

Hardware-accelerated integrated optoelectronic platform towards real-time high-resolution hyperspectral video understanding

Received: 7 April 2024

Accepted: 5 August 2024

Published online: 15 August 2024

 Check for updates

Maksim Makarenko^{1,2,5}, Arturo Burguete-Lopez^{1,5}, Qizhou Wang¹,
Silvio Giancola³, Bernard Ghanem³, Luca Passone⁴ & Andrea Fratalocchi¹ ✉

Recent advancements in artificial intelligence have significantly expanded capabilities in processing language and images. However, the challenge of comprehensively understanding video content still needs to be solved. The main problem is the requirement to process real-time multidimensional video information at data rates exceeding 1 Tb/s, a demand that current hardware technologies cannot meet. This work introduces a hardware-accelerated integrated optoelectronic platform specifically designed for the real-time analysis of multidimensional video. By leveraging optical information processing within artificial intelligence hardware and combining it with advanced machine vision networks, the platform achieves data processing speeds of 1.2 Tb/s. This capability supports the analysis of hundreds of frequency bands with megapixel spatial resolution at video frame rates, significantly outperforming existing technologies in speed by three to four orders of magnitude. The platform demonstrates effectiveness for AI-driven tasks, such as video semantic segmentation and object understanding, across indoor and aerial scenarios. By overcoming the current data processing speed limitations, the platform shows promise in real-time AI video understanding, with potential implications for enhancing human-machine interactions and advancing cognitive processing technologies.

The autonomous processing of big data via artificial intelligence (AI)^{1–3} is opening frontiers in medical^{4,5}, security^{6,7}, robotics^{8–11}, automated speech recognition¹², and natural language processing¹³ with human-like and—in some cases—better than human performances. Foundation models are accelerating this development significantly with the

emergence of an understanding of human languages, carrying out tasks the designer never trained the model on^{14–16}. The latest generation of foundation models in GPT-4 technology extend such learning abilities by combining information from languages and images using different data modalities^{14,17,18}, while the recently proposed Gato model

¹PRIMALIGHT, Faculty of Electrical Engineering; Applied Mathematics and Computational Science, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. ²AI & Advanced Computing Lab, EXPEC ARC, Saudi Aramco, 4143 Dhahran Blvd, Gharb Al Dhahran, Dhahran 34466, Saudi Arabia. ³Image and Video Understanding Lab, Faculty of Electrical Engineering; Applied Mathematics and Computational Science, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. ⁴Falconviz, King Abdullah University of Science and Technology Research Park Headquarters - Level 1 - Office 2225, Thuwal 23955-6900, Saudi Arabia. ⁵These authors contributed equally: Maksim Makarenko, Arturo Burguete-Lopez.

✉ e-mail: andrea.fratalocchi@kaust.edu.sa

allows generalist agents to communicate, process pictures, play video games, and control robotic arms at the same time^{16,19}.

The speed of these developments has significantly outpaced the velocity of hardware progress, with a roadblock looming on multi-modal AI applications if new technological platforms to acquire and process data do not materialize²⁰. This issue is particularly significant in the emerging horizon of video understanding, representing the subsequent large language model (LLM) development, which requires grasping context in four-dimensional spatial and temporal data^{21–23}.

Hyperspectral imaging is the closest existing technology to multidimensional data flow acquisitions^{24–27}. Hyperspectral imaging augments two-dimensional spatial pixels with a third dimension comprising hundreds of frequency channels corresponding to narrow portions of the spectrum within and beyond the visible range²⁸. These frequency bands encompass spectral signatures essential for the identification, measurement, and classification of objects, materials, and compounds while enabling remote monitoring of their properties in diverse processes of industrial interest^{4,5,29–32}.

Figure 1 provides an overview of the performance of current spectral imaging equipment, including both commercially available devices and those demonstrated in academic literature. We compute the resolution of each camera as the product of the megapixels per frequency band and the number of frequency bands they output, indicated in parenthesis for each point. Additionally, we list the frame rate of every device after the @ symbol. For each point, we determine the data rate as the product of the resolution, frame rate, and bit depth of the pixels in each band. In the current standard for digital video³³, a high-resolution hyperspectral datastream at 4 K (3840 × 2160 pixels), acquired with hundreds of bands in the visible range between the wavelengths of 350 nm and 750 nm, and with 12 bits per band, requires processing a data rate over 1 Tb/s.

Currently, the best snapshot hyperspectral devices capable of recording more than a hundred frequency bands possess between three and four orders of magnitude slower data rates and cannot record at video speed^{34,35} (Fig. 1). Faster hyperspectral and multispectral technologies with frame rates in the 100 Gb/s range reduce the spectral resolution by one order of magnitude^{36–42}. At the same time, accurate one-dimensional scanners^{43–47} do not meet the spatial resolution required to capture 2D image flows at video rates. The critical challenge to reaching real-time Tb/s multimodal data processing is the speed of data transmission in electronics. State-of-the-art DDR5

memory, with a bandwidth of 500 Gb/s⁴⁸, exemplifies this barrier that current technology cannot yet overcome (Fig. 1 dashed line).

While most existing research focuses on acquiring data from static objects in controlled indoor settings with fixed lighting conditions and scenes, dynamic outdoor data acquisition and processing presents a challenge^{49–51}. Indoor environments allow controlling factors such as specular reflections, unwanted shadows, humidity, uneven spatial and temporal illumination, scene and camera movement, and the presence of different illumination sources, which dynamically change outdoors and lead to degraded hyperspectral analysis^{52–54}. Supplementary Note 1 and Supplementary Fig. 1 provide experimental examples of these issues with commercially available hyperspectral devices.

Addressing these problems provides a substantial opportunity in research to significantly improve this technology, unlocking future advancements in a wide number of critical applications in the medical, life sciences, forensics, security, pharmaceutical, environmental, mining and oil industries that real-time cognitive processing of high-resolution multidimensional visual data flows could empower^{4,5,29–32,55–62}.

In this work, we present and validate a hardware-accelerated video understating platform for modern AI learning tasks. We demonstrate hyperspectral video recording at a rate of 1.175 Tb/s with >200 frequency bands at 30 FPS and 2% spectral reconstruction error. Additionally, we present a pipeline that combines motion and spectral information for video object segmentation and tracking, and show experimental results on aerial and indoor video sequences.

Results

Figure 2 illustrates the architecture of the hardware-accelerated platform for real-time hyperspectral video understanding we propose to address the aforementioned issues. A video sequence comprising a succession of frames ($\dots, \beta^{t-1}, \beta^t, \dots$) with β^t corresponding to the frame at time t represents the input of the system (Fig. 2a, b). Each frame β^t contains a three-dimensional data representation of the optical information flow, with $\beta_{ij}^t(\omega)$ representing the power density spectrum emanating from a single spatial point (i, j) in the scene (Fig. 2b, solid red area).

Figure 2c illustrates the physical components of the hardware-accelerated platform, comprising a camera sensor modified by integrating an encoder array E_s over the sensor's pixels. This array, represented by the color squares in the panel, consists of a repeating pattern of nanostructured encoders that extract spectral features from each video frame. Figure 2d, e present photographs of the experimental implementation of this concept. Figure 2d shows a view of the manufactured encoder array at 100 × magnification, with each colored square corresponding to a different encoder design. Figure 2e shows an encoder array integrated on a monochrome camera sensor board capable of recording 12 megapixels video at 30 frames per second (DFK 37AUX226 from The Imaging Source). The integrated array acts analogously to the Bayer array of a color camera, with each encoder covering one camera sensor pixel, and the designs tiling the sensor area in groups of nine. Figure 2f shows a scanning electron microscope image of one among the nine encoder groups that tile the sensor array. Each encoder consists of a repeating free-form nanoresonator geometry and possesses a transmission function $\hat{\Lambda}_k(\omega)$, with $k=1, \dots, 9$.

When the input data flow impinges on one encoder, the camera sensor converts the spectrum emanating from a spatial point of the video frame into a digital scalar coefficient \hat{S}_{ijk}^t read by the camera hardware:

$$\hat{S}_{ijk}^t = \sigma \left(\int \beta_{ij}^t(\omega) \hat{\Lambda}_k(\omega) d\omega \right), \quad (1)$$

where $\sigma(x)$ is the readout input-output response of the single camera pixel⁶³, or an added nonlinearity implemented in software, and $\hat{\Lambda}_k(\omega)$ is

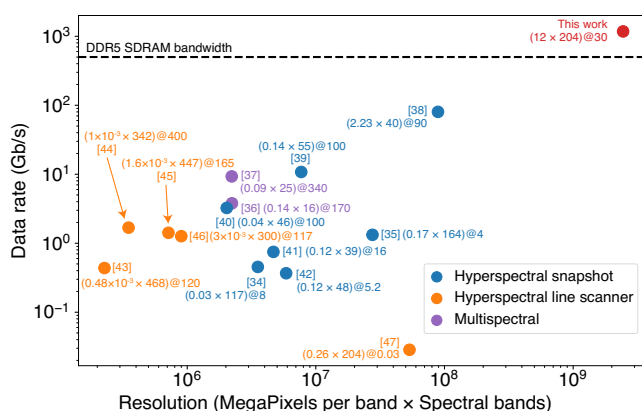


Fig. 1 | State-of-the-art in the acquisition of multidimensional optical information flow. Next to each technology, we indicate the number of megapixels times the number of spectral bands @ frames per second (FPS), as the producer specifies in the technology datasheet of the hardware used. The dashed line indicates a state-of-the-art DDR5 memory's ideal bandwidth, which provides the upper theoretical limit of any electronic technology requiring data transfer. Source data are provided as this figure data.

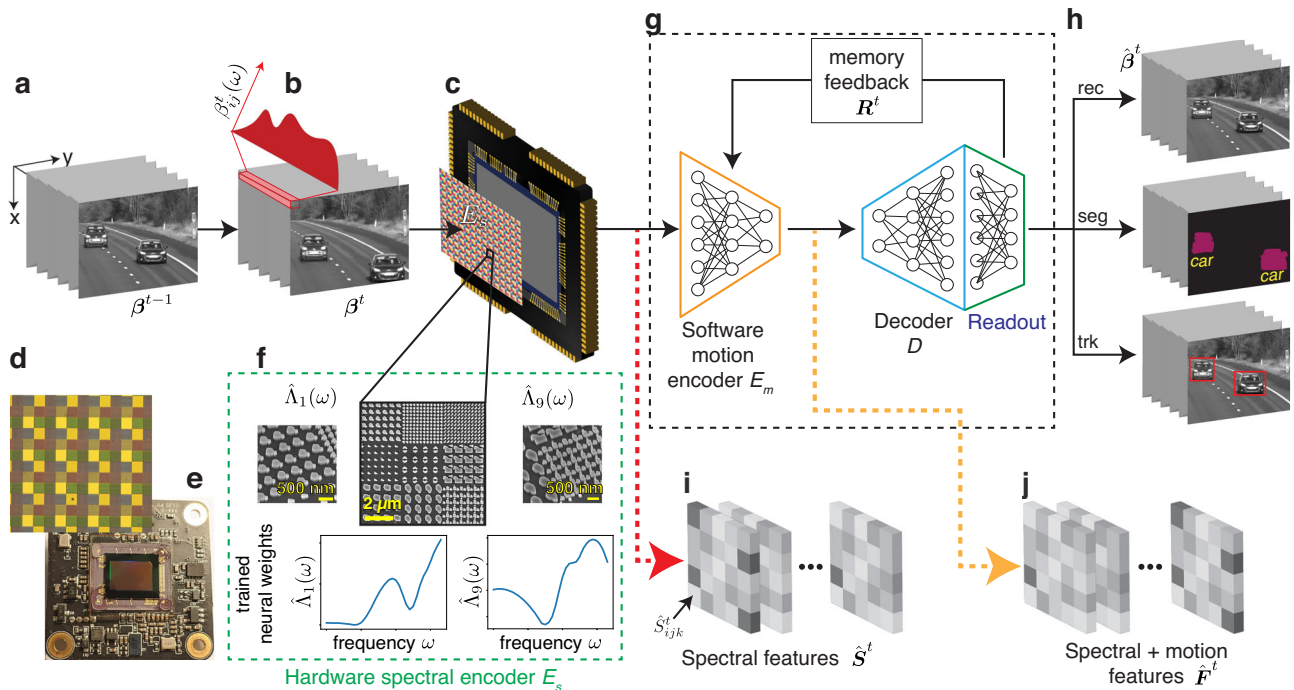


Fig. 2 | Hardware-accelerated video understanding platform. **a** Example of hyperspectral video comprising frames β^t over time. **b** Power density distribution $\beta_{ij}^t(\omega)$ emanating from one pixel (i, j) in the scene. **c** Schematics of hardware encoder placed on top of a camera sensor. **d** Microscope image of a fabricated $N_k = 9$ encoder array. **e** The visual appearance of an experimental board camera sensor integrated with the hardware encoder of (c). **f** Scanning electron microscope

(SEM) images of the hardware encoder E_s composed of a set of nanoresonators with trained transmission functions $\Lambda_k(\omega)$ acting as neural weights for feature extraction. **g** Recurrent AI module comprising a motion encoder, decoder unit, and readout. **h** Different video tasks computed by this platform (rec: reconstruction, seg: segmentation, trk: tracking). **i** Schematic illustration of spectral features read by (c). **j** Schematic representation of combined spectral and motion features.

the encoder pixel's transmission function. We inversely design the nanoresonators in each pixel using universal approximators based on the methods described in^{64–66}, however, in contrast to these works, we employ a pipeline that permits free-form optimization. Free-form nanostructure design permits achieving encoder responses that can capture spectral features inaccessible to cuboid-based designs (See Supplementary Note 2 and Supplementary Fig. 2), which results in a device better suited for non-ideal outdoor illumination. Fabrication of these free-form structures involves large-area electron beam lithography patterning, detailed in the Methods section.

Because these nanostructures can approximate arbitrary responses⁶⁵, it is possible to inverse-design the transmission $\hat{\Lambda}_k(\omega)$ of each encoder to represent a user-defined distribution of amplitude coefficients for the spectral coordinate ω . In this condition, the camera integration (1) implements the hardware equivalent of a neural network's multiply-accumulate (MAC) operation, where $\beta_{ij}^t(\omega)$ represents the input and $\hat{\Lambda}_k(\omega)$ represent neural weights distributed along the frequency axis ω . We train the weights $\hat{\Lambda}_k(\omega)$ to implement feature extraction tasks of a traditional software neural network (see Methods). The hardware encoder array performs feature extraction operations at optical speed and in parallel for every pixel, generating a flow of sparse spectral features \hat{S}^t (Fig. 2i). The camera hardware reads the feature flow \hat{S}^t and sends it to the software motion encoder E_m (Fig. 2g). The encoder E_m combines spectral and motion features extracted from the data flow into the feature flow tensor \hat{F}^t (Fig. 2j). The motion features comprise dynamic temporal changes between video frames, including the direction and speed of movement of objects, changes in the flow composition, and variations in illumination over time. The motion encoder processes these features in real-time with a memory feedback R^t comprising information extracted from previous time-frames (Fig. 2g, feedback loop). The feature flow projects sequentially into a decoder D terminating in a nonlinear

readout (Fig. 2g, right side). The decoder processes the feature tensor \hat{F}^t for different end-to-end optimizations of user-defined tasks, including spectral video reconstruction $\hat{\beta}^t$, video object segmentation⁵⁷, and spectral object tracking⁵⁸ (Fig. 2h).

Spectral video reconstruction

The goal of spectral reconstruction is to predict the visual flow $\hat{\beta}^t$ while minimizing the difference $\Delta_\omega = \|\beta^t - \hat{\beta}^t\|$ with the original input β^t (Fig. 2h, rec). We configure the system of Fig. 2 for this task by disconnecting the recurrent feedback R^t unit (Fig. 2g, top) and set the motion encoder E_m as the identity operator I , carrying out all encoding operations through the hardware encoder E_s . For the type of encoding used in this work (see Methods), we decode the information flow through the projector $D = \hat{A}^\dagger(\omega)$, with $\hat{A}(\omega) = [\hat{\Lambda}_1(\omega), \dots, \hat{\Lambda}_k(\omega)]$ being the set of trained encoder transmission functions. When the projector D operates on the features flow \hat{S}_{ij}^t arising from one camera pixel i, j , we obtain $\hat{\beta}_{ij}^t(\omega) = D\hat{S}_{ij}^t = \hat{A}^\dagger(\omega)\hat{S}_{ij}^t = \sum_k \hat{\Lambda}_k(\omega)\hat{S}_{ijk}^t$, which represents the optimal least square approximation of the spectral video information flow⁶⁷.

After integrating the hardware encoder, we perform a demosaicing operation similar to the debayering process used in traditional color cameras to preserve the sensor's resolution. We reconstruct full spectral information from the camera readings using a bilinear interpolation algorithm, where we estimate each pixel's missing spectral values through the arithmetic mean of adjacent pixels within the same spectral band. We apply the process iteratively to each channel using diluted convolution kernels, ensuring the recovery of a full-resolution hyperspectral image.

Figure 3 illustrates results of this process for encoder configurations ranging from 2×2 to 5×5 . Figure 3a compares the spatial interpolation performance for a single band (monochrome) image sampled with each encoder configuration. For each image, we report the

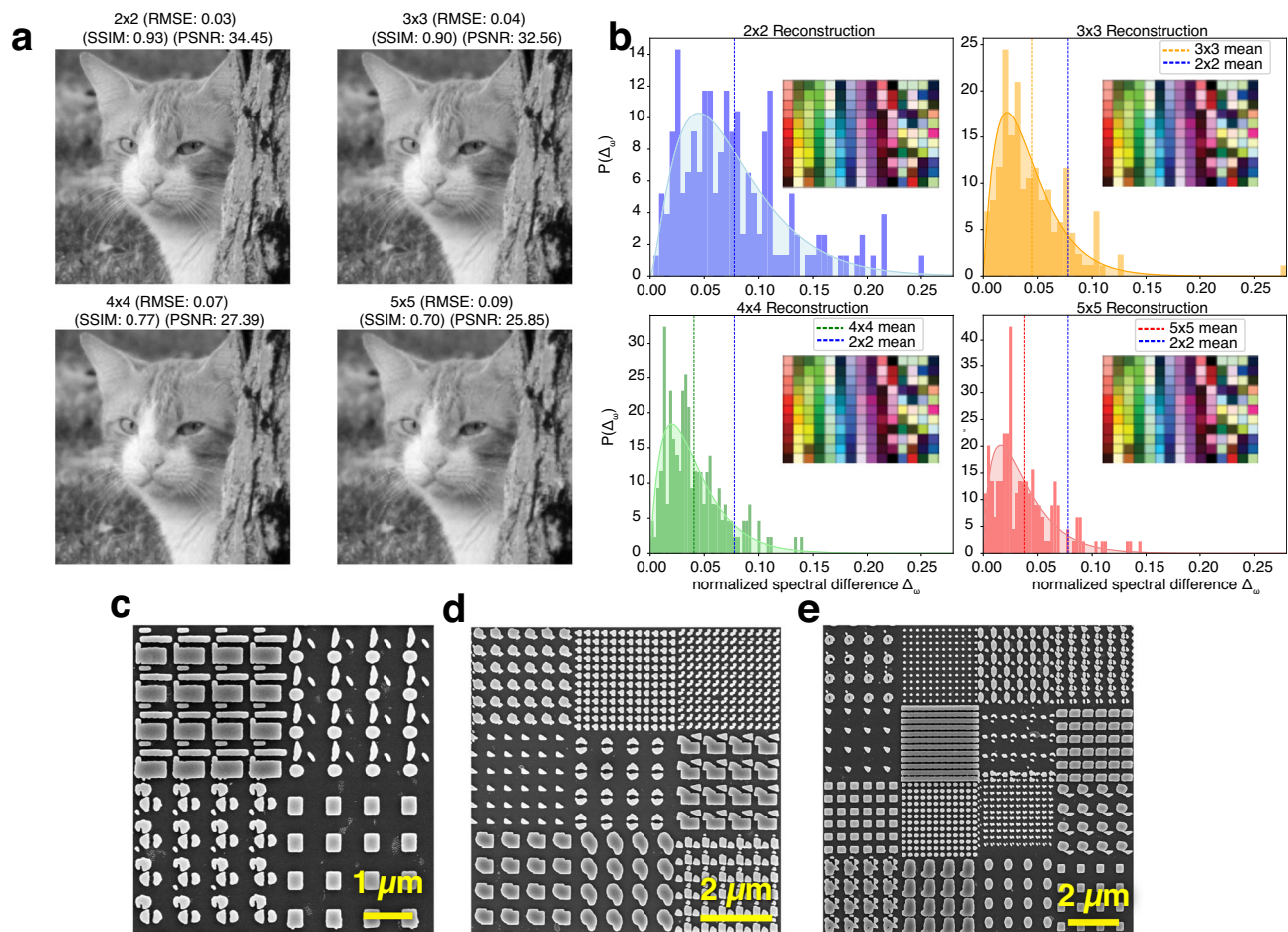


Fig. 3 | Comparison of different encoder numbers. a Spatial interpolation comparison for a single band image using 2×2 – 5×5 encoder configurations. **b** Spectral reconstruction performance comparison for 2×2 – 5×5 encoder configurations, the insets show the color palette for which each patch's spectra was reconstructed,

spatially interpolated as in (a). **(c–e)** Scanning electron microscope images of experimental realizations of 2×2 to 4×4 encoder configurations. Source data are provided as Fig. 3 data.

normalized root mean square error (NRMSE), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR) of the interpolation compared to the original graphic.

Figure 3b compares the spectral performance of 2×2 – 5×5 encoder configurations. The histograms display the distribution of the spectral reconstruction error for the spectra of each color panel of the color calibration palette inset in each plot. As in panel Fig. 3a, the palette image is interpolated according to the number of encoders. The results show a trade-off between spatial and spectral information quality for different encoder numbers. The mean spectral reconstruction difference decreases linearly as the number of encoders increases from 2×2 to 4×4 , with each additional encoder decreasing the difference by -0.0032 . However, the trend stops at the 4×4 – 5×5 transition, as the effect of adding nine additional encoders is less significant than adding one encoder in configurations below 4×4 .

Similarly, the RMSE error for spatial information increases linearly with the number of encoders. For 2×2 – 4×4 , each additional encoder increases the error by -0.0068 until the 4×4 – 5×5 transition shows a decreased benefit from adding more encoders. Overall, the total spatial reconstruction accuracy regarding NRMSE remains between 0.1 and 0.17 for all encoder configurations. In contrast, the mean spectral reconstruction error ranges between 0.36 and 0.76.

Figure 3c–e present scanning electron microscope (SEM) images of experimental realizations of 2×2 , 3×3 , and 4×4 encoder configurations, respectively. Supplementary Note 3 and Supplementary Fig. 3 discuss the hardware encoder integration with the camera and

the performance of 2×2 – 5×5 encoder configurations on classification and segmentation tasks. Table 1 summarizes the results.

Figure 4 presents experimental results of the hardware-accelerated hyperspectral platform in field applications. Figure 4a validates the reconstruction of a single video frame by using a calibration palette composed of various colors with known reflection spectra. The test compares two hyperspectral images with 204 frequency bands each, one captured with our hardware-accelerated camera with 12 Megapixels, $N_k=9$ trained encoders on a publicly available general hyperspectral dataset³⁴, and a video rate of 30 FPS, and the other obtained with a commercial SPECIM IQ (Specim, Spectral Imaging Ltd.) hyperspectral camera, which possesses 0.26 Megapixels and a 0.016 FPS acquisition rate. We acquired both images outdoors under direct sunlight illumination, with the SPECIM IQ mounted on a tripod to avoid motion blur from the low framerate. The solid area in Fig. 4a shows the distribution $P(\Delta_\omega)$ of the absolute spectral difference Δ_ω between the data retrieved with the two cameras. The hardware accelerated platform, while working at an acquisition rate ~ 2000 times faster, provides the same spectral prediction for the same number of bands, with an average reconstruction difference below 3%. The insets in Fig. 4a show reconstructed RGB images of the palette from hyperspectral data.

Figure 4b–g illustrate a field application for real-time video understanding using an Unmanned Aerial Vehicle (UAV). We integrate the hardware-accelerated platform of Fig. 2e on a Matrice 300 RTK drone (DJI). Figure 4b shows a photograph of the drone in flight, with

Table 1 | Performance metrics for different encoder configurations

Metric	Encoder configuration	Value
NRMSE (Set12)	2 × 2	0.097
	3 × 3	0.119
	4 × 4	0.170
	5 × 5	0.162
PSNR (dB) (Set12)	2 × 2	24.76
	3 × 3	23.12
	4 × 4	19.70
	5 × 5	20.34
SSIM (Set12)	2 × 2	0.82
	3 × 3	0.76
	4 × 4	0.61
	5 × 5	0.63
Classification Accuracy (Hyperspectral Flowers)	2 × 2	94.1%
	3 × 3	94.5%
	4 × 4	92.3%
	5 × 5	90.2%
mIoU (FVGNET)	2 × 2	0.71
	3 × 3	0.69
	4 × 4	0.66
	5 × 5	0.63

the inset showing a closeup of the hardware-accelerated hyperspectral camera. Supplementary Note 4 and Supplementary Fig. 4 present additional details of this component. We use the drone to record a total of 80 min of aerial hyperspectral video footage at 30 frames per second, capturing the terrain from an altitude of 50 m. Figure 4c displays a selection of raw frames from the original hyperspectral footage. Figure 4d shows an extracted hyperspectral video frame, while Fig. 4e illustrates the reflection spectra associated with one spatial point (panel d, yellow dot *p*). Figure 4f shows a complete 3D hyperspectral map representing a 0.5 km² area of the terrain obtained by processing the hyperspectral aerial footage. We compute the map using Pix4DMapper⁶⁸, a photogrammetry software that stitches video frames into 3D models of surveyed locations. We implement a K-means clustering algorithm to analyze the hyperspectral data further, segmenting it into four distinct clusters based on depth and spectral dimensions (Fig. 4g). We visualize these clusters by plotting each pixel of the 3D hyperspectral map with the corresponding spectral color in panel f. The red pixel cluster in the map represents high-elevation areas of the buildings hosting metallic heating, ventilation, and air conditioning (HVAC) units. Light violet pixels correspond to rooftop surfaces, which the clustering algorithm segments as the strongest light-reflecting scene objects. Yellow pixel clusters map concrete, which appears as the buildings' walls and the building's rooftop in the lower right area where dust has accumulated. The blue pixels combine information about objects above the ground that predominantly reflect green light, marking the signature of chlorophyll and indicating specific species of vegetation. Finally, green pixels map the ground plane.

Hyperspectral video segmentation and tracking

Video Object Segmentation (VOS) in AI video understanding aims to classify and monitor target objects distinct from the background across video frames over time. Figure 5 illustrates the two principal methodologies in VOS. One-Shot Video Object Segmentation (OVOS)⁶⁹ uses manually labeled reference frames to instruct the segmentation algorithm on the initial composition of targets (Fig. 5a, b). This technique is semi-supervised and necessitates human input to specify the

objects of interest. Target labeling uses either a segmentation map (Fig. 5b, cyan and yellow transparencies) or a bounding box for visual object tracking (Fig. 5b, cyan and yellow rectangles). Zero-Shot Video Object Segmentation (ZVOS), conversely, autonomously processes objects with different visual characteristics without human-defined labels (Fig. 5c).

The main limitation of current VOS processes is that AI cannot segment information that cameras cannot capture, that is, light beyond RGB colors. The hardware-accelerated hyperspectral platform introduced in this work addresses this problem by empowering AI with spectral features, which provide more comprehensive information than primary colors. In contrast to methods where the hypercube is recorded and fed to AI models afterwards, we obtain spectral features from a scene in real-time thanks to the optical processing of the hardware encoder array. This approach allows carrying out VOS on live video, or at a notable speedup in recordings compared to working with hypercubes.

Figure 6a illustrates the configuration of the hardware-accelerated video understanding platform for one-shot video semantic segmentation. In this system, the spectral features arising from the hardware encoders enter a motion encoder E_m comprising five integral modules that incorporate state-of-the-art spatial-time network models^{70–72}. The first module, the query key encoder (Fig. 6a, E_q^k unit), extracts spectral-spatial image features k^Q , which the query-memory projection (Fig. 6a, QMP) processes. The QMP computes similarities between the k^Q features and the spectral-spatial features k^M arising from previous frames extracted by the memory key encoder (Fig. 6a, E_m^k unit). The QMP evaluates the degree of affinity via a similarity matrix $\mathbb{W} \in \mathbb{R}^{HW \times HW}$:

$$\mathbb{W}(k^Q, k^M)_{ij} = \frac{\exp(k_i^Q \odot k_j^M)}{\sum_j \exp(k_i^Q \odot k_j^M)}, \quad (2)$$

where the \odot operator is the dot product. The matrix entries \mathbb{W}_{ij} furnish a similarity score between the input features k^Q and k^M ranging between zero and one. The mask adjustment module (Fig. 6a, MAM) processes the data by projecting the video mask v^M , computed from previous frames by the memory value encoder (Fig. 6a, E_m^v unit), to the QMP output using the following similarity matrix \mathbb{W} :

$$v^Q = \mathbb{W}(k^Q, k^M)v^M. \quad (3)$$

The mask adjustment module output represents the feature tensor $\tilde{\mathbf{F}}^f$ for the current timestep t (Fig. 2i). Both key and value encoders are implemented with the ResNet architecture⁷³, using ResNet50 and ResNet18 respectively. Adhering to the methodology outlined in the STM practice⁷⁰, we utilize res4 features with a stride of 16 from the foundational ResNets as our principal backbone features, while omitting res5. We employ a 3×3 convolutional layer, without nonlinearity as a projection mechanism from the backbone feature towards either the key space (dimensionality of $C_k = 128$) or the value space (dimensionality of $C_v = 512$). The output decoder further processes $\tilde{\mathbf{F}}^f$ to predict the mask for the frame at timestep t , and output the frame itself. We utilize the refinement module from⁷⁴ as the core component of our decoder. Initially, the read output is condensed to 256 channels through a convolutional layer and a residual block⁷⁵. Subsequently, several refinement modules sequentially double the size of the compressed feature map. Each stage of refinement incorporates the previous stage's output and a corresponding scale feature map from the query encoder via skip connections. The final refinement block's output undergoes reconstruction into the object mask with a concluding convolutional layer, succeeded by a softmax function. All convolutional layers within the decoder apply 3×3 filters, generating an output of 256 channels, except the ultimate layer, which yields a 2-channel

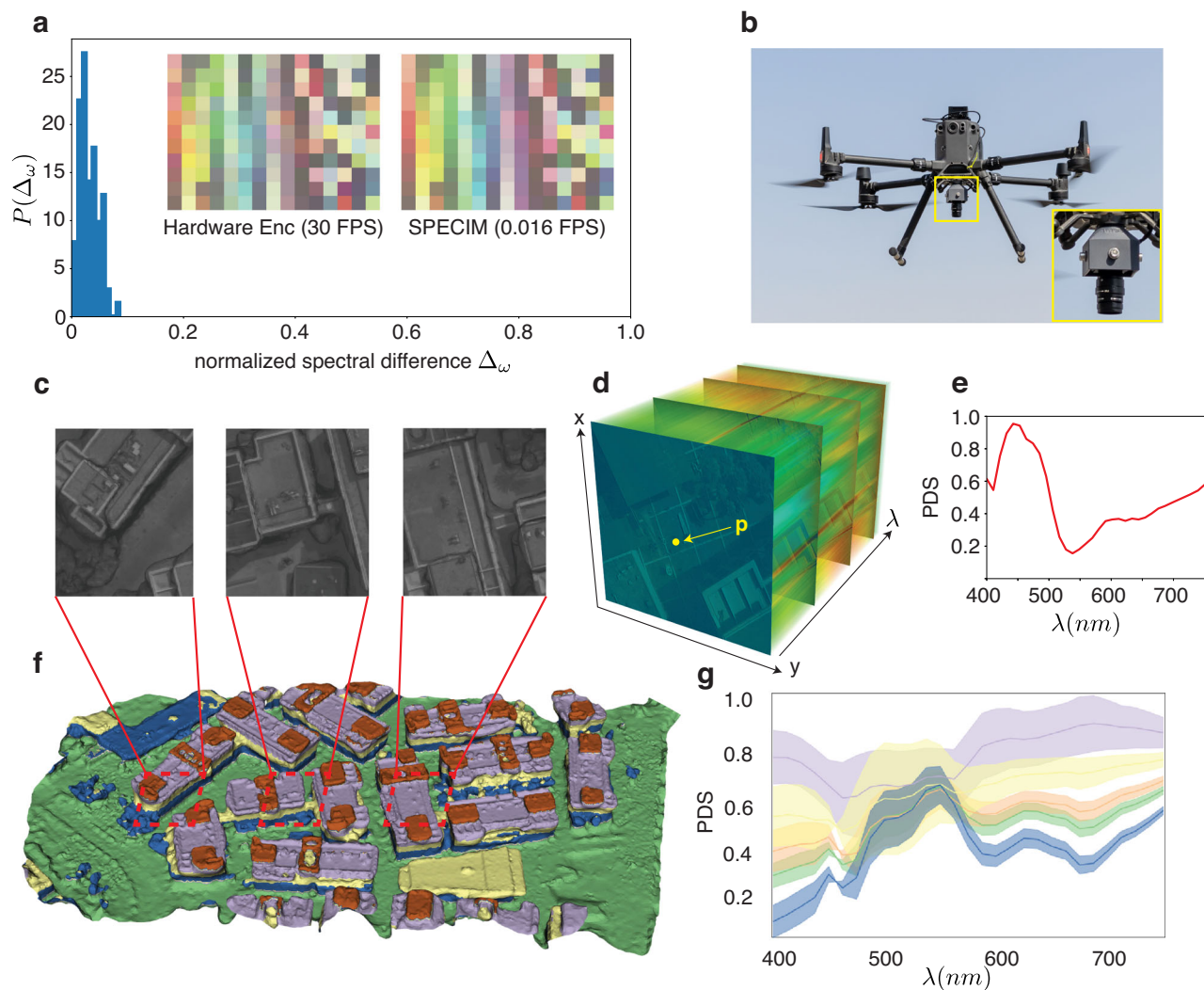


Fig. 4 | Real-time hyperspectral UAV scene reconstruction. **a** Distribution of spectral reconstruction difference for a reconstructed hyperspectral frame image (inset) at 204 bands using the platform of Fig. 2 at 30 FPS and a commercial SPECIM IQ operating at 0.016 FPS. **b** Camera module with hardware encoders integrated on DJI Matrice 300 RTK drone, the inset shows a closeup of the camera. **c** Raw frames

in the hyperspectral UAV video sequence. **d** Visualization of a single hyperspectral frame acquired by the drone at 30 FPS. **e** Power density spectrum (PDS) retrieved at the spatial pixel p in panel d. **f** Reconstructed three-dimensional hyperspectral map from the video sequences, with spectral distribution visualized via K-MEANS clustering. **g** Cluster's spectral distribution. Source data are provided as Fig. 4 data.

output. The decoder predicts the mask at a quarter scale of the original input image. The mask and frame then form feedback for the motion encoder's successive predictions.

Figure 6b–e presents the results on OVOS using the hardware-accelerated platform of this work operating with 204 bands (OVOS204) applied to two distinct segmentation tasks. The first task (Fig. 6b, c) uses hyperspectral video data acquired with the UAV of Fig. 4 to track and segment the hyperspectral signature characterizing a specific car from many with the same visual color appearance in the aerial footage (Fig. 6b). Accessing hyperspectral signatures allows the user to find and correctly label objects that color cameras cannot distinguish due to a lack of information. Figure 6c shows how the system successfully tracks the target's position over time. The images in Fig. 6c depict the data evolution directly from the spectral feature data stream received by the camera, which contains the features motion encoder will process. In the second task (Fig. 6d), we mounted the hardware accelerated camera inside a car and performed the same type of hyperspectral tracking, segmenting spectral signatures corresponding to specific vehicles (Fig. 6d, e).

Figure 7 showcases additional ZVOS examples of how hyperspectral video flows could empower AI understanding. Figure 7a, b

show the system configuration for ZVOS. A query encoder module (Fig. 7a, E_q^k unit) extracts spectral-spatial features k^Q using the same encoder architectures as in the OVOS task. A query memory correlation module processes these features nonlinearly (Fig. 7a, b, QMCM). The QMCM understands dense spatial relationships in the input frame features by using the correlation matrix \mathbb{W}_{corr} , defined as:

$$\mathbb{W}_{\text{corr}}(k^M, k^Q) = \frac{1}{C_k} \text{softmax}[k^M (k^Q)^\dagger], \quad (4)$$

to project the query key feature vector k^Q :

$$v_{\text{mem}}^Q = \mathbb{W}_{\text{corr}}(k^M, k^Q) k^Q. \quad (5)$$

The top QMCM understands relationships between present k^Q , and past k^M query features arising from the memory encoder E_m^k . The bottom QMCM understands the correspondence of k^Q features with themselves. The motion encoder E_m concatenates the output from both QMCMs with the initial features k^Q . The decoder uses the same architecture as in the OVOS scheme, outputting the predicted mask

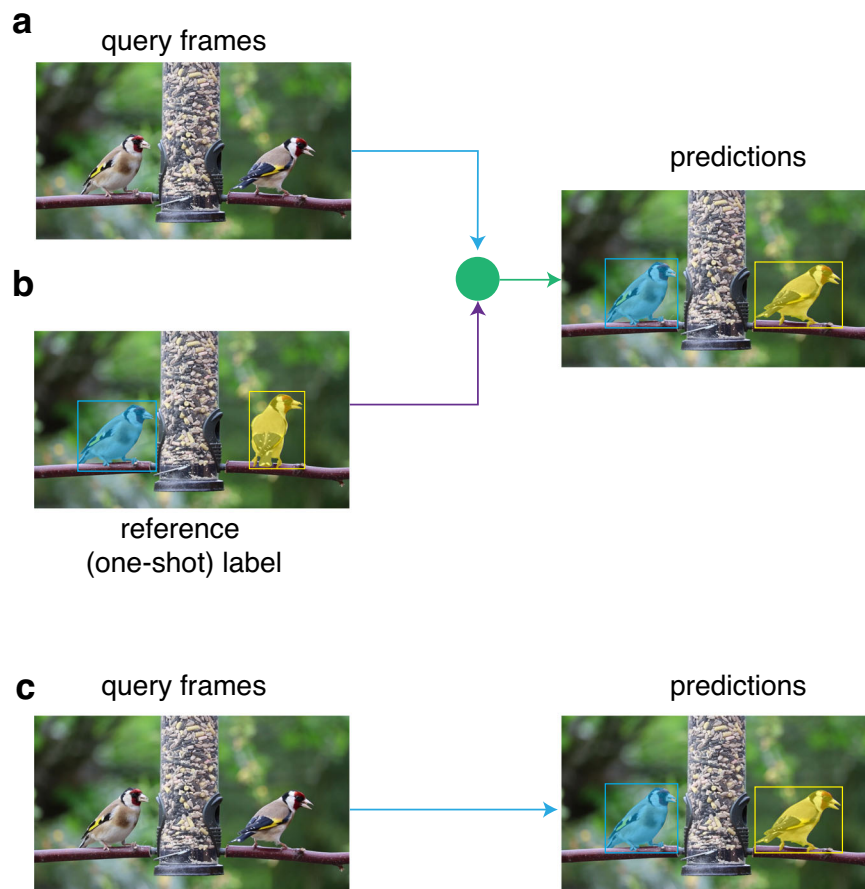


Fig. 5 | Approaches to AI video segmentation and tracking. **a, b** One-shot video object segmentation (OVOS) and **(c)** zero-shot video object segmentation (ZVOS).

alongside the frame, which becomes feedback for E_m as a memory frame R' .

We benchmark the hardware-accelerated platform on a hyperspectral video dataset (FVgNET-video) built from the public FVgNET dataset⁶⁶. FVgNET-video consists of 30 FPS hyperspectral video sequences of combinations of artificial and natural fruits and vegetables placed on a rotating turntable (Fig. 7c, d). Figure 7c–h illustrate the performance of hardware-accelerated hyperspectral ZVOS on samples from the dataset. We compare the performance of the hyperspectral camera with a simulated RGB camera that records at the same resolution and frame rate. Figure 7c, d show segmentation masks generated in real-time from a video sequence showing two grapes, one banana, one orange, and one potato on the turntable. One of the grape bundles is artificial, while the rest are natural. Figure 7c shows the result of segmentation masks created on RGB data, with each color marking a different object class. These images show that the RGB camera cannot distinguish between artificial and natural objects, predicting that both grapes are of the same type. Figure 7d shows the segmentation masks resulting from hyperspectral data, allowing AI to identify all items correctly. Figure 7e presents the CIE 1931 chromaticity diagram distribution of the RGB values for the artificial and natural grape bundles of the video sequence. The panel shows there is little chromaticity variation between these samples, causing the RGB VOS to fail. Figure 7d shows the reflection spectra for each grape bundle. In contrast to panel e, there is significant variation in the spectral response of the two objects, explaining the success of the hyperspectral VOS. Figure 7g, h quantify this performance difference further by showing the confusion matrices for the RGB and hyperspectral VOS tasks, respectively. In the RGB case, the segmentation fails in two of the eight categories, with a substantial number of incorrect pixels for

artificial oranges and real grapes. There is also significant confusion in the case of artificial grapes, with over 40% of all pixels classified incorrectly. Conversely, the segmentation is successful for all categories in the hyperspectral case, with <5% of the pixels incorrectly classified.

Discussion

This work implemented and field-validated a hardware-accelerated platform for real-time hyperspectral video understanding, demonstrating hyperspectral UAV scene reconstruction, hyperspectral video object segmentation, and classification using >200 frequency bands at 12-megapixel spatial resolution, and video rates of 30 FPS. This platform technology processes information beyond 1 Tb/s, enabling the current generation of AI to understand information that color video acquisition systems do not discern, and current hyperspectral imaging technologies cannot acquire in real-time and at these resolutions. This work opens research and application opportunities for AI video understanding utilizing broadband hyperspectral data flows for environmental monitoring, security, pharmaceutical, mining, and medical diagnostics that require processing high-resolution spectral and spatial information at video rates. Future research can focus on developing scalable systems that benefit both the adoption of this technology and the ease of acquiring real-world hyperspectral data flows for subsequent AI development. The principles and methodologies devised in this work can also be generalized across various fields and help impact the way AI interacts with the visual world, particularly for foundation models like GPT-4 and Claude2. The technology we have introduced could empower this generation's AI to apply its information processing capacities to a broader set of multimodal tasks, facilitating the

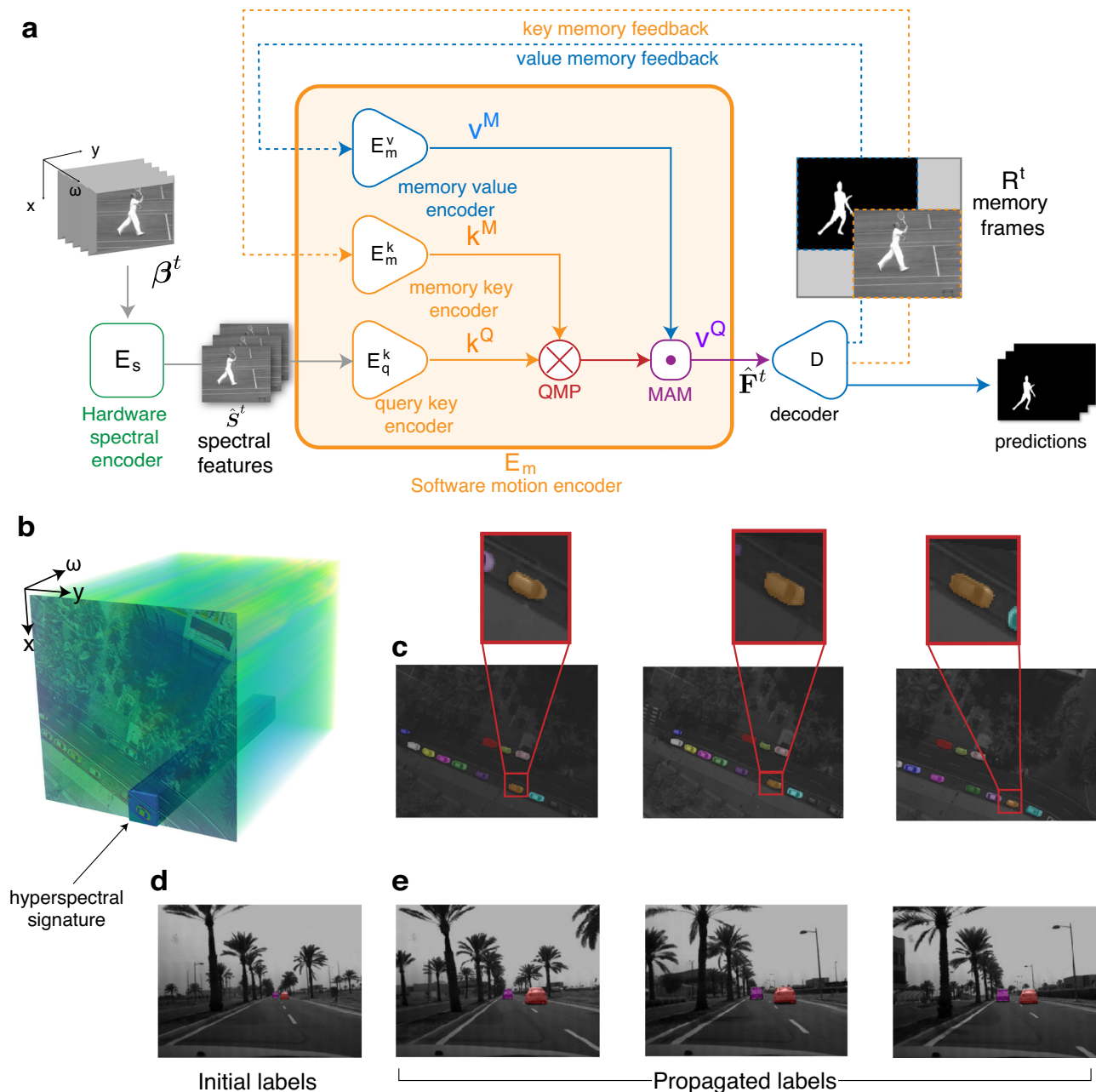


Fig. 6 | One-shot hyperspectral segmentation and object tracking.

a Configuration of the general architecture of Fig. 2 for one-shot video semantic segmentation. **b, c** Segmentation results on hyperspectral object tracking with (b) Hyperspectral signature selected from the datacube acquired by the UAV of Fig. 4.

c Segmentation results in real-time with detailed zoomed regions. **d** Video recording from inside a car and initially selected signatures. **e** A segment of hyperspectral data with signatures that the architecture tracks and propagates over time.

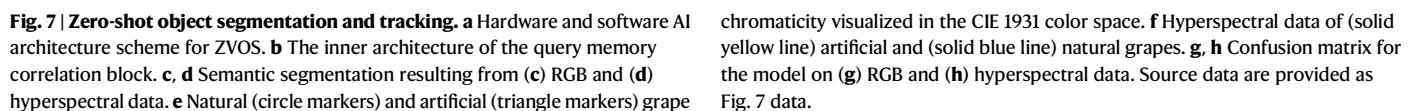
pursuit of more robust forms of AI, such as Artificial General Intelligence⁷⁶.

Methods

Hardware encoder nanofabrication

We start with a 15 mm wide and 500 μm thick square piece of fused silica (University Wafer) as our substrate. To ensure a clean surface, we sonicate the substrate in isopropyl alcohol at 25 KHz for 5 min, followed by additional sonication at 45 KHz for 5 min. We then grow a 200 nm thick layer of hydrogenated amorphous silicon (a-Si:H) using plasma-enhanced chemical vapor deposition (PECVD). Afterward, we spin coat a positive electron beam resist layer on the sample. Because of their complex geometry, exposing free-form nanostructures

requires conducting proximity effect correction prior to their writing. We conduct shape-based proximity effect correction by simulating the energy absorption of the nine encoders' design, as shown in Fig. 2b, on the resist using the software package BEAMER (GenISys GmbH). We then manually deform the shapes of each of the nine designs so that, upon exposure, the shape of the energy absorption patches on the resist matches that of the resonator designs. We print the nanostructures pattern using a JEOL JBX-6300FS electron beam lithography system at 100 kV accelerating voltage. Following this, we develop the resist and perform a liftoff process, creating a hard mask with the shape of the nanostructures on the silicon. Finally, we use reactive ion etching (RIE) to completely remove the unprotected silicon and remove the hard mask.



It is possible to train the encoders response functions $\hat{\lambda}_k(\omega)$ using both linear and nonlinear feature extraction schemes. In both cases, the starting point is to flatten the hyperspectral tensor into a single matrix containing the power density spectra of a set of camera pixels on each column, creating a dataset β for subsequent training. To provide feature extractions to this data, we can use any supervised or unsupervised techniques developed in deep learning to train the relevant distributions of coefficient $\Lambda(\omega)$. Once these distributions of values are found, we use ALFRED^{64,65}, an advanced optimization framework for the design and implementation of nanoresonators with

user-defined broadband responses across the visible and infrared. This article uses a linear encoder \mathbf{A} developed through an unsupervised learning approach utilizing Principal Component Analysis (PCA). This process entails hardware encoding E_s , in which we specifically selected the nine strongest principal components, denoted as $\hat{\mathbf{A}}^\dagger$, following the singular value decomposition of data tensor $\boldsymbol{\beta}$. This methodology is not limited to PCA and could be easily extended into a supervised learning framework. In order to do so, we refine the encoder's training process by incorporating differentiable spectra projector modules^{64,66}, facilitating an iterative refinement where the model learns to adjust the distributions of $\mathcal{I}(\omega)$ in order to find the

minimum loss $L(\beta^t, \hat{\beta}^t)$ in the iterative end-to-end optimization. This supervised framework enables a more targeted optimization of the encoders, ensuring that the output not only captures the principal variances within the dataset but also aligns closely with final application goals.

While it is possible to train the hardware encoder to the specific problem, due to the absence of hyperspectral video training datasets, we here use in every application the same set of encoders trained from a publicly available dataset of general hyperspectral images under various illumination conditions⁵⁴.

Data availability

Source data are provided with this paper.

Code availability

The code ALFRED used in this work is available at <https://github.com/makamoa/alfred>.

References

- Wu, C., Wu, F., Lyu, L., Huang, Y. & Xie, X. Communication-efficient federated learning via knowledge distillation. *Nat. Commun.* **13**, 2032 (2022).
- Wu, C. et al. A federated graph neural network framework for privacy-preserving personalization. *Nat. Commun.* **13**, 3091 (2022).
- Dayan, I. et al. Federated learning for predicting clinical outcomes in patients with covid-19. *Nat. Med.* **27**, 1735–1743 (2021).
- Afromowitz, M., Callis, J., Heimbach, D., DeSoto, L. & Norton, M. Multispectral imaging of burn wounds: a new clinical instrument for evaluating burn depth. *IEEE Trans. Biom. Eng.* **35**, 842–850 (1988).
- Panasyuk, S. V. et al. Medical hyperspectral imaging to facilitate residual tumor identification during surgery. *Cancer Biol. Therapy* **6**, 439–446 (2007).
- Guo, X. et al. Smartphone-based dna diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security. *Nat. Electron.* **4**, 615–624 (2021).
- Martini, G. et al. Machine learning can guide food security efforts when primary data are not available. *Nat. Food* **3**, 716–728 (2022).
- Yin, X. & Müller, R. Integration of deep learning and soft robotics for a biomimetic approach to nonlinear sensing. *Nat. Mach. Intell.* **3**, 507–512 (2021).
- Goddard, M. A. et al. A global horizon scan of the future impacts of robotics and autonomous systems on urban ecosystems. *Nat. Ecol. Evol.* **5**, 219–230 (2021).
- Soenksen, L. R. et al. Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. *Sci. Transl. Med.* **13**, eabb3652 (2021).
- de Croon, G. C. H. E., De Wagter, C. & Seidl, T. Enhancing optical-flow-based control by learning visual appearance cues for flying robots. *Nat. Mach. Intell.* **3**, 33–41 (2021).
- Pang, B. et al. Complex sequential understanding through the awareness of spatial and temporal concepts. *Nat. Mach. Intell.* **2**, 245–253 (2020).
- Hirschberg, J. & Manning, C. D. Advances in natural language processing. *Science* **349**, 261–266 (2015).
- OpenAI. Gpt-4 Technical Report. <https://cdn.openai.com/papers/gpt-4.pdf> (2023).
- Brown, T. et al. Language Models are Few-Shot Learners. in *Advances in Neural Information Processing Systems*, vol. 33, 1877–1901 (Curran Associates, Inc., 2020).
- Reed, S. et al. A generalist agent. *arXiv* <https://doi.org/10.48550/arXiv.2205.06175> (2022).
- Lu, J., Clark, C., Zellers, R., Mottaghi, R. & Kembhavi, A. Unified-io: A unified model for vision, language, and multi-modal tasks. *arXiv* <https://doi.org/10.48550/arXiv.2206.08916> (2022).
- Aghajanyan, A. et al. Cm3: A causal masked multimodal model of the internet. *arXiv* <https://doi.org/10.48550/arXiv.2201.07520> (2022).
- Moor, M. et al. Foundation models for generalist medical artificial intelligence. *Nature* **616**, 259–265 (2023).
- Makarenko, M., Wang, Q., Burguete-Lopez, A. & Fratalocchi, A. Photonic optical accelerators: the future engine for the era of modern AI? *APL Photonics* **8**, 110902 (2023).
- Zhao, Y., Misra, I., Krähenbühl, P. & Girdhar, R. Learning video representations from large language models. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 6586–6597 (2023).
- Buch, S. et al. Revisiting the “video” in video-language understanding. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2917–2927 (2022).
- Wang, J. et al. Omnivl: One foundation model for image-language and video-language tasks. In *Advances in Neural Information Processing Systems*. (eds. Koyejo, S. et al.) 5696–5710 (Curran Associates, Inc., 2022).
- Yu, N. & Capasso, F. Flat optics with designer metasurfaces. *Nat. Mater.* **13**, 139–150 (2014).
- Chen, W. T. et al. A broadband achromatic metalens for focusing and imaging in the visible. *Nat. Nanotechnol.* **13**, 220–226 (2018).
- Burguete Lopez, A., Fratalocchi, A., Getman, F., Makarenko, M. & Wang, Q. *Hyperspectral Imaging Apparatus and Methods*. <https://patents.google.com/patent/WO2023084401A1/en?inventor=Arturo+BURGUETE+LOPEZ> (2023).
- Kim, I. et al. Metasurfaces-driven hyperspectral imaging via multiplexed plasmonic resonance energy transfer. *Adv. Mater.* **35**, 2300229 (2023).
- Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K. & Abbas, A. Modern trends in hyperspectral image analysis: a review. *IEEE Access* **6**, 14118–14129 (2018).
- Moncrieff, M., Cotton, S., Claridge, E. & Hall, P. Spectrophotometric intracutaneous analysis: a new technique for imaging pigmented skin lesions. *British J. Dermatol.* **146**, 448–457 (2002).
- Johansen, T. H. et al. Recent advances in hyperspectral imaging for melanoma detection. *WIREs Comput. Stat.* **12**, e1465 (2020).
- Chennu, A., Färber, P., De’ath, G., de Beer, D. & Fabricius, K. E. A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Sci. Rep.* **7**, 7122 (2017).
- Dumke, I. et al. Underwater hyperspectral imaging as an in situ taxonomic tool for deep-sea megafauna. *Sci. Rep.* **8**, 12860 (2018).
- International Telecommunication Union—Radiocommunication Sector. Recommendation ITU-R BT.2020-2 (10/2015) Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange. Technical Recommendation BT.2020-2 https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2020-2-201510-!-!PDF-E.pdf (2015).
- Nguyen, T.-U., Pierce, M. C., Higgins, L. & Tkaczyk, T. S. Snapshot 3D optical coherence tomography system using image mapping spectrometry. *Opt. Express* **21**, 13758–13772 (2013).
- Cubert Ultris X20 Plus. <https://www.cubert-hyperspectral.com/products/ultris-x20-plus>
- XIMEA. Hyperspectral Snapshot USB3 Camera 16 Bands 460–600 nm. <https://www.ximea.com/en/products/hyperspectral-cameras-based-on-usb3-xispec/mq022hg-im-sm4x4-vis> (2024).
- Photonfocus. MV4-D2048x1088-C01-HS02-GT Hyperspectral Camera. <https://www.photonfocus.com/products/camerafinder/camera/mv4-d2048x1088-c01-hs02-gt/> (2024).

38. Imec. SNAPSHOT UAV VIS+NIR Hyperspectral Camera. <https://www.imechyperspectral.com/en/cameras/snapshot-uav-vis-nir> (2024).
39. Dwight, J. G. et al. Compact snapshot image mapping spectrometer for unmanned aerial vehicle hyperspectral imaging. *J. Appl. Remote Sens.* **12**, 044004 (2018).
40. Pawlowski, M. E., Dwight, J. G., Nguyen, T.-U. & Tkaczyk, T. S. High performance image mapping spectrometer (IMS) for snapshot hyperspectral imaging applications. *Opt. Express* **27**, 1597–1612 (2019).
41. Amann, S., Haist, T., Gatto, A., Kamm, M. & Herkommer, A. Design and realization of a miniaturized high resolution computed tomography imaging spectrometer. *J. Eur. Opt. Soc. Rapid Publ.* **19**, 34 (2023).
42. Kester, R. T., Bedard, N., Gao, L. & Tkaczyk, T. S. Real-time snapshot hyperspectral imaging endoscope. *J. Biomed. Opt.* **16**, 056005 (2011).
43. Corning. microHSI™ 425 Sensor and microHSI™ 425 SHARK.pdf. https://www.corning.com/media/worldwide/csm/documents/microHSI_425_Sensor_and_microHSI_425_SHARK.pdf (2024).
44. Headwall. MV.C VNIR Imaging System. <https://www.headwallphotonics.com/knowledge-center/product/pdf/mvc-vnir-sensor> (2024).
45. Resonon Pika XC2. Hyperspectral Imaging Cameras - Resonon. <https://resonon.com/Pika-XC2> (2024).
46. HySpex. VNIR-3000 N. <https://www.hyspex.com/hyspex-products/hyspex-classic/hyspex-vnir-3000-n/> (2024).
47. Specim. Specim IQ. <https://www.specim.com/iq/> (2024).
48. JEDEC. Solid State Technology Association. DDR5 SDRAM | JEDEC. Standard JESD79-5B <https://www.jedec.org/standards-documents/docs/jesd79-5b> (2022).
49. Bioucas-Dias, J. M. et al. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **1**, 6–36 (2013).
50. Adão, T. et al. Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens.* **9**, 1110 (2017).
51. Eckhard, J., Eckhard, T., Valero, E. M., Nieves, J. L. & Contreras, E. G. Outdoor scene reflectance measurements using a bragg-grating-based hyperspectral imager. *Appl. Opt.* **54**, D15–D24 (2015).
52. Glatt, O. et al. Beyond RGB: a real world dataset for multispectral imaging in mobile devices. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 4332–4342 (Waikoloa, HI, 2024).
53. Catelli, E. et al. Can hyperspectral imaging be used to map corrosion products on outdoor bronze sculptures? *J. Spectral Imaging* **7**, a10 (2018).
54. Li, Y., Fu, Q. & Heidrich, W. Multispectral illumination estimation using deep unrolling network. In *2021 IEEE Int. Conf. Computer Vision (ICCV)*. 2652–2661 (IEEE, 2021).
55. Lu, G. & Fei, B. Medical hyperspectral imaging: a review. *J. Biomed. Opt.* **19**, 010901 (2014).
56. Antonik, P., Marsal, N., Brunner, D. & Rontani, D. Human action recognition with a large-scale brain-inspired photonic computer. *Nat. Mach. Intell.* **1**, 530–537 (2019).
57. Sekh, A. A. et al. Physics-based machine learning for subcellular segmentation in living cells. *Nat. Mach. Intell.* **3**, 1071–1080 (2021).
58. Ma, F. et al. Unified transformer tracker for object tracking. In *Proceedings of the IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*. 8781–8790 (2022).
59. Gowen, A. A., Feng, Y., Gaston, E. & Valdramidis, V. Recent applications of hyperspectral imaging in microbiology. *Talanta* **137**, 43–54 (2015).
60. Mengu, D., Tabassum, A., Jarrahi, M. & Ozcan, A. Snapshot multispectral imaging using a diffractive optical network. *Light Sci. Appl.* **12**, 86 (2023).
61. Li, J. et al. Spectrally encoded single-pixel machine vision using diffractive networks. *Sci. Adv.* **7**, eabd7690 (2021).
62. Choi, E. et al. Neural 360° structured light with learned meta-surfaces. *arXiv* <https://doi.org/10.48550/arXiv.2306.13361> (2023).
63. Pierangeli, D., Marcucci, G. & Conti, C. Photonic extreme learning machine by free-space optical propagation. *Photon. Res.* **9**, 1446–1454 (2021).
64. Makarenko, M., Burguete-Lopez, A., Getman, F. & Fratalocchi, A. Robust and scalable flat-optics on flexible substrates via evolutionary neural networks. *Adv. Intell. Syst.* **3**, 2100105 (2021).
65. Getman, F., Makarenko, M., Burguete-Lopez, A. & Fratalocchi, A. Broadband vectorial ultrathin optics with experimental efficiency up to 99% in the visible region via universal approximators. *Light Sci. Appl.* **10**, 1–14 (2021).
66. Makarenko, M. et al. Real-time hyperspectral imaging in hardware via trained metasurface encoders. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12692–12702 (IEEE, 2022).
67. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Transact. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
68. PIX4D. PIX4Dmapper: Professional Photogrammetry Software for Drone Mapping. <https://www.pix4d.com/product/pix4dmapper-photogrammetry-software> (2024).
69. Gao, M. et al. Deep learning for video object segmentation: a review. *Artif. Intell. Rev.* **56**, 457–531 (2023).
70. Oh, S. W., Lee, J.-Y., Xu, N. & Kim, S. J. Video Object Segmentation Using Space-Time Memory Networks. *arXiv* <https://doi.org/10.48550/arXiv.1904.00607> (2019).
71. Cheng, H. K., Tai, Y.-W. & Tang, C.-K. Rethinking space-time networks with improved memory coverage for efficient video object segmentation. *arXiv* <https://doi.org/10.48550/arXiv.2106.05210> (2021).
72. Park, K., Woo, S., Oh, S. W., Kweon, I. S. & Lee, J.-Y. Per-clip video object segmentation. *arXiv* <https://doi.org/10.48550/arXiv.2208.01924> (2022).
73. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778 (IEEE, 2015).
74. Oh, S. W., Lee, J.-Y., Sunkavalli, K. & Kim, S. J. Fast video object segmentation by reference-guided mask propagation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2018).
75. He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. *arXiv* <https://doi.org/10.48550/arXiv.1603.05027> (2016).
76. Shevlin, H., Vold, K., Crosby, M. & Halina, M. The limits of machine intelligence: despite progress in machine intelligence, artificial general intelligence is still a major challenge. *EMBO Rep.* **20**, e49177 (2019).

Acknowledgements

We carried out the hardware encoder's fabrication, scanning electron microscope analysis, and the fabrication of the drone integration hardware at KAUST's Nanofabrication, Imaging and Characterization, and Prototyping and Product Development Core Lab's respectively. We thank Falvonvz's staff for their assistance with the aerial video recording and KAUST's Core Labs staff for their help in troubleshooting the fabrication processes.

Author contributions

A.F. supervised the work. A.B.L. fabricated the hardware encoders, hardware-accelerated camera, UAV integration hardware, and conducted the ground-based outdoors data acquisition. Q.W. developed the software for the UAV-based data recording. M.M., Q.W., A.B.L. and

L.P. conducted the UAV-based aerial footage recording. A.B.L. and L.P. processed the UAV recordings into 3D maps. M.M., Q.W. and A.B.L. conducted the indoors data acquisition and car-based recordings. M.M. developed the framework for spectral reconstruction, classification, segmentation, and tracking analysis. M.M. designed and trained the hardware encoders. S.G. and B.G. provided feedback on the segmentation and tracking applications. M.M., Q.W., A.B.L. and A.F. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51406-6>.

Correspondence and requests for materials should be addressed to Andrea Fratalocchi.

Peer review information *Nature Communications* thanks Junsuk Rho and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024