac{r}{\Delta} - n, \frac{c}{\Delta} - m\right) dr dc \\ &= \iiint R_i(\lambda) X_i(r, c; \lambda) \text{rect}\left(\frac{r}{\Delta} - n, \frac{c}{\Delta} - m\right) dr dc d\lambda \\ &= \int R_i(\lambda) X_i(n, m; \lambda) d\lambda \end{aligned} \quad (3.2)$$

Where the sampled spectrum by i -th unit is $X_i(n, m; \lambda) = \iint X_i(r, c; \lambda) \text{rect}\left(\frac{r}{\Delta} - n, \frac{c}{\Delta} - m\right) dr dc$. Let us consider the K spectral bands in discrete form as $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_K]$, ($K = 251$), then we represent Eq. (3.2) into a linear equation as follows.

$$\begin{aligned}
y_i(n, m) &= \sum_{k=1}^K R_i(\lambda_k) X_i(n, m; \lambda_k) \\
&= \mathbf{R}_i \mathbf{X}_i(n, m)
\end{aligned} \tag{3.3}$$

Where $\mathbf{R}_i \in \mathbb{R}^{1 \times K}$ is a row vector with spectral sensitivity of i -th unit and $\mathbf{X}_i(n, m) \in \mathbb{R}^{K \times 1}$ is an incident spectrum at pixel position (n, m) to be measured by i -th unit. Considering that we have 36 units, there are linear equations for $i = 1, \dots, 36$. Besides, lenses are placed at different positions of MLA, resulting in shifts of the object between the sub-images as shown in Fig. 3.2. To align them in the exact location, we calculate disparities between reference sub-image ($i = 1$) and destination sub-images ($i = 2, \dots, 36$) using patch matching as described in [10]. We denote the disparity between the reference sub-image and the i -th destination sub-image in the direction of n axis as d_i^n and of m axis as d_i^m . Then we calculate 35 disparities in each direction. With these disparities, we obtain aligned measurements as $y_i(n - d_i^n, m - d_i^m) = \mathbf{R}_i \mathbf{X}_i(n - d_i^n, m - d_i^m)$. Since we aligned the measurements with the reference sub-image, we rewrite Eq. (3.3) as $y_i(n - d_i^n, m - d_i^m) = \mathbf{R}_i \mathbf{X}_1(n, m) = \mathbf{R}_i \mathbf{X}(n, m)$. Therefore, we express 36 linear equations in one matrix form as follows.

$$\begin{aligned}
\begin{bmatrix} y_1(n, m) \\ \vdots \\ y_{36}(n - d_{36}^n, m - d_{36}^m) \end{bmatrix} &= \begin{bmatrix} R_1(\lambda_1) & \cdots & R_1(\lambda_K) \\ \vdots & \ddots & \vdots \\ R_{36}(\lambda_1) & \cdots & R_{36}(\lambda_K) \end{bmatrix} \begin{bmatrix} X(n, m; \lambda_1) \\ \vdots \\ X(n, m; \lambda_K) \end{bmatrix} \\
\mathbf{y}(n, m) &= \mathbf{R} \mathbf{X}(n, m)
\end{aligned} \tag{3.4}$$

Where $\mathbf{y}(n, m) \in \mathbb{R}^{36 \times 1}$ is a column vector with measurements of incident spectrum $\mathbf{X}(n, m) \in \mathbb{R}^{K \times 1}$ at position (n, m) , multiplied by spectral sensitivity matrix $\mathbf{R} \in \mathbb{R}^{36 \times K}$. Because the length of the incident spectrum vector is larger than that of the measurement vector ($36 < K$), the system is underdetermined. Therefore, the reconstruction of the HS image is an ill-posed problem.

3.3 DL-based point-wise spectral reconstruction

In this paper, we adopt DL-based spectral reconstruction networks [32, 33, 34, 35] to reconstruct an HS image. These networks take measurements \mathbf{y} as an input and output the reconstructed spectrum $\hat{\mathbf{X}}$. These CNN-based models learn the mapping from measurements to a target spectrum, enabling the reconstruction of the spectral information. The input \mathbf{y} is divided by its maximum value and fed into the model. The model output reconstructed spectrum signal in normalized form between zero and one, and then the value that was divided into the input is multiplied again to reflect the intensity of the input spectrum. By applying this process to every point (n, m) in the measurement cube, a full HS image can be reconstructed. This process can be accomplished in a single forward operation of the model by exploiting batch processing, making this approach computationally efficient.

3.3.1 Training Dataset

To train the models, we generated a synthetic dataset with 8,000 spectra for training the models. Each spectrum sample, denoted as \mathbf{X} , is synthesized as a combination

of Gaussian distribution functions with varying parameters including peak location, height, and Full Width Half Maximum (FWHM). We used the Gaussian distribution function, as it is generally used to represent spectral signals in nature. The number of Gaussian functions in each combination is determined by sampling from a geometric distribution with the probability parameter of 0.3. To ensure that there is at least one function, we add one to the sampled value. The peak locations are sampled from a uniform distribution of integers from 500 to 750, meaning the spectra are reconstructed at 1 nm intervals within the 500 to 750 nm range. The height of each peak is sampled from a uniform distribution ranging from zero to one and the FWHM of each peak is sampled from a uniform distribution with the interval [2, 50]. Finally, all Gaussian functions are summed and normalized between zero to one thereby generating a spectrum sample. To generate measurements \mathbf{y} , \mathbf{X} is multiplied by spectral sensitivity matrix \mathbf{R} , as described in Eq. (3.4). Then, measurements \mathbf{y} is divided by its maximum value, ensuring the maximum value within input vectors of the models to be one.

3.3.2 Data Augmentation with Uniform Noise Level

In this work, we propose a data augmentation scheme to mitigate the measurement noise in the proposed HS camera. The data augmentation process is conducted as follows.

$$\hat{\mathbf{y}} = \mathbf{y} + \sigma \mathbf{n} \quad (3.5)$$

Where $\hat{\mathbf{y}} \in \mathbb{R}^{36 \times 1}$ is a noisy measurement vector, and $\mathbf{n} \in \mathbb{R}^{36 \times 1}$ is a random noise vector sampled from a normal distribution with zero mean and one variance. The scale of the noise $\sigma \in \mathbb{R}^+$ is determined by the following equation.

$$\sigma = \frac{\sqrt{E[\mathbf{y}^2]}}{10^{\frac{SNR(dB)}{20}}} \quad (3.6)$$

The value of SNR is determined by sampling from a uniform distribution within the range $[a, b]$, which are the design parameters for this method. This process is repeated each time the training samples are drawn from the dataset. Instead of sampling the noise level from a single Gaussian noise distribution, we determine the noise level by sampling from a uniform distribution. This prevents the model from being trained on specific noise levels, improving the reconstruction accuracy.

Chapter 4

Experimentation

For evaluation of the proposed system, we conducted experiments reconstructing the HS image of the real scenes. We compare the reconstruction results using DL-based methods [32, 33, 34, 35] based on the noise level configuration of the data augmentation scheme. Moreover, reconstruction results for both DL-based and a numerical optimization method, Sparse Recovery (SR) [19] are compared. The experimental setup is described in the following paragraphs and summarized in Table 4.1.

4.1 Experimental Setting

Fig. 4.1 shows the experimental setup of the proposed camera for capturing HS images. As targets, we captured the image of a checkerboard (ColorGauge Micro Target, Image Science Associates) and a hand-made clay doll to see the ability to retrieve a spectrum of multiple colors. A commercial spectrometer (Black-Comet, StellarNet) was used to measure the reference spectra of reflected light from the target. A ring-shaped LED light (Advanced Illumination) was used as a light source. We used a variable aperture (SM2D25, Thorlabs) to prevent crosstalk among the adjacent sub-images. Since our target wavelength range begins at 500 nm, we attached a 500 nm long-pass filter on a C mount of the camera module. Fig. 4.2 shows the sub-image array, part of a 4000×3000 sized raw image. Each sub-image captured by microlens passes through the

underlying MTF filter and is modulated according to the transmittance of the filter. By aligning and cropping these sub-images as described in section 3.2, we can obtain a $250 \times 250 \times 36$ measurement cube, from which an HS image with a dimension of $250 \times 250 \times 251$ can be recovered.

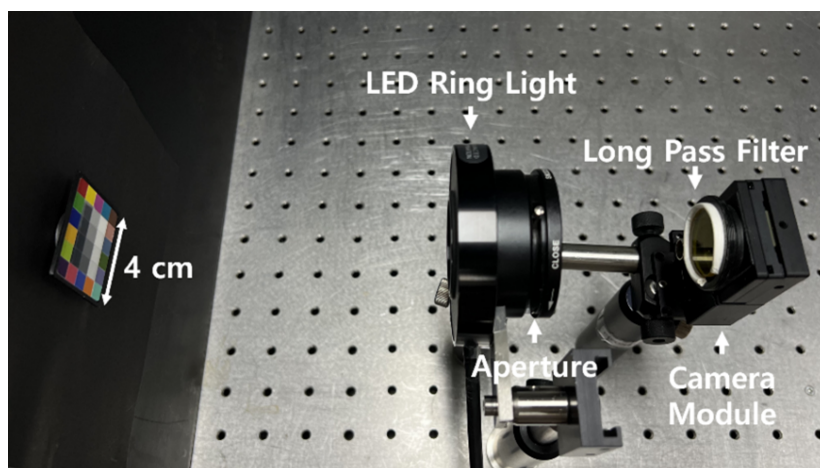


Figure 4.1: Experimental setup for the proposed computational HS camera.

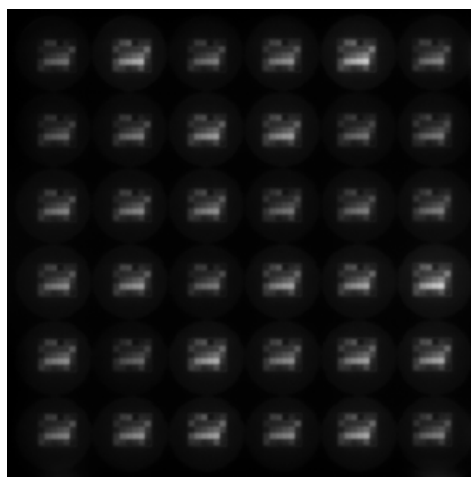


Figure 4.2: Raw image captured by the proposed camera module.

As reconstruction methods, we used DL-based spectral reconstruction networks CNN, ResCNN, U-Net, and ReSpecNN and compared with SR. The DL-based models were trained on the synthetic dataset, minimizing the Mean Square Error (MSE) between the reconstructed spectrum $\hat{\mathbf{X}}$ and target spectrum \mathbf{X} . The following is the equation of MSE.

$$MSE = \frac{1}{N} \|\hat{\mathbf{X}} - \mathbf{X}\|_2^2$$

Where N is the number of wavelengths. We conducted training using an NVIDIA RTX TITAN GPU, Intel i7-11700, and the PyTorch framework. We utilized the Adam [36] optimizer with a fixed learning rate of 0.0001, and a batch size of 1024 and trained the models for 500 epochs. To demonstrate the effectiveness of our data augmentation scheme, we compared the performance of the model under varying noise levels according to Eq. (3.5) and Eq. (3.6) during training. We chose noise levels as follows: no additive noise, fixed noise level (SNR of 15, 20, 25, and 30dB), and uniform noise level (range of [10, 20], [10, 25], [15, 20], and [15, 25] dB). These parameters were determined based on the assumption that the noise level in the proposed system is approximately 20 dB, and they were selected to cover this range. The performance of reconstruction methods was measured by the MSE value between reference spectra and reconstructed spectra. Finally, we compared the reconstruction speed between the methods.

Table 4.1: Summary of experimental setup.

Targets	Checkerboard (ColorGuage Micro Target, Image Science Associates), Clay Doll (handmade)
Commercial spectrometer (reference spectra)	Black-Comet (StellarNet)
Image sensor	IMX 226 (Sony)
Light source	Ring-shaped LED light (Advanced Illumination)
Aperture	SM2D25 (Thorlabs)
Long-pass filter	500 nm \sim
Baseline	SR [19]
DL models (parameters)	CNN (1M) [32], ResCNN (3.5M) [33], U-Net (18.5M) [34], ReSpecNN (4M) [35]
Fixed noise level	15 dB, 20 dB, 25 dB, 30 dB
Uniform noise level	[10 dB, 20 dB], [10 dB, 25 dB], [15 dB, 20 dB], [15 dB, 25 dB]
Training samples	8000 spectra
Optimizer	Adam [36]
Learning rate	0.0001
Batch size	1024
Epochs	500

Chapter 5

Results and Discussion

5.1 Results of Spectral Reconstruction

Table 5.1: Average MSE values of reconstruction results in 18 colors on the checkerboard according to data augmentation configuration.

Model		CNN	ResCNN	U-Net	ReSpecNN
W/O Noise		0.22742	0.15252	0.02924	0.01532
Fixed Noise Level	15 dB	0.01867	0.06079	0.01776	0.0096
	20 dB	0.01284	0.02989	0.01427	0.00994
	25 dB	0.01875	0.0341	0.01493	0.00878
	30 dB	0.02517	0.03482	0.01343	0.0117
Uniform Noise Level	10-20 dB	0.00904	0.03396	0.00823	0.00685
	10-25 dB	0.01138	0.02897	0.0073	0.00878
	15-20 dB	0.01505	0.0327	0.01373	0.00658
	15-25 dB	0.01158	0.03242	0.00983	0.00733

Table 5.1 provides the average MSE values between the reference and reconstructed spectra of 18 colors on the target checkerboard according to the noise level configuration during training. Without data augmentation, all models show high MSE. On the other hand, the fixed noise level configuration resulted in significantly reduced MSE in all models. Specifically, both CNN and ResCNN achieved their best results with data augmentation at a noise level of 20 dB SNR. Additionally, U-Net produced the lowest MSE value at a noise level of 30 dB and ReSpecNN at 25 dB. Finally, configuration with uniform noise level yielded the best results of all models. CNN showed an MSE value of 0.00904 with noise levels ranging from 10 dB to 20 dB. ResCNN and U-Net showed

MSE values of 0.02897 and 0.0073, respectively, with noise levels ranging from 10 dB to 25 dB. ReSpecNN achieved an MSE value of 0.00653 with noise levels ranging from 15 dB to 20 dB, resulting in a 2.3-fold reduction as compared to the configuration without data augmentation. All the best MSE values obtained through data augmentation with a uniform noise level were lower compared to the MSE values achieved under the fixed noise level configuration.

Fig. 5.1 and 5.2 illustrate the reconstruction results of the checkerboard, using SR and DL-based spectral reconstruction networks. DL-based spectral reconstruction models were selected based on the results in Table 5.1, each of which was the best model according to the noise configuration. The colorized images synthesized at 500 nm, 550 nm, and 650 nm wavelengths from reconstructed HS images are presented in Fig. 5.1. DL-based models that provide the best results under uniform noise level data augmentation based on Table 5.1 are used to synthesize color images. Fig. 5.2 provides the comparison between reference and reconstructed spectra at the four points on the target checkerboard (#1, #6, #10, and #15 shown in Fig. 5.1), with each row representing the noise level configuration during training. The reconstruction results of 18 colors are provided in the Appendix. Without any data augmentation, both CNN and ResCNN failed to recover the spectra, whereas U-Net and ReSpecNN achieved slightly better performance. However, both U-Net and ReSpecNN exhibited undesirable peaks and noisy fluctuations in the reconstructed spectra. This failure occurs because the dataset used for training the models does not account for the inherent noise of the system, preventing the model from recovering the input signal from the actual measurements

under the noisy environment. On the other hand, fixed noise level data augmentation resulted in reduced MSE. However noisy fluctuations persisted in reconstructed spectra by CNN, ResCNN, and U-Net. Furthermore, undesirable peaks remained in the spectra reconstructed by ReSpecNN. In contrast, the uniform noise level configuration yielded the most enhanced results. The noisy fluctuations and undesirable peaks were significantly suppressed in the spectra reconstructed by CNN, U-Net, and ReSpecNN, achieving performance comparable to that of SR.

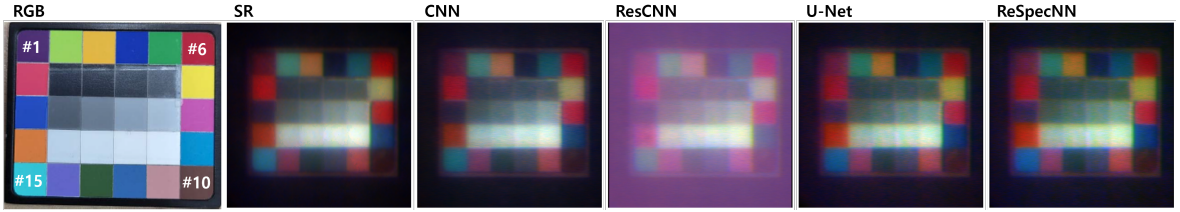


Figure 5.1: Image of the color checkerboard (left) and colored image (rest) of reconstructed HS cube using 500, 550, and 650 nm band images.

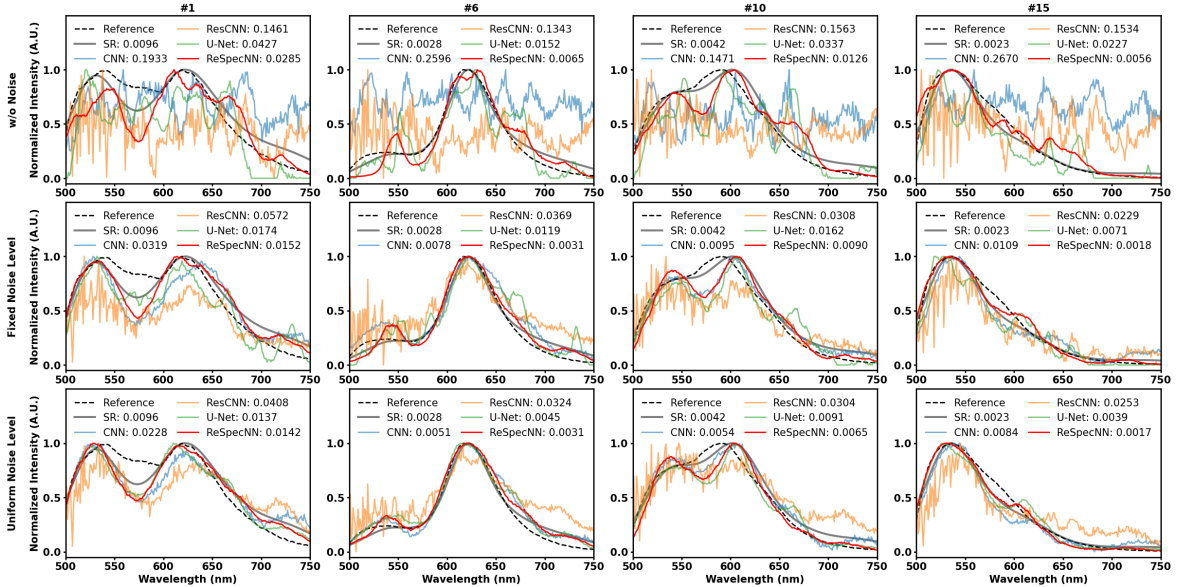


Figure 5.2: Reconstructed spectra of four colors of the checkerboard (#1, #6, #10, and #15 shown in Fig. 5.1). The MSE between the reconstructed and reference spectra is written on the upper side of each graph.

The proposed data augmentation method using sampled noise levels from a uniform distribution showed enhanced spectral reconstruction for DL-based methods. This improvement is attributed to the fact that the models were effectively trained to the inherent noise of the system through data augmentation with uniform noise levels. Without the data augmentation (where σ is zero), the models learn the mapping from measurement \mathbf{y} to input spectrum \mathbf{X} , that is, one-to-one mapping, thereby leading to an inability to recover the input spectra from the unseen noisy measurements. Under the fixed noise level configuration (where σ is constant), the models learn to map from the vectors located at a distance $\sigma\|\mathbf{n}\|_2$ from the measurements \mathbf{y} to input spectrum \mathbf{X} . It is a many-to-one mapping, where the vectors mapped to \mathbf{X} have a ring-shaped line with the radius of $\sigma\|\mathbf{n}\|_2$. Therefore, the model learns the noise distribution and can reconstruct the input signal from the noisy measurements. To improve the ability to recover input spectra under noisy measurements, in the proposed method, the scale of noise σ is determined based on SNR levels sampled from a uniform distribution. As a result, the vectors mapped to \mathbf{X} form a ring-shaped region that covers broader noise distribution compared to the fixed noise level configuration, allowing the models to be trained on the inherent noise of the system effectively.

5.2 Results of Hyperspectral Image Reconstruction

Fig. 5.3, 5.4, and 5.5 show the image clay doll and HS image reconstructed by 5 different methods. DL-based spectral reconstruction models were selected based on the results in Table 5.1, each of which was the best model under uniform noise level

configuration. Fig. 5.3 shows the color image of the clay doll and Fig. 5.4 shows the spectral reconstruction results of target points on the clay doll (#1 and #2 as shown in Fig. 5.3). ResCNN shows noisy fluctuations, whereas the other models successfully reconstructed target spectra. Fig. 5.5 shows the colorized image of the reconstructed HS image (leftmost column) and grayscale band images at 500, 550, 600, 650, 700, and 750nm (the rest). CNN, U-Net, and ReSpecNN show results comparable to those of the SR, whereas the noise remains in images reconstructed by ResCNN. Chromatic aberration can be seen in all colorized images. Since the refractive angle varies depending on the wavelength of light, the angle of incident light entering the filter varies. This aberration distorts the transmittance of the filters, thereby hindering accurate HS imaging.

Table 5.2 shows the average time duration of the HS imaging by 5 different methods. The numerical optimization method, SR shows the slowest speed among CPU times. ReSpecNN demonstrated the lowest reconstruction time on a CPU, achieving approximately a 17-fold reduction. Moreover, reconstruction results of CNN, ResCNN, and ReSpecNN accelerated by GPU show a time duration of less than a second. Furthermore, these networks are relatively lightweight, thereby video-rate HS imaging could be realized by optimized implementation of the models, enabling the application in real-time detection.

Table 5.2: Time duration of HS imaging.

	SR	CNN	ResCNN	U-Net	ReSpecNN
CPU (seconds)	450	87	125	123	27
GPU (seconds)	-	0.44	0.28	5.79	0.18



Figure 5.3: Image of the clay doll

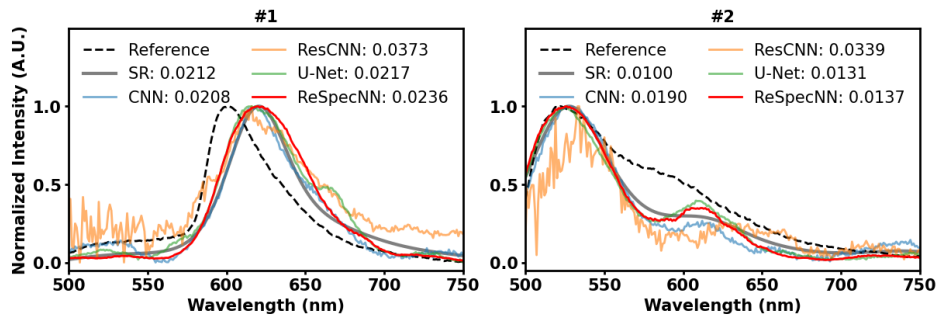


Figure 5.4: Reconstructed spectra at two points of the clay doll (#1 and #2 as shown in 5.3).

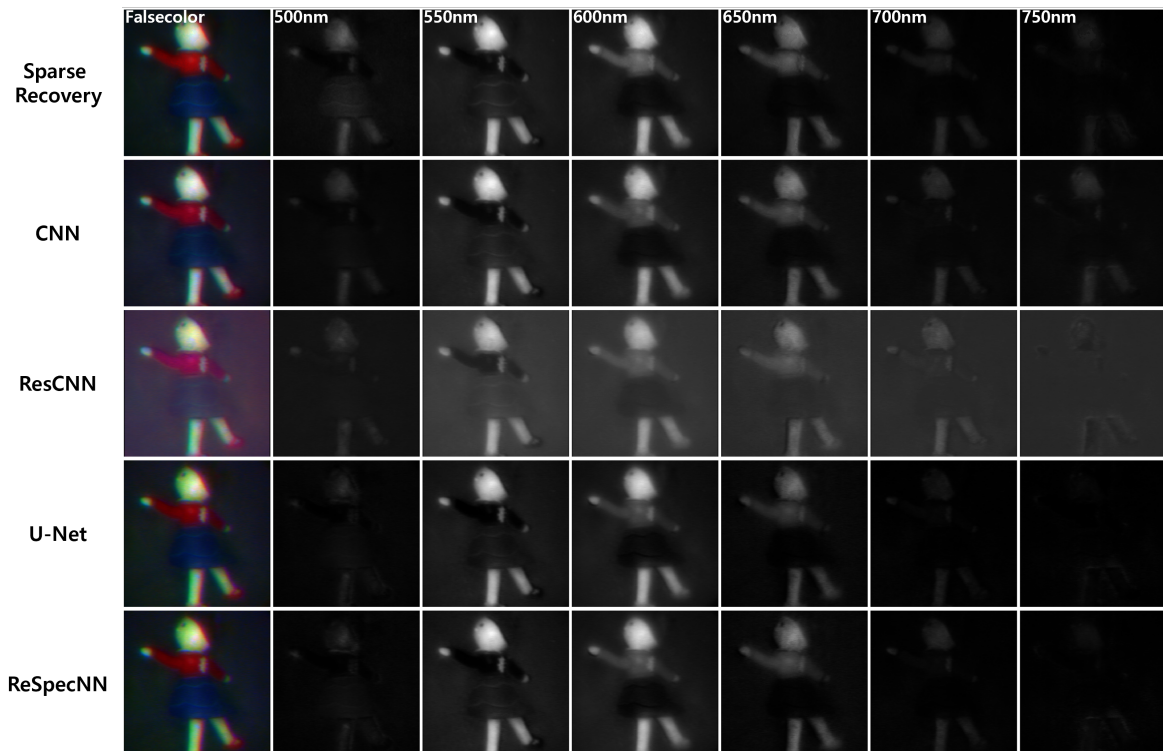


Figure 5.5: Colorized images using 500, 550, and 650 nm band images and grayscale images corresponding to different wavelengths (500, 550, 600, 650, 700, and 750 nm).

Chapter 6

Limitations

Despite the improvements to the proposed system, still there is room for enhancement. First, the spatial resolution is reduced due to inefficient utilization of the CMOS image sensor area by the MLA. Second, an aperture is used to prevent crosstalk between sub-images, which may compromise the system's compactness. Exploring alternatives, such as implementing a signal separator, could address this issue while maintaining compactness. Third, the system exhibits chromatic aberration, which is not considered in the current reconstruction algorithm, leading to a reduction in the quality of the reconstructed HS images. Addressing chromatic aberration within the reconstruction process could significantly enhance image quality.

Chapter 7

Conclusion

In this study, we have developed a DL-based computational HS camera using an MLA and MTF filter array. A total of 36 sub-images could be captured and processed by aligning sub-images using disparity and cropping them into a size of 250×250 pixels, thereby producing a measurement cube with a size of $250 \times 250 \times 36$. The DL-based spectral reconstruction networks are applied to each individual pixel of a measurement cube, producing an HS image with a dimension of $250 \times 250 \times 251$. The developed system improves the reconstruction efficiency while maintaining compact and low-cost hardware configuration by adopting MLA and DL. Our system incorporating MLA can capture an HS image in a single exposure without the requirement of multiple implementations of filter array with a multilayer structure, thereby reducing the complexity of the reconstruction process. By leveraging batch processing, the DL-based spectral reconstruction networks can recover an HS image with a single forward operation, thereby increasing the efficiency of the HS imaging. The proposed data augmentation scheme effectively represents real-world environmental noise. Sampling noise levels from a uniform distribution during training enables DL models to learn the mapping from noisy measurements to the target spectrum, thereby minimizing reconstruction error. The experimental results demonstrated that the proposed method achieves an efficient HS imaging system while maintaining a compact size and low fabrication cost. There-

fore, the proposed HS camera can provide a promising solution for mobile applications such as on-site detection and consumer electronics. In future work, we aim to use the proposed camera to develop a real-time face anti-spoofing system.

References

1. M. Govender, K. Chetty, and H. Bulcock, “A review of hyperspectral remote sensing and its application in vegetation and water resource studies,” *Water Sa*, vol. 33, no. 2, pp. 145–151, 2007.
2. P. M. Mehl, Y.-R. Chen, M. S. Kim, and D. E. Chan, “Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations,” *Journal of food engineering*, vol. 61, no. 1, pp. 67–81, 2004.
3. J. Qin, M. S. Kim, K. Chao, D. E. Chan, S. R. Delwiche, and B.-K. Cho, “Line-scan hyperspectral imaging techniques for food safety and quality applications,” *Applied Sciences*, vol. 7, no. 2, p. 125, 2017.
4. F. Cai, D. Wang, M. Zhu, and S. He, “Pencil-like imaging spectrometer for bio-samples sensing,” *Biomedical optics express*, vol. 8, no. 12, pp. 5427–5436, 2017.
5. U. Khan, S. Paheding, C. P. Elkin, and V. K. Devabhaktuni, “Trends in deep learning for medical hyperspectral image analysis,” *IEEE Access*, vol. 9, pp. 79534–79548, 2021.
6. A. ul Rehman and S. A. Qureshi, “A review of the medical hyperspectral imaging systems and unmixing algorithms’ in biological tissues,” *Photodiagnosis and Photodynamic Therapy*, vol. 33, p. 102165, 2021.
7. C. Song, Y. Hong, J. Lan, H. Zhu, W. Wang, and J. Zhang, “Supervised contrastive learning for snapshot spectral imaging face anti-spoofing,” in *Proceedings of the*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 980–985, 2024.
8. S. Rao, Y. Huang, K. Cui, and Y. Li, “Anti-spoofing face recognition using a metasurface-based snapshot hyperspectral image sensor,” *Optica*, vol. 9, no. 11, p. 1253, 2022.
 9. B. Boldrini, W. Kessler, K. Rebner, and R. W. Kessler, “Hyperspectral imaging: a review of best practice, performance and pitfalls for in-line and on-line applications,” *Journal of near infrared spectroscopy*, vol. 20, no. 5, pp. 483–508, 2012.
 10. M. Zhang, L. Wang, L. Zhang, and H. Huang, “High light efficiency snapshot spectral imaging via spatial multiplexing and spectral mixing,” *Optics Express*, vol. 28, no. 14, pp. 19837–19850, 2020.
 11. F. Huang, R. Cao, P. Lin, B. Zhou, and X. Wu, “High-efficiency multispectral-polarization imaging system using polarization camera array with notch filters,” *IEEE Transactions on Instrumentation and Measurement*, 2023.
 12. F. Huang, Y. Chen, X. Wang, S. Wang, and X. Wu, “Spectral clustering super-resolution imaging based on multispectral camera array,” *IEEE Transactions on Image Processing*, vol. 32, pp. 1257–1271, 2023.
 13. J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka, “Thin observation module by bound optics (tombo): concept and experimental verification,” *Applied optics*, vol. 40, no. 11, pp. 1806–1813, 2001.

14. R. Shogenji, Y. Kitamura, K. Yamada, S. Miyatake, and J. Tanida, “Multispectral imaging using compact compound optics,” *Optics Express*, vol. 12, no. 8, pp. 1643–1655, 2004.
15. S. A. Mathews, “Design and fabrication of a low-cost, multispectral imaging system,” *Applied optics*, vol. 47, no. 28, pp. F71–F76, 2008.
16. Q. Hao, Y. Song, J. Cao, H. Liu, Q. Liu, J. Li, Q. Luo, Y. Cheng, H. Cui, and L. Liu, “The development of snapshot multispectral imaging technology based on artificial compound eyes,” *Electronics*, vol. 12, no. 4, p. 812, 2023.
17. J. Bacca, E. Martinez, and H. Arguello, “Computational spectral imaging: a contemporary overview,” *JOSA A*, vol. 40, no. 4, pp. C115–C125, 2023.
18. D. L. Donoho, “Compressed sensing,” *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
19. C. Kim, P. Ni, K. R. Lee, and H.-N. Lee, “Mass production-enabled computational spectrometers based on multilayer thin films,” *Scientific Reports*, vol. 12, no. 1, p. 4053, 2022.
20. J. M. Bioucas-Dias and M. A. Figueiredo, “A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Transactions on Image processing*, vol. 16, no. 12, pp. 2992–3004, 2007.
21. D. Gedalin, Y. Oiknine, and A. Stern, “Deepcubenet: reconstruction of spectrally

- compressive sensed hyperspectral images with deep neural networks,” *Optics express*, vol. 27, no. 24, pp. 35811–35822, 2019.
22. Y. Zhao, H. Guo, Z. Ma, X. Cao, T. Yue, and X. Hu, “Hyperspectral imaging with random printed mask,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10149–10157, 2019.
23. Z. Yu, D. Liu, L. Cheng, Z. Meng, Z. Zhao, X. Yuan, and K. Xu, “Deep learning enabled reflective coded aperture snapshot spectral imaging,” *Optics Express*, vol. 30, no. 26, pp. 46822–46837, 2022.
24. W. Zhang, H. Song, X. He, L. Huang, X. Zhang, J. Zheng, W. Shen, X. Hao, and X. Liu, “Deeply learned broadband encoding stochastic hyperspectral imaging,” *Light: Science & Applications*, vol. 10, no. 1, p. 108, 2021.
25. M. Yako, Y. Yamaoka, T. Kiyohara, C. Hosokawa, A. Noda, K. Tack, N. Spooren, T. Hirasawa, and A. Ishikawa, “Video-rate hyperspectral camera based on a cmos-compatible random array of fabry–pérot filters,” *Nature Photonics*, vol. 17, no. 3, pp. 218–223, 2023.
26. J. Wen, H. Gao, W. Shi, S. Feng, L. Hao, Y. Liu, L. Xu, Y. Shao, Y. Zhang, and W. Shen, “On-chip real-time hyperspectral imager with full cmos resolution enabled by massively parallel neural network,” *arXiv preprint arXiv:2404.09500*, 2024.
27. M. Hubold, E. Montag, R. Berlich, R. Brunner, and R. Brünig, “Multi-aperture

- system approach for snapshot multispectral imaging applications,” *Optics Express*, vol. 29, no. 5, pp. 7361–7378, 2021.
28. X. Yu, J. Hao, J. Zhou, Y. Su, S. Karim, and Y. Yu, “Modular snapshot multispectral-panchromatic imager (mspi) with customized filter arrays,” *Optics Express*, vol. 31, no. 2, pp. 1475–1485, 2023.
 29. M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, “Single-shot compressive spectral imaging with a dual-disperser architecture,” *Optics express*, vol. 15, no. 21, pp. 14013–14027, 2007.
 30. A. Wagadarikar, R. John, R. Willett, and D. Brady, “Single disperser design for coded aperture snapshot spectral imaging,” *Applied optics*, vol. 47, no. 10, pp. B44–B51, 2008.
 31. J. Oliver, W.-B. Lee, and H.-N. Lee, “Filters with random transmittance for improving resolution in filter-array-based spectrometers,” *Optics Express*, vol. 21, no. 4, pp. 3969–3989, 2013.
 32. C. Kim, D. Park, and H.-N. Lee, “Convolutional neural networks for the reconstruction of spectra in compressive sensing spectrometers,” in *Optical Data Science II*, vol. 10937, pp. 61–66, Spie, 2019.
 33. C. Kim, D. Park, and H.-N. Lee, “Compressive sensing spectroscopy using a residual convolutional neural network,” *Sensors*, vol. 20, no. 3, p. 594, 2020.
 34. C. Kim, D. Park, J. Lee, and H.-N. Lee, “Deep learning-based single-shot computa-

- tional spectrometer using multilayer thin films,” *arXiv preprint arXiv:2204.02669*, 2022.
35. J. Chen, P. Li, Y. Wang, P.-C. Ku, and Q. Qu, “Sim2real in reconstructive spectroscopy: Deep learning with augmented device-informed data simulation,” *arXiv preprint arXiv:2403.12354*, 2024.
36. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

Appendix A

Spectral Reconstruction Results of 18 Colors in the Checkerboard



Figure A.1: Color checkerboard with 18 colors.

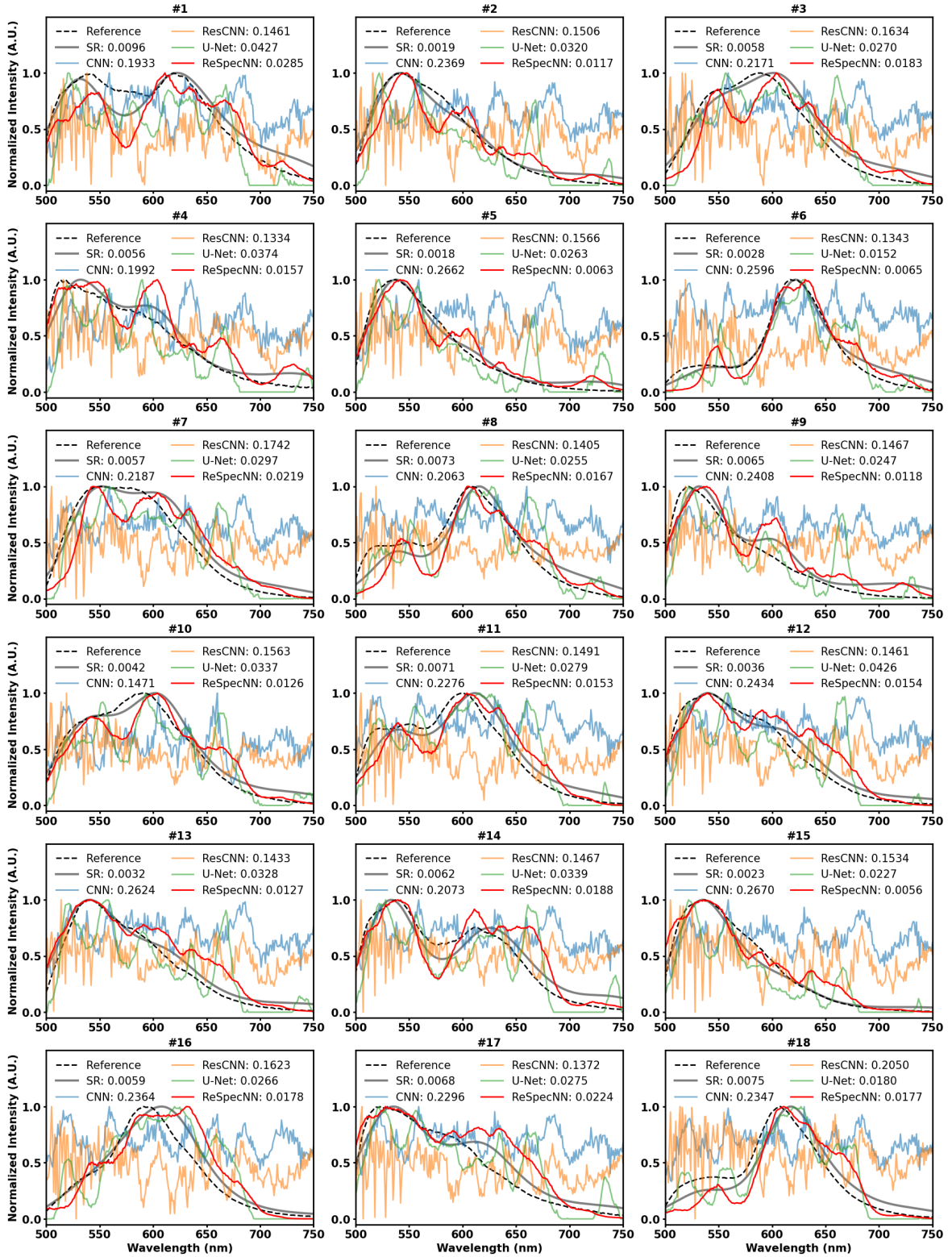


Figure A.2: Reconstruction results without data augmentation.

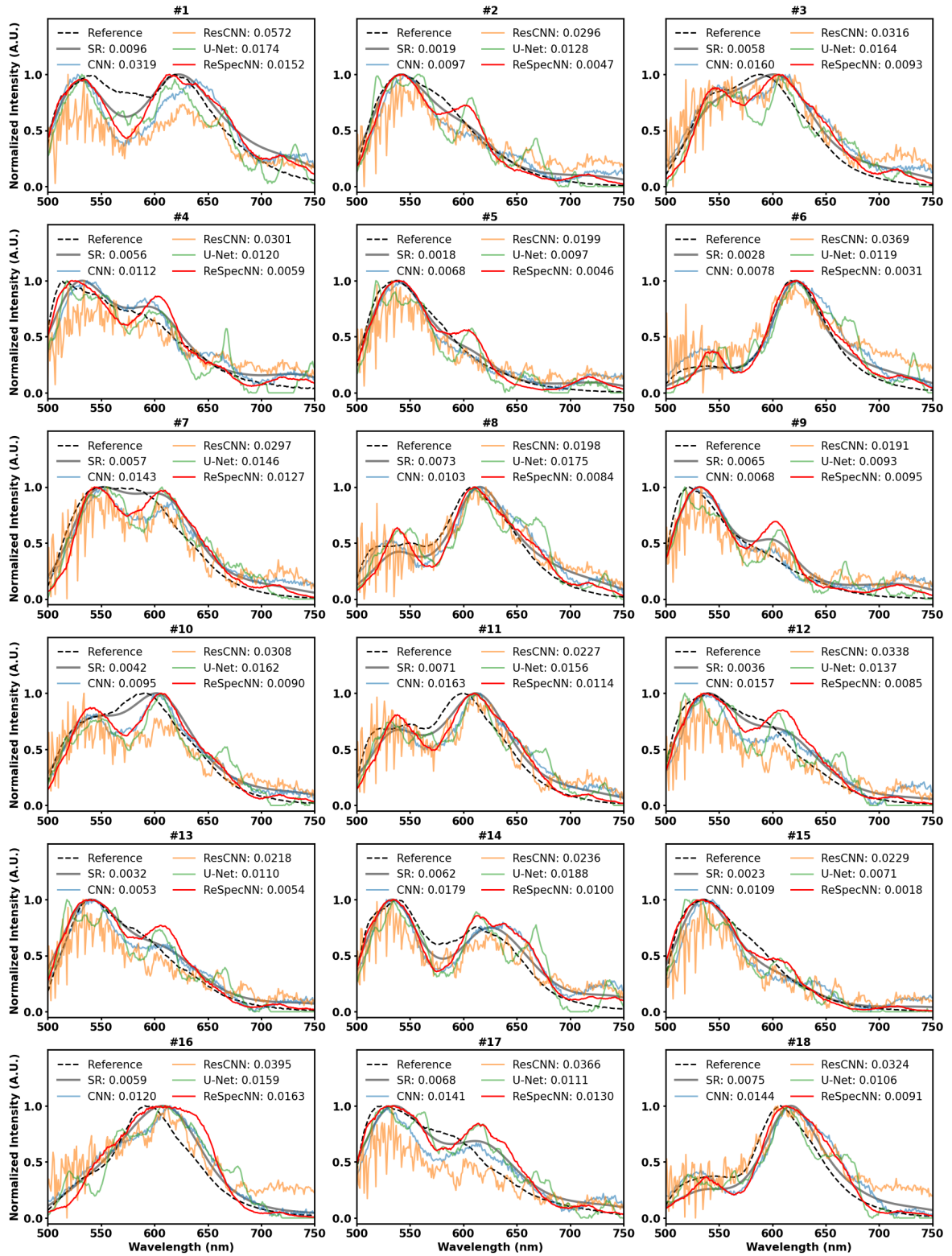


Figure A.3: Reconstruction results with fixed noise level data augmentation.

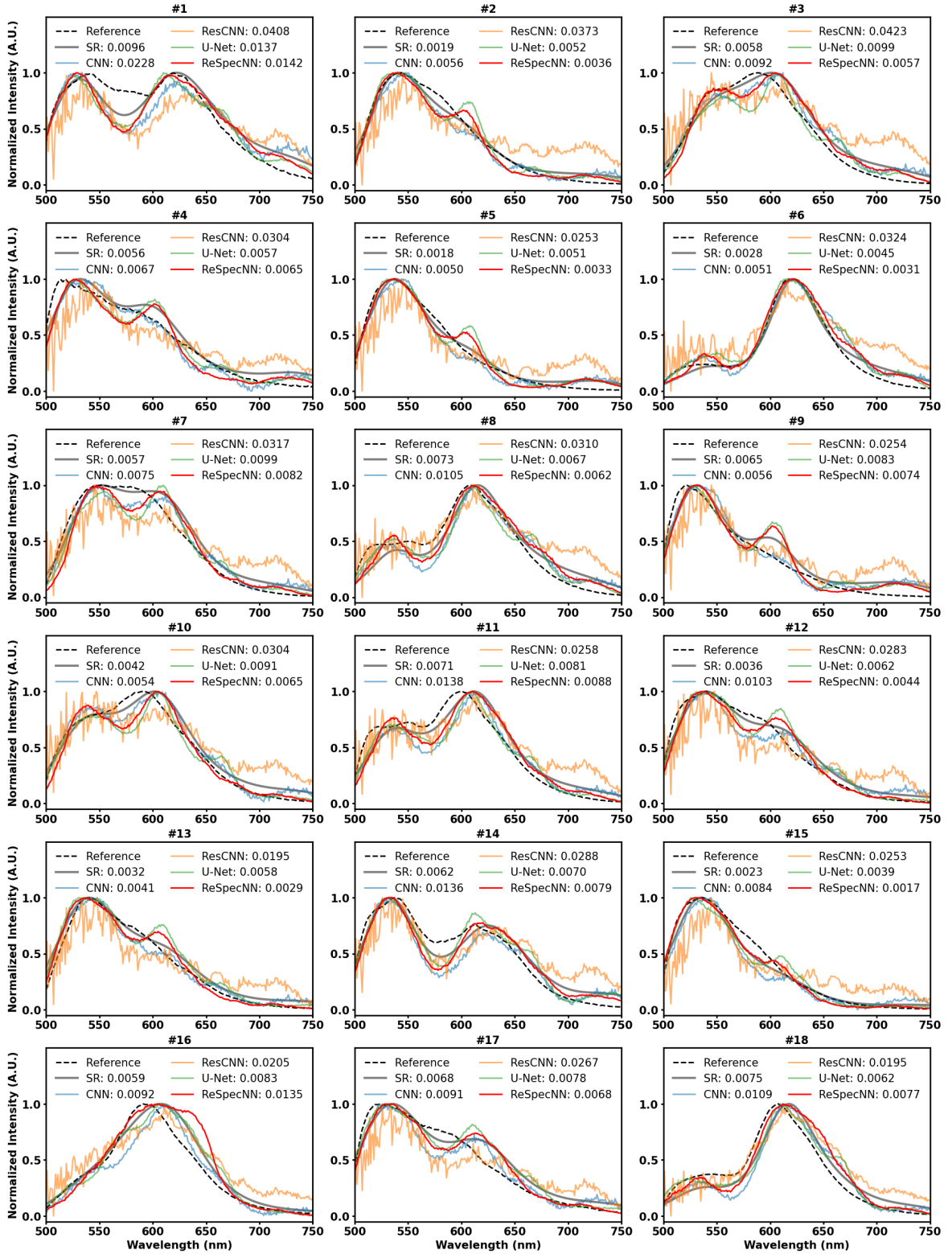


Figure A.4: Reconstruction results with uniform noise level data augmentation.

Acknowledgements

막연한 기대로 연구자의 길을 걸어 보고자 GIST에 입학하였습니다. 박사 학위를 목표로 통합과정으로 입학하였지만, 많은 고민 끝에 다른 결정을 하였습니다. 선택을 달리 하여도, 연구실에서 보낸 수 많은 밤이 앞으로의 길에 양분이 될 것이라 기대하면서 석사로 학위를 마치고 졸업합니다.

제가 사랑하고 존경하는 부모님 덕분에 연구라는 도전을 해 볼 수 있었던것 같습니다. 부족한 아들이지만 부모님의 응원 덕에 여기까지 올 수 있었습니다. 본가에 돌아 갈 때 마다 격하게 반겨주는 귀여운 내 동생들 지민이, 지한이에게 고마운 마음이 듭니다. 멋진 형, 오빠가 되고자 하는 마음이 항상 저의 원동력이 되고 있습니다.

연구자의 길 가운데 지도해주신 이흥노 교수님께도 감사드립니다. 교수님이 주신 가르쳐 주신 "읽고", "생각하고", "쓰고" 의 3가지가 연구의 기본임을 깨닫게 되었습니다. 고집불통의 말 안듣는 MZ세대다운 학생이지만, 끝까지 지도해 주셨습니다. 귀중한 시간 내어 학위 논문 심사를 맡아주신 윤훈한 교수님, 이규빈 교수님께도 감사드립니다.

연구의 처음부터 끝까지 모두 챙겨주신 김철순 박사님께도 감사드립니다. 졸업하신지 3년이 다 되어 가지만, 회사 일로 바빠신 와중에도 못난 후배에게 해주신 조언이 어려움을 맞닥뜨릴 때 마다 큰 도움이 되었습니다. I also appreciate Dr. David Samuel Bhatti for every advice, discussion, and teaching. It was great time to work with you.

학위 과정 동안 함께 생활한 INFONET 구성원들에게 감사드립니다. 광주에서 생활했던 3년간 여러분들과 함께하면서 많은 것을 경험했기에 오늘 날의 제가 있다고 생각합니다. 소중한 가르침, 잊지 않고 앞으로의 일을 해 나가려고 합니다. 다시한번 감사드립니다.

