Real-time Hyperspectral Imaging in Hardware via Trained Metasurface Encoders

Maksim Makarenko[1†], Arturo Burguete-Lopez[1], Qizhou Wang[1], Fedor Getman[1], Silvio Giancola[1],
Bernard Ghanem[1], Andrea Fratalocchi[1]
[1]King Abdullah University of Science and Technology (KAUST)
Thuwal, 23955-6900, KSA
[†]maksim.makarenko@kaust.edu.sa

Abstract

Hyperspectral imaging has attracted significant attention to identify spectral signatures for image classification and automated pattern recognition in computer vision. State-of-the-art implementations of snapshot hyperspectral imaging rely on bulky, non-integrated, and expensive optical elements, including lenses, spectrometers, and filters. These macroscopic components do not allow fast data processing for, e.g. *real-time and high-resolution videos. This work introduces Hyplex™, a new integrated architecture addressing the limitations discussed above. Hyplex™ is a CMOS-compatible, fast hyperspectral camera that replaces bulk optics with nanoscale metasurfaces inversely designed through artificial intelligence. Hyplex™ does not require spectrometers but makes use of conventional monochrome cameras, opening up the possibility for real-time and high-resolution hyperspectral imaging at inexpensive costs. Hyplex™ exploits a model-driven optimization, which connects the physical metasurfaces layer with modern visual computing approaches based on end-to-end training. We design and implement a prototype version of Hyplex™ and compare its performance against the state-of-the-art for typical imaging tasks such as spectral reconstruction and semantic segmentation. In all benchmarks, Hyplex™ reports the smallest reconstruction error. We additionally present what is, to the best of our knowledge, the largest publicly available labeled hyperspectral dataset for semantic segmentation.* [1]

1. Introduction

Hyperspectral imaging is gaining considerable interest in many areas including civil, environmental, aerial, military, and biological sciences for estimating spectral features that allow the identification and remote sensing of complex materials [10, 29]. Ground-based hyperspectral imaging enables automated classification for food inspection, surgery, biology, dental and medical diagnosis [1, 21, 34, 42]. Likewise, aerial and submarine hyperspectral imaging are currently opening new frontiers in agriculture and marine biology for the taxonomic classification of fauna, and through aerial drone footage for precision agriculture [2, 10, 13, 14]. The present state-of-the-art in hyperspectral imaging, however, is still affected by problems of expensive setup costs, time-consuming post-data processing, low speed of data acquisition, and the needs of macroscopic optical and mechanical components [41, 58]. A single hyperspectral image obtained from a high-resolution camera typically requires gigabytes of storage space, making it impossible to perform real-time

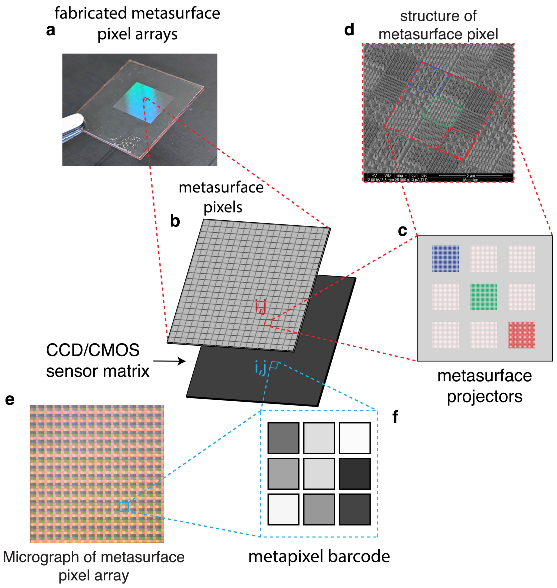

a
fabricated metasurface pixel arrays
d
structure of metasurface pixel
metasurface pixels
b
c
i,j
CCD/CMOS sensor matrix
i,j
metasurface projectors
e
f
Micrograph of metasurface pixel array
metapixel barcode

Figure 1. **Hardware implemented Hyplex™ imaging system.** (a) Example of metasurface pixel arrays (blue squares). (b) Schematic of meta-pixel array on top of a camera sensor. (c) Closeup showing the metasurface projectors as subpixels of the array. (d) Scanning electron microscope image of a fabricated metasurface pixel. (e) Optical micrograph of the metasurface pixel array. (f) Illustration of the barcode generated by (e).

[1] Dataset available on https://github.com/makamoa/hyplex.

video analysis with today's computer vision techniques [28].

Computational hyperspectral reconstruction from a single RGB image is a promising technique to overcome some of the challenges mentioned above [4, 7, 18, 22, 25, 26, 38, 44, 54, 63]. Heidrich *et al*. [24] proposed hyperspectral cameras based on integrated diffractive optical elements, while other groups [12, 60] leveraged deep neural networks for designing spectral reconstruction filters. While these approaches could help address the problem of speed, they are not yet able to tackle the issues of high cost and slow data processing. Other bottlenecks are the use of elementary filter responses, which are not optimized beyond primitive thin-film interference patterns, and the lack of integrated structures that could exploit the modern footprint of CCD/CMOS sensors.

We here introduce the Hyplex™ system (Fig. 1), a data-driven hyperspectral imaging camera (Fig. 1, a-b), which uses state-of-the-art metasurfaces to replace macroscopic components with highly integrated dielectric nanoresonators that manipulate light as a feed-forward neural network [9, 17, 19]. Metasurfaces have successfully demonstrated the ability to integrate various basic optical components for different applications [48–50]. Hyplex™ leverages this technology to compress high-dimensional spectral data into a low-dimensional space via suitably defined projectors (Fig. 1, c-d), designed with end-to-end learning of large hyperspectral datasets. ALFRED [16, 19, 37], an open-source, inverse-design software exploiting artificial intelligence (AI), provides the means to design the metasurface projectors. These nanostructures encode broadband information carried by incoming spectra into a barcode composed of a discrete pattern of intensity signals (Fig. 1, e-f). A physical model-aware framework finds the optimal projectors' response with various learning schemes, designed based on user end tasks.

We summarize our contribution as follows: **(i)** We propose and implement an inexpensive and fast-processing data-driven snapshot hyperspectral camera that uses two integrated components: inverse-designed spectral encoders and a monochrome camera. **(ii)** We implement an end-to-end framework for hyperspectral semantic image segmentation and spectral reconstruction, and benchmark it against the state-of-the-art, reporting the highest performance to date. **(iii)** We create FVgNET, the largest publicly available dataset of 317 samples of labeled hyperspectral images for semantic segmentation and classification.

2. Related Work

Hyperspectral reconstruction is an ill-posed problem demanding the inverse projection from low-dimensional RGB images to densely sampled hyperspectral images (HSI) [5, 39]. Metamerism [15], which projects different spectral distributions to similar activation levels of visual sensors, represents a significant challenge. Traditional RGB cameras project the entire visible spectra into only three primary colors. This process eliminates critical information making it challenging to distinguish different objects [38]. For the specific task of hyperspectral reconstruction, we can partially recover such lost information. Spectral projections are similar to autoencoders in the sense that they downsample the input to a low-dimensional space. If we design a suitable algorithm that explores this space efficiently, we could retrieve sufficient data to reconstruct the initial input.

Reconstruction by sparse coding and deep learning: Sparse coding [32, 45] represents perhaps the most intuitive approach to this idea. These methods statically discover a set of basis vectors from HSI datasets known *a priori*. Arad *et al*. [5] implemented the K-SVD algorithm to create overcomplete HSI and RGB dictionaries. The HSI is reconstructed by decomposing the input image into a linear combination of basis vectors, then transferred into the hyperspectral dictionary. A limit of sparse-coding methods is their applied matrix decomposition algorithms, which are vulnerable to outliers and show degraded performance [27]. Recently, research groups extended the capabilities of sparse coding by investigating deep learning. Galliani *et al*. [18] demonstrated a supervised learning method, where a UNet-like architecture [46] is trained to predict HSI out of single RGB images. Nguyen [38] trained a radial basis function network to translate white-balanced RGB values to reflection spectra. In another work, Xiong *et al*. [55] introduced a 2-stage reconstruction approach comprising an interpolation-based upsampling method on RGB images. The end-to-end training proposed recovers true HSI from the upsampled images. Wug *et al*. [40] used different RGB cameras to acquire non-overlapping spectral information to reconstruct the HSI. These approaches reconstruct spectral information from highly non-linear prediction models, limited by their supervised learning structure. The models constrain data downsampling to non-optimal RGB images by applying a color generation function on HSI or generic RGB cameras. With Hyplex™, we avoid all the issues of the sparse coding and deep-learning reconstruction methods by exploring a new concept, which performs spectral downsampling with optimally designed metasurface projectors.

Hyperspectral imaging with trainable projectors: Optical projectors in cameras mimic the chromatic vision of humans based on primary colors [23]. In hyperspectral imaging, however, the design of projectors requires further study to identify their optimal number and response. Human eyes are not the best imaging apparatus for every possible real-world scenario. The works of [6, 47, 53] expand the concept of RGB cameras to arbitrary low-dimensional sampling of reflectance spectra. These works employ different variants of optimization routines, which converge to a set of optimal projectors from an initial number of candidates. The selected projectors provide a three-channel reconstruction of the HSI with superior performance. Nie *et al*. [39] demonstrated that a 1×1 convolution operation achieves similar functionality to optical projectors while processing multi-spectral data frames. The network is like an autoencoder, where the input HSI is downsampled and then reconstructed by a decoder network. Zhang *et al.* [62] designed and fabricated a broadband encoding stochastic camera containing 16 trainable projectors that map high-dimensional spectra to lower-dimensional intensity matrices. Recently, Liutao *et al.* [56] proposed FS-Net, a filter-selection network for task-specific hyperspectral image analysis. In [59], the authors showcased an idea of filter optimization for hyperspectral-informed image segmentation tasks.

Inverse design of metasurface projectors: Optimizing best-fit filters is a dimensionality reduction problem, which requires finding the principal component directions that encode eigenvectors showing the lowest loss. The state-of-the-art results are generated either from theoretical calculation or experimental measurement on thin-film filters, representing a rough approximation of the precise principal components. In hyperspectral imaging, these components typically exhibit frequency-dependent irregular patterns composed of complex distributions of sharp and broad resonances, indicating the need for more dedicated control of material structures, *e.g.* metasurface technology. Modern metasurface design approaches [51,52] usually rely on a library of pre-computed metasurface responses and polynomial fitting to further generalize the relationship between design parameters and the device performance. We, instead, design our metasurface optical projectors via ALFRED [20], a hybrid inverse design scheme that combines classical optimization and deep learning [52]. In this work, we significantly extend the capabilities of the original code by adding differentiability, physical-model regularization, and complex decoder projectors able to tackle different computer vision tasks and perform thousands of parameter optimizations through the supervised end-to-end learning process.

3. Methodology

The Hyplex™ hyperspectral imaging system consists of two parts: a hardware linear spectral encoder $\mathcal{E}$ and a software decoder (Fig. 2). The encoder compresses an input high-dimensional HSI $\boldsymbol{\beta}$ to a lower multispectral image tensor $\hat{S} = \mathcal{E}(\beta)$, while the decoder maps the tensor $\hat{S}$ to user-defined task-specific outputs. In this work, we consider two types of tasks: hyperspectral reconstruction and semantic segmentation. Spectral reconstruction aims to reconstruct with minimum losses the input HSI tensor. We define the loss via the Root Mean Squared Error (RMSE) $\tilde{\beta} = \mathcal{D}_{rec}(\mathcal{E}(\beta))$ between reconstructed and input spectra. Semantic segmentation, conversely, provides a pixel-by-pixel classification of HSI. In this task, we use as decoder $\mathcal{D}_{seg}$ the U-Net architecture, with adjusted input and output layers to meet the dimensionality of the HSI tensor. The decoder outputs softmax logits $\hat{y}$, representing the probability of observing each pixel ground-truth label y. We assess these predictions quantitatively by using the Cross-Entropy loss function $\mathcal{L}_{seg}$.

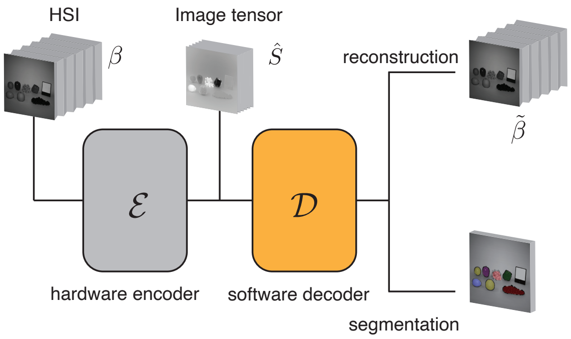


Figure 2. **Conceptual sketch of Hyplex™ system.** The system is constructed by a hardware optical encoder $\mathcal{E}$ that is implemented via trainable metasurface arrays and a software decoder $\mathcal{D}$ optimized for two different tasks, including hyperspectral reconstruction and spectral-informed semantic segmentation.

3.1. Hardware encoder

Recent work demonstrates that the transfer function of an array of sub-micron nanostructured geometries can approximate arbitrarily defined continuous functions [19, 36]. In Hyplex™, we exploit such universal approximation ability to design and implement an optimal linear spectral encoder hardware for a specific hyperspectral information-related imaging task.

Figure 3 summarizes the data workflow of Hyplex™ for a generic linear encoder operator $\mathcal{E} = \hat{\mathbf{\Lambda}}^{\dagger}$. Panel (a) shows an example hyperspectral image. The data is represented as a tensor $\boldsymbol{\beta}$ with three dimensions: two spatial dimensions (x, y), corresponding to the camera virtual image plane, and one frequency axis ω, measuring the power density spectra retrieved at one camera pixel (Fig. 3b). Following a data-driven approach, we implement a linear dimensionality reduction operator that finds a new equivalent encoded representation of $\boldsymbol{\beta}$ (Fig. 3c). The hyperspectral tensor of a dataset of images is flattened to a matrix $\mathbf{B}$ that contains, on each column, the power density spectra of a set of camera pixels. We then the apply the linear encoding $\mathbf{\Lambda}^{\dagger}$ to obtain an approximation of $\mathbf{B}$ [8] via a set of linear projectors $\mathbf{\Lambda}(\omega)$, which map pixel-by-pixel the spectral coordinate β_{ij} to a set of scalar coefficients S_{ijk}:

$$S_{ij} = \tilde{\mathbf{\Lambda}}(\omega)\beta ij(\omega), \quad S_{ijk} = \int \Lambda_k(\omega)\beta_{ij}(\omega)\,\mathrm{d}\omega\,. \quad (1)$$

The spectral information contained in $\beta_{ij}(\omega)$ is embedded into an equivalent barcode S_{ijk} of a few components.

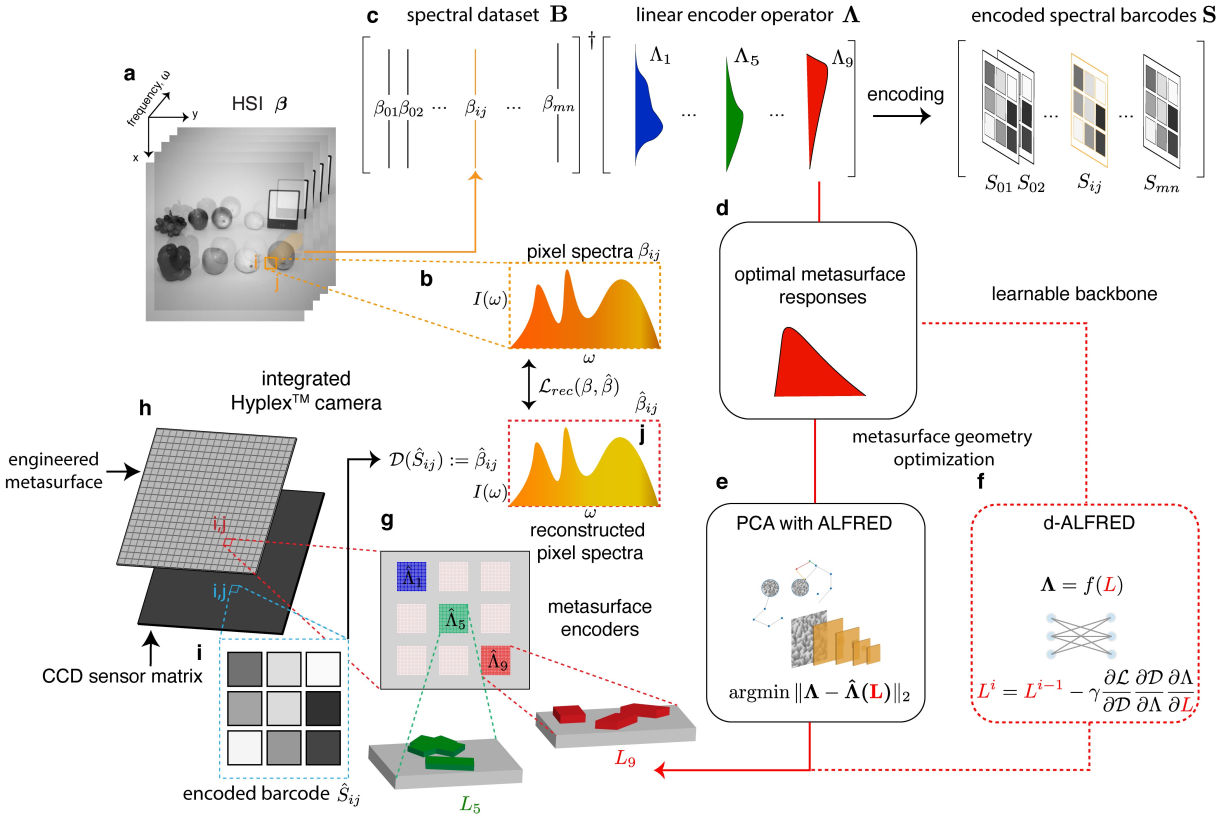

a
frequency, ω
y
x
HSI β
c
spectral dataset $\mathbf{B}$
linear encoder operator $\mathbf{\Lambda}$
encoded spectral barcodes $\mathbf{S}$
$\beta_{01}\beta_{02}$ ··· β_{ij} ··· β_{mn}
†
Λ_1 ··· Λ_5 ··· Λ_9
encoding
S_{01} S_{02} ··· S_{ij} ··· S_{mn}
d
optimal metasurface responses
learnable backbone
pixel spectra β_{ij}
b
$I(\omega)$
ω
$\mathcal{L}_{rec}(\beta, \hat{\beta})$
$\hat{\beta}_{ij}$
integrated Hyplex™ camera
h
engineered metasurface
$\mathcal{D}(\hat{S}_{ij}) := \hat{\beta}_{ij}$
j
$I(\omega)$
ω
reconstructed pixel spectra
metasurface geometry optimization
e
f
PCA with ALFRED
d-ALFRED
$\mathbf{\Lambda} = f(L)$
i,j
i,j
g
$\hat{\Lambda}_1$
$\hat{\Lambda}_5$
$\hat{\Lambda}_9$
metasurface encoders
CCD sensor matrix
i
argmin $\|\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}(\mathbf{L})\|_2$
$L^i = L^{i-1} - \gamma \frac{\partial \mathcal{L}}{\partial \mathcal{D}} \frac{\partial \mathcal{D}}{\partial \Lambda} \frac{\partial \Lambda}{\partial L}$
L_9
L_5
encoded barcode $\hat{S}_{ij}$

Figure 3. **Metasurface subpixel array as a linear spectral encoder.** (a) A spectral image tensor ($\boldsymbol{\beta}$) is captured by a hyperspectral camera. (b) The corresponding pixel spectra ($\hat{\beta_{ij}}$) at position ij in the xy camera plane. (c) Example of dimensional reduction linear operator $\mathbf{\Lambda}^\dagger$ of a flattened matrix $\mathbf{B}$ with the resulted projected encoder barcode for a pixel spectra at the ij position. (d) Optimal encoder functions $\mathbf{\Lambda}^\dagger$. (e) Non-differentiable inverse design optimization framework implemented via ALFRED utilized to find a set of metasurfaces $\mathbf{L}$ with desired response Λ_i. (f) Differentiable backbone enabling simultaneous optimization of responses $\mathbf{\Lambda}$ and structures $\mathbf{L}$. Metasurface pixel (g) composed of a two-dimensional array of resonant metapixels with corresponding fitted transmission responses $\hat{\mathbf{\Lambda}}$. (h) Conceptual sketch of the Hyplex™ system with an enlarged spectral-specific barcode (i) produced by an imaging-based readout of the metasurface's transmission response.(j) Recovered pixel spectra through decoder $\mathcal{D}_{rec}$ projection $\hat{\beta}_{ij}$.

To implement the $\mathbf{\Lambda}$ encoder projectors into hardware, Hyplex™ uses two different engineering lines (Fig. 3e-f). When the user end task does not require additional constraints, such as in, *e.g.* spectral reconstruction, Hyplex™ implements the projector by utilizing optimization frameworks to minimize the norm between the physical metasurface response $\hat{\mathbf{\Lambda}}$ and the target $\mathbf{\Lambda}$ (Fig. 3e). Conversely, in tasks that require further conditions such as, *e.g.* hyperspectral semantic segmentation, Hyplex™ uses a learnable backbone (Fig. 3f). This optimization exploits d-ALFRED, a new version of ALFRED that creates a differentiable physical model that is trained with an end-to-end approach. d-ALFRED designs metasurface geometries with an iterative process that minimizes the loss function $\mathcal{L}_{seg}$ by optimizing simultaneously the projector responses $\mathbf{\Lambda}$ and the vector $\mathbf{L}$ containing all the parameters defining the metasurface. A single Hyplex™ pixel (Fig. 3g) integrates various metasurface projectors in a two-dimensional array of sub-pixels, which are replicated in space to form the Hyplex™ hardware encoder (Fig. 3h). The encoder transforms a reflection spectra arising from a scene into a barcode $\hat{S}_{ij}$ (Fig. 3i), composed of a set of intensity signals proportional to the overlap between the input spectra and each projector's response as defined in Eq. (1). A standard monochromatic camera, placed behind the metasurfaces, acts as an imaging readout layer. Each pixel of the camera matches the sub-pixel of the hardware encoder and retrieves one intensity signal of the barcode $\hat{S}_{ij}$ (Fig. 3j).

PCA projectors engineered with ALFRED: We use a linear encoder $\mathbf{\Lambda}$ obtained through an unsupervised learning technique via principal component analysis (PCA). The PCA performs hardware encoding $\mathcal{E}$ by selecting the k strongest ($k = 9$ for this work) principal components $\tilde{\mathbf{\Lambda}}^\dagger$ from the singular value decomposition of $\mathbf{B} = \mathbf{\Lambda\Sigma V}^\dagger$ [8], and approximating $\mathbf{B}$ as follows:

$$\mathbf{B} \approx \tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Sigma}}\tilde{\boldsymbol{V}}^\dagger \tag{2}$$

Equation (2) offers the closest linear approximation of $\mathbf{B}$ in least square sense. We implement the decoder $\mathcal{D}$ with the linear projector $\hat{\beta}_{ij} = \tilde{\mathbf{\Lambda}}\hat{S}_{ij}$, which recovers the best least square approximation of the pixel spectra $\hat{\beta}_{ij}(\omega) \approx \beta_{ij}(\omega)$ (Fig. 3j) from the selected PCA component.

3.2. Learnable backbone via differentiable physical model

In this approach, we represent the decoder operator $\mathcal{D}$ as a set of hierarchical nonlinear operators $\mathcal{F}$, which project the input tensor $\hat{S}$ into an output measurement tensor $\hat{y}$. This process is iteratively trained via supervised learning, comparing the measurement $\hat{y}$ with some ground-truth tensor $\tilde{y}$. This end-to-end training finds the optimal feature space $\hat{S}$ and the associated linear projectors $\mathbf{\Lambda}$. To train Hyplex™ in this framework with backpropagation, the encoder $\mathcal{E}$ needs to be differentiable.

In the inverse-design of projectors, the encoder $\mathcal{E} = \mathbf{H}$, with $\mathbf{H}(\omega)$ representing the output transmission function of the metasurface response, which is obtained from the solution of the following set of coupled-mode equations [36]:

$$\begin{cases} \tilde{\mathbf{a}}(\omega) = \frac{\tilde{K}}{i(\omega - W) + \frac{\tilde{K}\tilde{K}^\dagger}{2}}\tilde{\mathbf{s}}_+ \\ \tilde{\mathbf{s}}_-(\omega) = \tilde{C}(\omega)\cdot\left(\tilde{\mathbf{s}}_+ - \tilde{K}^\dagger\cdot\tilde{\mathbf{a}}\right) \end{cases} \tag{3}$$

where W is a diagonal matrix with resonant frequencies ω_n of the modes $W_{nn} = \omega_n$, $\tilde{C}(\omega)$ is a scattering matrix modeling the scattering of impinging waves $\tilde{\mathbf{s}}_+$ on the resonator space, and $\tilde{K}$ is a coupling matrix representing the interaction between traveling waves $\tilde{\mathbf{s}}_\pm(t)$ and resonator modes $\tilde{\mathbf{a}}(t)$. Equations (3) describe the dynamics of a network of resonator modes $\tilde{\mathbf{a}} = [\tilde{a}_1(\omega), \dots, \tilde{a}_n(\omega)]$, interacting with $\tilde{\mathbf{s}_\pm} = [\tilde{s}_{1\pm}(\omega), \dots, \tilde{s}_{m\pm}(\omega)]$ incoming $(+)$ and reflected $(-)$ waves. Section 1 of the Supplementary Material provides more details on the quantities appearing in Eq. (3).

The input-output transfer function $\mathbf{H} = \tilde{\mathbf{s}}_-/\tilde{\mathbf{s}}_+$ resulting from the solution of Eq. (3) is the superposition of two main terms: a propagation term defined by the scattering matrix $\tilde{C}(\omega)$ and a nonlinear term containing the rational function $\frac{\tilde{K}}{\sigma(\omega - W)}$. Equation (3) represents a differentiable function of W through which it is possible to backpropagate (Fig. 4 b). **d-ALFRED:** To project the resonator quantities in Eq. (3) to metasurface input parameters $\mathbf{L}$, we use a supervised optimization process. We train a deep neural network to learn the relationship between $\mathbf{L}$ and the resonator variables in Eq. (3). Following the same approach of [37], we train the network with a supervised spectral prediction task by using arrays of silicon boxes with simulated transmission/reflection responses (see Sec. 2 of Supplementary Material).

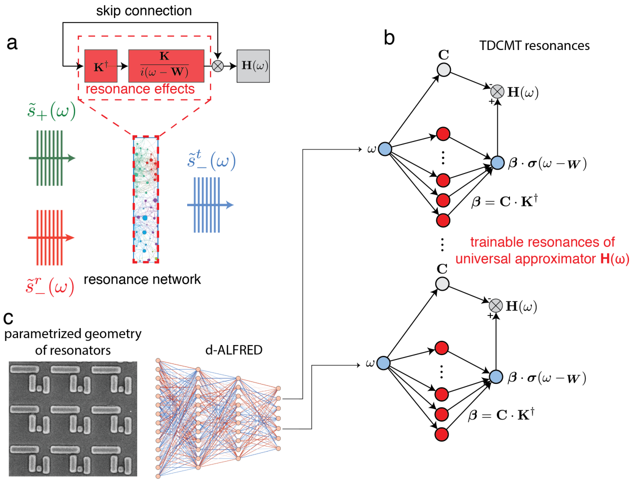

a
skip connection
$\mathbf{K}^\dagger$
$\frac{\mathbf{K}}{i(\omega - \mathbf{W})}$
$\mathbf{H}(\omega)$
resonance effects
$\tilde{s}_+(\omega)$
$\tilde{s}^t_-(\omega)$
$\tilde{s}^r_-(\omega)$
resonance network
b
TDCMT resonances
C
$\mathbf{H}(\omega)$
ω
$\boldsymbol{\beta}\cdot\boldsymbol{\sigma}(\omega - \boldsymbol{W})$
$\boldsymbol{\beta} = \mathbf{C}\cdot\mathbf{K}^\dagger$
trainable resonances of
universal approximator H(ω)
C
$\mathbf{H}(\omega)$
ω
$\boldsymbol{\beta}\cdot\boldsymbol{\sigma}(\omega - \boldsymbol{W})$
$\boldsymbol{\beta} = \mathbf{C}\cdot\mathbf{K}^\dagger$
c
parametrized geometry of resonators
d-ALFRED

Figure 4. **Coupled mode network as a differentiable metasurface physical model.** (a) Coupled-mode photonic network as a feedback-loop with skip connection. (b) trainable coupled resonance layer. (c) d-ALFRED: trained differentiable projections from parametric geometry shapes to resonances.

4. Datasets

To train and validate the Hyplex™ system, we use three publicly available datasets: the CAVE dataset, consisting of 32 indoor images covering 400 nm to 700 nm, and the Harvard and KAUST sets, which contain both indoor and outdoor scenes, and amount to 75 and 409 images, respectively, with spectral bands covering 420 nm to 720 nm and 400 nm to 700 nm respectively. We create an additional hyperspectral dataset FVgNET. FVgNET is comprised of 317 scenes showing fruits and vegetables, both natural and artificial, taken indoors under controlled lighting conditions, and covering the 400 nm to 1000 nm range. We acquired the images using a setup consisting of a white paper sheet arranged in an infinity curve, a configuration employed in photography to isolate objects from the background. We achieve good spectral coverage while minimizing the presence of shadows in the final images by illuminating the objects with overhead white LED indoor lighting, a 150 W halogen lamp (OSL2 from Thorlabs) equipped with a glass diffuser and a 100 W tungsten bulb mounted in a diffuse reflector.

Figure 5a-b shows the distribution of object classes in the dataset. For each class of objects (*e.g.*, apple, orange, pepper), we generated an approximately equal number of scenes showing: natural objects only and artificial objects only. The dataset consists of 12 classes, represented in the images proportionally to their chromatic variety. Furthermore, we annotated 80% of our images with addititional segmentation masks. We incorporate semantic segmentation masks into the dataset by processing the RGB images generated from the 204 spectral channels. We acquired the

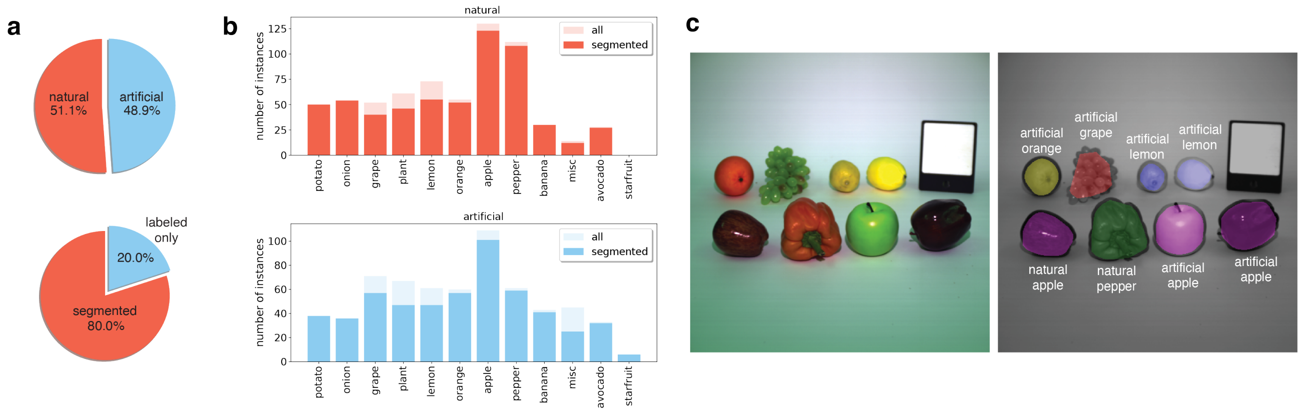


Figure 5. **Example and statistical analysis on our dataset** (a) Overview of the composition of the dataset. There exist a near equal number of natural and artificial objects in the scenes, 80% of the images are with segmentation masks and the rest with labels only. (b) Distribution of scene objects in classes. Each class has a roughly equal number of instances in the dataset with the exception of apples and peppers, as they have more chromatic variety. (c) Left: RGB visualization of hyperspectral image. Right: Segmentation mask and labels for each object.

images in such a way to avoid the intersection of objects, allowing for automatic generation of masks for the areas occupied by each object. We then annotated each marked object, identifying each object class and whether they are natural or artificial. Figure 5c illustrates the implementation of the semantic segmentation mask on an image of the dataset. For more details about the FVgNET dataset please refer to Sec. 3 of Supplementary Material.

5. Results

5.1. Hardware implementation

We fabricate arrays of metasurface projectors by patterning thin layers of amorphous silicon deposited on optical grade fused silica glass slides. Figure 6a shows a scanning electron microscope (SEM) image of a manufactured metasurface pixel, detailing the nanoscale structure of each of the nine projectors. We produce each projector of the 3 × 3 subarray, so it occupies the area of a $2.4\,\mu\mathrm{m}$ wide square, a size typical for the pixels present in modern digital camera sensors, which allows integration with the camera in the scheme of Fig. 1b. We characterize the optical response of each projector by using linearly polarized light with wavelengths from $400\,\mathrm{nm}$ to $1000\,\mathrm{nm}$. Figure 3 in the Supplementary Material shows the experimentally measured responses of the metasurfaces, illustrating excellent agreement with the expected theoretical responses. We utilize the fabricated projector as a fixed encoder to optimize the reconstruction ability of the neural network decoders.

5.2. Spectral Reconstruction

We perform spectral reconstruction from the barcodes obtained from both the theoretical and experimental responses of the fabricated metasurface projectors. Figure 6b shows a

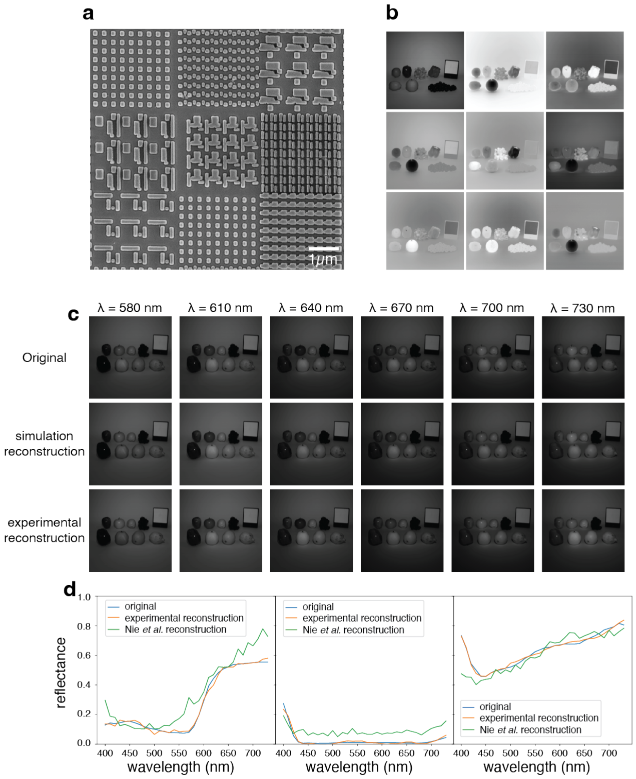


Figure 6. **Spectral reconstruction.** (a) Scanning electron microscope image of the array of projectors. (b) The output of the scene processed by our projectors. (c) Comparison between acquired and recovered hyperspectral image using the theoretical (middle row) and experimental (lower row) responses of our projectors. (d) Comparison between the original spectra and their reconstruction using the projectors and the reconstruction algorithm by Nie *et al*. [39] for random pixels of the scene in (c).

Model	Dataset				
	CAVE [57]	Harvard(out) [11]	Harvard(in) [11]	KAUST [33]	FVgNET
Nguyen *et al.* [38]	14.91±11.09	9.06±9.69	15.61±8.76	-	-
Arad and Ben-Shahar [5]	8.84±7.23	14.89±13.23	9.74±7.45	-	-
Jia *et al.* [25]	7.92±3.33	8.72 ±7.40	9.50±6.32	-	-
Nie *et al.* [39]	4.48 ± 2.97	7.57±4.59	8.88 ± 4.25	-	-
Hyplex™	**2.05**± **1.82**	**2.13** ± **1.81**	**6.65** ± **5.88**	2.23 ± 3.35	1.73 ± 1.35

Table 1. **Comparison of baselines.** We report the RMSE from spectral reconstruction in multiple hyperspectral datasets

scene from the FVgNET dataset as perceived through each of the projectors based on experimental data. In Fig. 6c we present a qualitative comparison between the hyperspectral reconstruction of this scene based on both the simulated and experimental barcodes against the original. Figure 6d illustrates a quantitative comparison between the original spectra and its reconstructions as obtained from the experimental implementation Hyplex™ and the algorithm by Nie *et al.* [39]. The reconstruction is carried out through the use of the connected MLP decoder introduced in Sec. 3. We designate 80% of our dataset for training the decoder and the remainder for validation purposes.

Table 1 presents a performance comparison of Hyplex™ against state-of-the-art reconstruction approaches. We present the results of the reconstruction from the datasets described in Sec. 4, as well as for the validation part of our own dataset. For the consistency of the comparison, we adapted the metrics and data reported in [39], where the calculated RMSE is normalized into the range [0, 255] to approximately represent the error in pixel intensity. The reconstruction error of Hyplex™ is the lowest value among CAVE and both indoor and outdoor images in the Harvard dataset, showing superior performance against all state-of-the-art models. We further tested our model on the KAUST dataset and FVgNET dataset by using the optical response of the fabricated metasurfaces.

5.3. Hyperspectral Semantic Segmentation

Here we present labeling of artificial and real fruits from scenes of the FVgNET dataset. Artificial and real fruits have similar RGB colors. However, they differ significantly in their reflection spectra. Supplementary Fig. 4 provides an example of this. We showcase the learning ability of the proposed physical encoders by training two classification networks. One model uses the spectral encoders for semantic segmentation labeling, and the second the RGB channels. Both models use an identical U-Net-like decoder and identical parameters (number of epochs, batch size, learning rate). The results are summarized in Fig. 7, where the panel a shows a qualitative comparison of the segmentation prediction quality for both models against the ground-truth mask.

	RMSE	mIoU
Simulation	4.23	0.812
Experiment	5.41	0.741

Table 2. **Simulation and experiment results.** We report RMSE and mIoU seperately for reconstruction and segmentation tasks

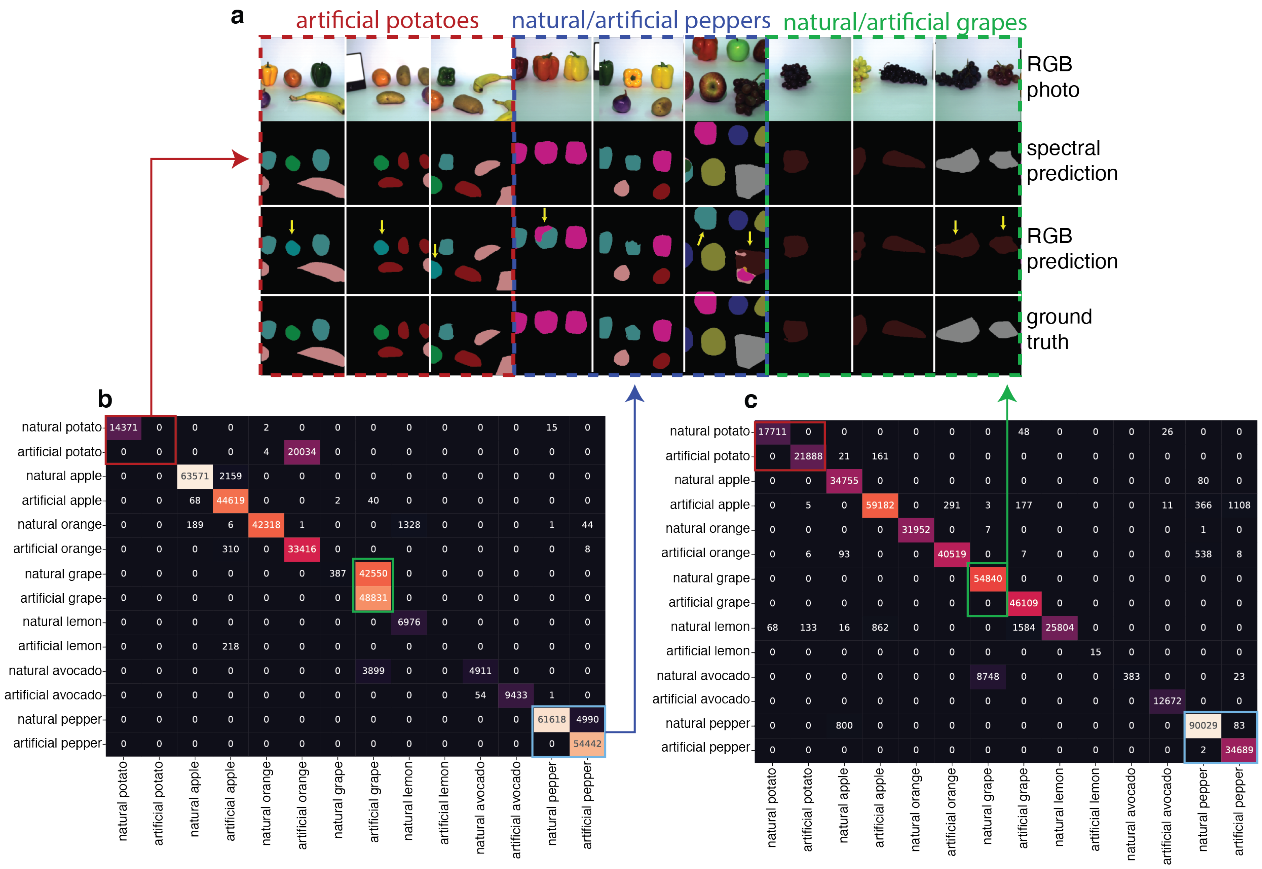
Object class	Hyperspectral segmentation		RGB segmentation	
	IoU	F1	IoU	F1
real orange	**0.979**	**0.989**	0.935	0.966
artificial orange	**0.954**	**0.976**	0.609	0.757
real grape	**0.829**	**0.907**	0.009	0.017
artificial grape	**0.897**	**0.946**	0.494	0.661

Table 3. **Quantitative comparison.** We report 4 examples of object classes segmented with HSI and RGB images.

While the mask quality is similar for both methods, the mean Intersection over Union (IoU) score for the spectral-informed model is significantly higher compared to the RGB one. The mIoU computed with the theoretical and experimental responses of encoders reaches 81%, and 74%, as shown in Tab. 2. With the RGB model, conversely, the mIoU decreases to 68%. The confusion matrix of the RGB trained model shows that the RGB model struggles to predict correct results for real-artificial pairs of fruits with similar colors (Fig. 7b). The spectral-informed model, conversely, generates correct labels for most real-artificial pairs (Fig. 7c) and outperforms the RGB model in IoU and F1 (Tab. 3). These results demonstrate that the small-sized barcodes generated by Hyplex™ efficiently compress spectral features that convey key information about the objects imaged. Table 1 and 2 in Supplementary Material provide detailed metrics for each object type (apple, potato, etc.) on both models.

6. Discussion and Limitations

In this work, we designed and implemented Hyplex™, a new hardware system for real-time and high-resolution hyperspectral imaging. We validated Hyplex™ against current



b

	natural potato	artificial potato	natural apple	artificial apple	natural orange	artificial orange	natural grape	artificial grape	natural lemon	artificial lemon	natural avocado	artificial avocado	natural pepper	artificial pepper
natural potato	14371	0	0	0	2	0	0	0	0	0	0	0	15	0
artificial potato	0	0	0	0	4	20034	0	0	0	0	0	0	0	0
natural apple	0	0	63571	2159	0	0	0	0	0	0	0	0	0	0
artificial apple	0	0	68	44619	0	0	2	40	0	0	0	0	0	0
natural orange	0	0	189	6	42318	1	0	0	1328	0	0	0	1	44
artificial orange	0	0	0	310	0	33416	0	0	0	0	0	0	0	8
natural grape	0	0	0	0	0	0	387	42550	0	0	0	0	0	0
artificial grape	0	0	0	0	0	0	0	48831	0	0	0	0	0	0
natural lemon	0	0	0	0	0	0	0	0	6976	0	0	0	0	0
artificial lemon	0	0	0	218	0	0	0	0	0	0	0	0	0	0
natural avocado	0	0	0	0	0	0	0	3899	0	0	4911	0	0	0
artificial avocado	0	0	0	0	0	0	0	0	0	0	54	9433	1	0
natural pepper	0	0	0	0	0	0	0	0	0	0	0	0	61618	4990
artificial pepper	0	0	0	0	0	0	0	0	0	0	0	0	0	54442

c

	natural potato	artificial potato	natural apple	artificial apple	natural orange	artificial orange	natural grape	artificial grape	natural lemon	artificial lemon	natural avocado	artificial avocado	natural pepper	artificial pepper
natural potato	17711	0	0	0	0	0	0	48	0	0	0	26	0	0
artificial potato	0	21888	21	161	0	0	0	0	0	0	0	0	0	0
natural apple	0	0	34755	0	0	0	0	0	0	0	0	0	80	0
artificial apple	0	5	0	59182	0	291	3	177	0	0	0	11	366	1108
natural orange	0	0	0	0	31952	0	7	0	0	0	0	0	1	0
artificial orange	0	6	93	0	0	40519	0	7	0	0	0	0	538	8
natural grape	0	0	0	0	0	0	54840	0	0	0	0	0	0	0
artificial grape	0	0	0	0	0	0	0	46109	0	0	0	0	0	0
natural lemon	68	133	16	862	0	0	0	1584	25804	0	0	0	0	0
artificial lemon	0	0	0	0	0	0	0	0	0	15	0	0	0	0
natural avocado	0	0	0	0	0	0	8748	0	0	0	383	0	0	23
artificial avocado	0	0	0	0	0	0	0	0	0	0	0	12672	0	0
natural pepper	0	0	800	0	0	0	0	0	0	0	0	0	90029	83
artificial pepper	0	0	0	0	0	0	0	0	0	0	0	0	2	34689

Figure 7. **Spectral and RGB-based semantic segmentations.** (a) Comparison between segmentation masks generated from a spectral-informed model, an RGB-only model, and the ground truth. (b) Confusion matrix for RGB only model. (c) Confusion matrix for the spectral-informed model. Each value in the confusion matrix represents the number of pixels of the segmentation mask of the item in the column that was classified as the item in the row.

state-of-the-art approaches and proved it to be outperforming in all benchmarks. Additionally, we demonstrated the superiority of hyperspectral features and trainable encoders by designing a model for spectral-informed semantic segmentation and comparing its performance against RGB models.

One of the limitations in the current implementation of Hyplex™ is the linear structure of the physical encoder. The study of nonlinear encoders [3] could enable more complex feature embeddings. This topic may stimulate future research that could generalize the Hyplex™ framework to include nonlinear metasurfaces, an essential area of research in the field of meta-optics [30, 35]. The second area of improvement is the spectral sparsity assumption at the core idea of efficient dimensionality reduction. While this assumption is practically verified in the majority of computer vision problems [43,61], it may not hold for specialized tasks. Fabrication errors are also an essential aspect that, if not adequately considered, can limit performance. In this work, we mitigate this effect by tuning the software decoder to best use the experimental response of the projectors. Future work could investigate techniques from robustness control in inverse design, a new promising area of research [31, 37].

Improved results could also be obtained if we augment the publicly available hyperspectral datasets with more scenes obtained at different wavelengths and in different settings such as, e.g., medical. Such study could generalize the results of Hyplex™ to provide high impact systems for personalized healthcare and precision medicine. Hyplex™ could provide a game-changer technology in this field, leveraging its vast capacity to fast-process high-resolution hyperspectral images (see Sec. 7 of Supplementary Material) at speed comparable with current RGB cameras.

Acknowledgements. This work was supported by the King Abdullah University of Science and Technology (KAUST) through the Artificial Intelligence Initiative (AII) funding. This research received funding from KAUST (Award OSR-2016-CRG5-2995). Parallel simulations are performed on KAUST's Shaheen supercomputer.

References

[1] M.A. Afromowitz, J.B. Callis, D.M. Heimbach, L.A. DeSoto, and M.K. Norton. Multispectral imaging of burn wounds: A new clinical instrument for evaluating burn depth. *IEEE Transactions on Biomedical Engineering*, 35(10):842–850, Oct. 1988. 1

[2] Airobot. Hyperspectral drone and software for agriculture, 2021. https://airobot.eu/solutions/hyperspectral-drone-and-software-for-agriculture/. 1

[3] Xavier Alameda-Pineda, Elisa Ricci, Yan Yan, and Nicu Sebe. Recognizing emotions from abstract paintings using non-linear matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 8

[4] Aitor Alvarez-Gila, Joost Van De Weijer, and Estibaliz Garrote. Adversarial networks for spatial context-aware spectral image reconstruction from rgb. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Oct 2017. 2

[5] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016. 2, 7

[6] B. Arad and O. Ben-Shahar. Filter selection for hyperspectral estimation. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3172–3180, 2017. 2

[7] Arad B. and Ben-Shahar O. Sparse recovery of hyperspectral signal from natural rgb images. In Leibe B., Matas J., Sebe N., and Welling M, editors, *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, volume 9911. Springer, Cham, 2016. 2

[8] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. 3, 5

[9] Marcella Bonifazi, Valerio Mazzone, Ning Li, Yi Tian, and Andrea Fratalocchi. Free-Electron Transparent Metasurfaces with Controllable Losses for Broadband Light Manipulation with Nanometer Resolution. *Advanced Optical Materials*, 8(1):1900849, 2020. 2

[10] Renfu Lu Bosoon Park. *Hyperspectral Imaging Technology in Food and Agriculture*. Springer, 2015. 1

[11] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011. 7

[12] Wei Ting Chen, Alexander Y Zhu, Vyshakh Sanjeev, Mohammadreza Khorasaninejad, Zhujun Shi, Eric Lee, and Federico Capasso. A broadband achromatic metalens for focusing and imaging in the visible. *Nature Nanotechnology*, 13(3):220–226, Jan. 2018. 2

[13] Arjun Chennu, Paul Färber, Glenn De'ath, Dirk de Beer, and Katharina E. Fabricius. A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Scientific Reports*, 7(1):7122, Aug. 2017. 1

[14] Ines Dumke, Autun Purser, Yann Marcon, Stein M. Nornes, Geir Johnsen, Martin Ludvigsen, and Fredrik Søreide. Underwater hyperspectral imaging as an in situ taxonomic tool for deep-sea megafauna. *Scientific Reports*, 8(1):12860, Aug. 2018. 1

[15] David H Foster, Kinjiro Amano, Sérgio MC Nascimento, and Michael J Foster. Frequency of metamerism in natural scenes. *Josa a*, 23(10):2359–2372, 2006. 2

[16] Andrea Fratalocchi, Fedor Getman, Maksim Makarenko, and Arturo Burguete-Lopez. Flat optics polarizer beam splitter. 2

[17] Henning Galinski, Gael Favraud, Hao Dong, Juan S. Totero Gongora, Grégory Favaro, Max Döbeli, Ralph Spolenak, Andrea Fratalocchi, and Federico Capasso. Scalable, ultra-resistant structural colors based on network metamaterials. *Light: Science & Applications*, 6(5):e16233–e16233, May 2017. 2

[18] Silvano Galliani, Charis Lanaras, Dimitrios Marmanis, Emmanuel Baltsavias, and Konrad Schindler. Learned spectral super-resolution. *arXiv preprint arXiv:1703.09470*, 03 2017. 2

[19] F. Getman, M. Makarenko, A. Burguete-Lopez, and A. Fratalocchi. Broadband vectorial ultrathin optics with experimental efficiency up to 99% in the visible region via universal approximators. *Light: Science & Applications*, 10(1):1–14, Mar. 2021. 2, 3

[20] Fedor Getman, Maksim Makarenko, Arturo Burguete-Lopez, and Andrea Fratalocchi. Broadband vectorial ultrathin optics with experimental efficiency up to 99% in the visible region via universal approximators. *Light: Science & Applications*, 10(1):1–14, 2021. 3

[21] Aoife A. Gowen, Yaoze Feng, Edurne Gaston, and Vasilis Valdramidis. Recent applications of hyperspectral imaging in microbiology. *Talanta*, 137:43–54, May 2015. 1

[22] Qiong He, Shulin Sun, Shiyi Xiao, and Lei Zhou. High-efficiency metasurfaces: Principles, realizations, and applications. *Advanced Optical Materials*, 6(19):1800415, 2018. 2

[23] Noor A Ibraheem, Mokhtar M Hasan, Rafiqul Z Khan, and Pramod K Mishra. Understanding color models: a review. *ARPN Journal of science and technology*, 2(3):265–275, 2012. 2

[24] Daniel S. Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H. Kim. Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Transactions on Graphics (Proc. SIGGRAPH 2019)*, 38(4):117:1–13, 2019. 2

[25] Yan Jia, Yinqiang Zheng, Lin Gu, Art Subpa-Asa, Antony Lam, Yoichi Sato, and Imari Sato. From rgb to spectrum for natural scenes via manifold-based mapping. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 2, 7

[26] Jun Jiang, Dengyu Liu, Jinwei Gu, and S. Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 168–179, 2013. 2

[27] Rei Kawakami, Yasuyuki Matsushita, John Wright, Moshe Ben-Ezra, Yu-Wing Tai, and Katsushi Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *CVPR 2011*, pages 2329–2336. IEEE, 2011. 2

[28] Nasser Kehtarnavaz and Mark Gamadia. Real-Time Image and Video Processing: From Research to Reality. *Synthesis*

Lectures on Image, Video, and Multimedia Processing, 2(1):1–108, Jan. 2006. 2

[29] Muhammad Jaleed Khan, Hamid Saeed Khan, Adeel Yousaf, Khurram Khurshid, and Asad Abbas. Modern trends in hyperspectral image analysis: A review. *IEEE Access*, 6:14118–14129, 2018. 1

[30] Yuri Kivshar. All-dielectric meta-optics and non-linear nanophotonics. *National Science Review*, 5(2):144–158, Mar 2018. 8

[31] Julius Kühne, Juan Wang, Thomas Weber, Lucca Kühner, Stefan A Maier, and Andreas Tittl. Fabrication robustness in bic metasurfaces. *Nanophotonics*, 2021. 8

[32] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808, 2007. 2

[33] Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *2021 IEEE International Conference on Computer Vision(ICCV)*, pages 1–8. IEEE, 2021. 7

[34] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: A review. *Journal of Biomedical Optics*, 19(1):010901, Jan. 2014. 1

[35] Stefan A. Maier. Dielectric and low-dimensional-materials nanocavities for non-linear nanophotonics and sensing. In *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF) (2018), paper SeW2E.4*, page SeW2E.4. Optical Society of America, Jul 2018. 8

[36] M. Makarenko, A. Burguete-Lopez, F. Getman, and A. Fratalocchi. Generalized maxwell projections for multi-mode network photonics. *Scientific Reports*, 10(1):9038, Dec 2020. 3, 5

[37] Maksim Makarenko, Qizhou Wang, Arturo Burguete-Lopez, Fedor Getman, and Andrea Fratalocchi. Robust and scalable flat-optics on flexible substrates via evolutionary neural networks. *Advanced Intelligent Systems*, page 2100105, Aug 2021. 2, 5, 8

[38] Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *European Conference on Computer Vision*, pages 186–201. Springer, 2014. 2, 7

[39] Shijie Nie, Lin Gu, Yinqiang Zheng, Antony Lam, Nobutaka Ono, and Imari Sato. Deeply learned filter response functions for hyperspectral reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4767–4776, 2018. 2, 6, 7

[40] Seoung Wug Oh, Michael S Brown, Marc Pollefeys, and Seon Joo Kim. Do it yourself hyperspectral imaging with everyday digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2461–2469, 2016. 2

[41] Yaniv Oiknine, Isaac August, Vladimir Farber, Daniel Gedalin, and Adrian Stern. Compressive Sensing Hyperspectral Imaging by Spectral Multiplexing with Liquid Crystal. *Journal of Imaging*, 5(1):3, Jan. 2019. 1

[42] Svetlana V. Panasyuk, Shi Yang, Douglas V. Faller, Duyen Ngo, Robert A. Lew, Jenny E. Freeman, and Adrianne E. Rogers. Medical hyperspectral imaging to facilitate residual tumor identification during surgery. *Cancer Biology & Therapy*, 6(3):439–446, Mar. 2007. 1

[43] Arun P.V., Krishna Mohan B., and Porwal A. Spatial-spectral feature based approach towards convolutional sparse coding of hyperspectral images. *Computer Vision and Image Understanding*, 188:102797, 2019. 8

[44] Antonio Robles-Kelly. Single image spectral reconstruction for multimedia applications. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, page 251–260, New York, NY, USA, 2015. Association for Computing Machinery. 2

[45] Antonio Robles-Kelly. Single image spectral reconstruction for multimedia applications. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 251–260, 2015. 2

[46] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2

[47] Hui-Liang Shen, Jian-Fan Yao, Chunguang Li, Xin Du, Si-Jie Shao, and John H Xin. Channel selection for multispectral color imaging using binary differential evolution. *Applied optics*, 53(4):634–642, 2014. 2

[48] Andreas Tittl, Aurelian John-Herpin, Aleksandrs Leitis, Eduardo R. Arvelo, and Hatice Altug. Metasurface-Based Molecular Biosensing Aided by Artificial Intelligence. *Angewandte Chemie International Edition*, 58(42):14810–14822, Oct. 2019. 2

[49] Andreas Tittl, Aleksandrs Leitis, Mingkai Liu, Filiz Yesilkoy, Duk-Yong Choi, Dragomir N Neshev, Yuri S Kivshar, and Hatice Altug. Imaging-based molecular barcoding with pixelated dielectric metasurfaces. *Science*, 360(6393):1105–1109, 2018. 2

[50] Andreas Tittl, Aleksandrs Leitis, Mingkai Liu, Filiz Yesilkoy, Duk-Yong Choi, Dragomir N. Neshev, Yuri S. Kivshar, and Hatice Altug. Imaging-based molecular barcoding with pixelated dielectric metasurfaces. *Science*, 360(6393):1105–1109, 2018. 2

[51] Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. Neural nano-optics for high-quality thin lens imaging. *Nature communications*, 12(1):1–7, 2021. 3

[52] Qizhou Wang, Maksim Makarenko, Arturo Burguete Lopez, Fedor Getman, and Andrea Fratalocchi. Advancing statistical learning and artificial intelligence in nanophotonics inverse design. *Nanophotonics*, 2021. 3

[53] Renjie Wu, Yuqi Li, Xijiong Xie, and Zhijie Lin. Optimized multi-spectral filter arrays for spectral reconstruction. *Sensors*, 19(13):2905, 2019. 2

[54] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 518–525, 2017. 2

[55] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 518–525, 2017. 2

[56] Liutao Yang, Zhongnian Li, Zongxiang Pei, and Daoqiang Zhang. Fs-net: Filter selection network for hyperspectral reconstruction. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2933–2937. IEEE, 2021. 3

[57] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, Columbia University, Nov 2008. 7

[58] Seung Chul Yoon, Bosoon Park, Kurt C. Lawrence, William R. Windham, and Gerald W. Heitschmidt. Line-scan hyperspectral imaging system for real-time inspection of poultry carcasses with fecal material and ingesta. *Computers and Electronics in Agriculture*, 79(2):159–168, Nov. 2011. 1

[59] Jun Yu, Toru Kurihara, and Shu Zhan. Optical filter net: A spectral-aware rgb camera framework for effective green pepper segmentation. *IEEE Access*, 9:90142–90152, 2021. 3

[60] Nanfang Yu and Federico Capasso. Flat optics with designer metasurfaces. *Nature Materials*, 13(2):139–150, Feb. 2014. 2

[61] Lei Zhang, Wei Wei, Yanning Zhang, Chunhua Shen, Anton van den Hengel, and Qinfeng Shi. Cluster sparsity field: An internal hyperspectral imagery prior for reconstruction. *International Journal of Computer Vision*, 126(8):797–821, Aug 2018. 8

[62] Wenyi Zhang, Hongya Song, Xin He, Longqian Huang, Xiyue Zhang, Junyan Zheng, Weidong Shen, Xiang Hao, and Xu Liu. Deeply learned broadband encoding stochastic hyperspectral imaging. *Light: Science & Applications*, 10(1):1–7, 2021. 3

[63] Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. Hierarchical regression network for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 422–423, 2020. 2