Real-time Hyperspectral Imaging in Hardware via Trained Metasurface Encoders

Maksim Makarenko, Arturo Burguete-Lopez, Qizhou Wang, Fedor Getman, Silvio Giancola, Bernard Ghanem, Andrea Fratalocchi

1King 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

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).
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 designed 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 x 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 $\beta$ 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) $\hat{\beta} = D_{\text{rec}}(\mathcal{E}(\beta))$ between reconstructed and input spectra. Semantic segmentation, conversely, provides a pixel-by-pixel classification of HSI. In this task, we use as decoder $D_{\text{seg}}$ 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 $L_{\text{seg}}$.

### 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 $\mathcal{E} = \hat{\Lambda}^\dagger$. Panel (a) shows an example hyperspectral image. The data is represented as a tensor $\beta$ with three dimensions: two spatial dimensions $(x, y)$, corresponding to the camera virtual image plane, and one frequency axis $\omega$, 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 $\beta$ (Fig. 3c). The hyperspectral tensor of a dataset of images is flattened to a matrix $B$ that contains, on each column, the power density spectra of a set of camera pixels. We then apply the linear encoding $\Lambda^\dagger$ to obtain an approximation of $B$ [8] via a set of linear projectors $\Lambda(\omega)$, which map pixel-by-pixel the spectral coordinate $\beta_{ij}$ to a set of scalar coefficients $S_{ijk}$:

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

The spectral information contained in $\beta_{ij}(\omega)$ is embedded into an equivalent barcode $S_{ijk}$ of a few components.
Figure 3. **Metasurface subpixel array as a linear spectral encoder.** (a) A spectral image tensor ($\beta$) 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 $\Lambda^\dagger$ of a flattened matrix $B$ with the resulted projected encoder barcode for a pixel spectra at the $ij$ position. (d) Optimal encoder functions $\Lambda^\dagger$. (e) Non-differentiable inverse design optimization framework implemented via ALFRED utilized to find a set of metasurfaces $L$ with desired response $\Lambda_i$. (f) Differentiable backbone enabling simultaneous optimization of responses $\Lambda$ and structures $L$. Metasurface pixel (g) composed of a two-dimensional array of resonant metapixels with corresponding fitted transmission responses $\hat{\Lambda}$. (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 $D_{rec}$ projection $\hat{\beta}_{ij}$.

To implement the $\Lambda$ 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 $\Lambda$ and the target $\Lambda$ (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 $L_{seg}$ by optimizing simultaneously the projector responses $\Lambda$ and the vector $L$ 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 $\Lambda$ obtained through an unsupervised learning technique via principal component analysis (PCA). The PCA performs hardware encoding $E$ by selecting the $k$ strongest ($k = 9$ for this work) principal components $\hat{\Lambda}^1$ from the
singular value decomposition of $B = \Lambda \Sigma V^\dagger$ [8], and approximating $B$ as follows:

$$B \approx \tilde{\Lambda} \tilde{\Sigma} \tilde{V}^\dagger$$

(2)

Equation (2) offers the closest linear approximation of $B$ in least square sense. We implement the decoder $D$ with the linear projector $\beta_{ij} = \tilde{\Lambda} S_{ij}$, 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 $D$ as a set of hierarchical nonlinear operators $F$, which project the input tensor $\tilde{S}$ 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 $\tilde{S}$ and the associated linear projectors $\Lambda$. To train Hyplex™ in this framework with backpropagation, the encoder $E$ needs to be differentiable.

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

$$\begin{align*}
\begin{cases}
\tilde{a}(\omega) &= \frac{K}{i(\omega - W) + KR^\dagger} \tilde{s}_+ \\
\tilde{s}_-(\omega) &= \tilde{C}(\omega) \cdot \left(\tilde{s}_+ - \tilde{K}^\dagger \cdot \tilde{a}\right)
\end{cases}
\end{align*}$$

(3)

where $W$ is a diagonal matrix with resonant frequencies $\omega_n$ of the modes $W_{nn} = \omega_n$, $\tilde{C}(\omega)$ is a scattering matrix modeling the scattering of impinging waves $\tilde{s}_+$ on the resonator space, and $\tilde{K}$ is a coupling matrix representing the interaction between traveling waves $\tilde{s}_+ (t)$ and resonator modes $\tilde{a}(t)$. Equations (3) describe the dynamics of a network of resonator modes $\tilde{a} = [\tilde{a}_1(\omega), \ldots, \tilde{a}_n(\omega)]$, interacting with $\tilde{s}_+ = [\tilde{s}_1(\omega), \ldots, \tilde{s}_m(\omega)]$ 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 $H = \tilde{s}_- / \tilde{s}_+$ 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 $\frac{K}{i(\omega - W)}$. Equation (3) represents a differentiable function of $W$ through which it is possible to backpropagate (Fig. 4b).

**d-ALFRED:** To project the resonator quantities in Eq. (3) to metasurface input parameters $L$, we use a supervised optimization process. We train a deep neural network to learn the relationship between $L$ 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).

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 classes 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 additional 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.

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 $3 \times 3$ sub-array, so it occupies the area of a $2.4\mu m$ 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 nm to 1000 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
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.

<table>
<thead>
<tr>
<th>Object class</th>
<th>Hyperspectral segmentation</th>
<th>RGB segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IoU</td>
<td>F1</td>
</tr>
<tr>
<td>real orange</td>
<td>0.979</td>
<td>0.989</td>
</tr>
<tr>
<td>artificial orange</td>
<td>0.954</td>
<td>0.976</td>
</tr>
<tr>
<td>real grape</td>
<td>0.829</td>
<td>0.907</td>
</tr>
<tr>
<td>artificial grape</td>
<td>0.897</td>
<td>0.946</td>
</tr>
</tbody>
</table>

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
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


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


12701


