APPLIED SCIENCES AND ENGINEERING

Spectrally encoded single-pixel machine vision using diffractive networks

Jingxi Li[1,2,3]*, Deniz Mengu[1,2,3]*, Nezih T. Yardimci[1,3], Yi Luo[1,2,3], Xurong Li[1,3], Muhammed Veli[1,2,3], Yair Rivenson[1,2,3], Mona Jarrahi[1,3], Aydogan Ozcan[1,2,3]†

We demonstrate optical networks composed of diffractive layers trained using deep learning to encode the spatial information of objects into the power spectrum of the diffracted light, which are used to classify objects with a single-pixel spectroscopic detector. Using a plasmonic nanoantenna-based detector, we experimentally validated this single-pixel machine vision framework at terahertz spectrum to optically classify the images of handwritten digits by detecting the spectral power of the diffracted light at ten distinct wavelengths, each representing one class/digit. We also coupled this diffractive network-based spectral encoding with a shallow electronic neural network, which was trained to rapidly reconstruct the images of handwritten digits based on solely the spectral power detected at these ten distinct wavelengths, demonstrating task-specific image decompression. This single-pixel machine vision framework can also be extended to other spectral-domain measurement systems to enable new 3D imaging and sensing modalities integrated with diffractive network-based spectral encoding of information.

INTRODUCTION

Engineering of materials and material properties has opened a myriad of new opportunities for designing new components and devices with unique functionalities that were not possible before (*1*–*10*). Precise control of light-matter interaction at different scales has been the key behind the success of these engineered material systems, including, e.g., plasmonics, metamaterials, and photonic crystals, which led to various new capabilities for nanoscopic imaging and sensing as well as light generation, modulation, and detection (*11*–*20*). This quest to harness engineered light-matter interactions has also led to all-optical processors that perform a desired computational task through wave propagation within specially designed materials (*21*–*23*). All the way from solving equations to performing statistical inference and machine learning, these approaches highlight the emergence of engineered and trained matter as a building block of optical computation. Considering the rapid advances being made in, e.g., autonomous vehicles, robotic systems, and medical imaging, there is a growing need for performing computation optically to benefit from the low power, low latency, and scalability that passive optical systems can offer.

Here, we report deep learning–based design of diffractive networks that perform machine vision and statistical inference by encoding the spatial information of objects into optical spectrum through learnable diffractive layers that collectively process the information contained at multiple wavelengths to perform optical classification of objects using a single-pixel detector (Fig. 1A). Unlike conventional optical components used in machine vision systems, we use diffractive layers that are composed of two-dimensional (2D) arrays of passive pixels, where the complex-valued transmission or reflection coefficients of individual pixels are independent learnable parameters that are optimized using a computer through deep learning and error backpropagation (*24*). The use of deep learning in optical information processing systems has emerged in various exciting directions including integrated photonics solutions (*25*–*32*) and free-space optical platforms (*33*–*42*) involving, e.g., the use of diffraction (*21*, *43*–*46*). In this work, we harnessed the native dispersion properties of matter and trained a set of diffractive layers using deep learning to all-optically process a continuum of wavelengths to transform the spatial features of different object classes into a set of unique wavelengths, each representing one data class. This enabled us to use a single-pixel spectroscopic detector to perform optical classification of objects based on the spectral power encoded at these class-specific wavelengths. It should be emphasized that the task-specific spectral encoding provided through a trained diffractive optical network is a single shot encoding for, e.g., image classification, without the need for variable or structured illumination or dynamic spatial light modulators.

We demonstrated this novel machine vision framework by designing broadband diffractive optical networks that operate with pulsed illumination at terahertz wavelengths to achieve >96% blind testing accuracy for optical classification of handwritten digits (never seen by the network before) based on the spectral power at 10 distinct wavelengths, each assigned to one digit/class. Using a plasmonic nanoantenna-based source and detector as part of a terahertz time-domain spectroscopy (THz-TDS) system (*47*, *48*) and 3D printed diffractive models, our experiments provided very good match to our numerical results, successfully inferring the classes/digits of the input objects by maximizing the power of the wavelength corresponding to the true label.

In addition to optical classification of objects through spectral encoding of data classes, we also demonstrate a shallow artificial neural network (ANN) with two hidden layers that is successively trained (after the diffractive network training) to rapidly reconstruct the images of the classified objects based on their diffracted power spectra detected by a single-pixel spectroscopic detector. Using only 10 inputs, one for each class-specific wavelength, this shallow ANN is shown to successfully reconstruct images of the input objects even if they were incorrectly classified by the trained diffractive

[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA. [2]Bioengineering Department, University of California, Los Angeles, CA 90095, USA. [3]California NanoSystems Institute (CNSI), University of California, Los Angeles, CA 90095, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: ozcan@ucla.edu

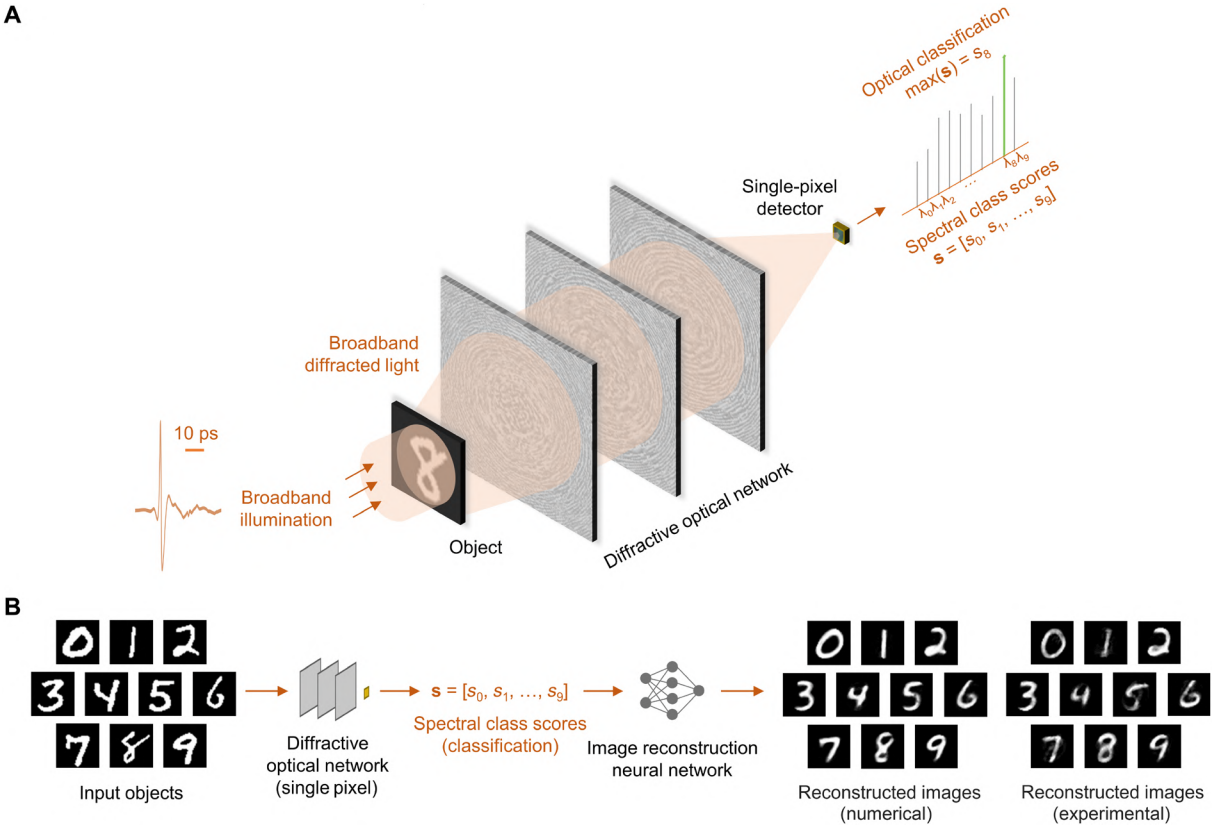

Fig. 1. Schematics of spectral encoding of spatial information for object classification and image reconstruction. (**A**) Optical layout of the single-detector machine vision concept for spectrally encoded classification of objects, e.g., the images of handwritten digits. As an example, a handwritten digit 8 is illuminated with a broadband pulsed light, and the subsequent diffractive optical network transforms the object information into the power spectrum of the diffracted light collected by a single detector. The object class is determined by the maximum of the spectral class scores, **s**, defined over a set of discrete wavelengths, each representing a data class (i.e., digit). (**B**) Schematic of task-specific image reconstruction using the diffractive network's spectral class scores as input. A separately trained shallow artificial neural network (ANN; with two hidden layers) recovers the images of handwritten digits from the spectral information encoded in **s**. Each reconstructed image is composed of >780 pixels, whereas the input vector, **s**, has 10 values.

network. Considering the fact that each image of a handwritten digit is composed of 784 pixels, this shallow image reconstruction ANN, with an input vector size of 10, performs a form of image decompression to successfully decode the task-specific spectral encoding of the diffractive network (i.e., the optical front-end). Despite being a modest ANN with two hidden layers, the success of this task-specific image reconstruction network, i.e., the decoder, also emphasizes the vital role of the collaboration between a trainable optical front-end and an all-electronic ANN-based back-end (*39*, *44*). Our results also demonstrate that once the reconstructed images of the input objects that were initially misclassified by the diffractive network are fed back into the same network as new inputs, their optical classification is corrected, significantly improving the overall inference accuracy of the trained diffractive network.

We believe that the framework presented in this work would pave the way for the development of various new machine vision systems that use spectral encoding of object information to achieve a specific inference task in a resource-efficient manner, with low latency, low power, and low pixel count. These features are particularly important since large pixel count of optical sensor arrays can put a burden on computational resources such as the allocated memory and the number of multiply-accumulate units required for statistical inference or classification over a large image size; furthermore, such high-resolution image sensors are not readily available at various parts of the electromagnetic spectrum, including, for example, far/midinfrared and terahertz bands. The teachings of this work can also be extended to spectral domain interferometric measurement systems, such as optical coherence tomography (OCT), Fourier transform infrared spectroscopy (FTIR), and others, to create fundamentally new 3D imaging and sensing modalities integrated with spectrally encoded classification tasks performed through trained diffractive networks. While the presented approach used solely the native dispersion properties of matter, we also envision harnessing metamaterials and their engineered dispersion to design diffractive spectral encoding systems with additional degrees of freedom.

RESULTS

Figure 1 illustrates our machine vision framework for spectral encoding of spatial information. A broadband diffractive network composed of layers is trained to transform the spatial information of the objects into the spectral domain through a preselected set of class-specific wavelengths measured by a single-pixel spectroscopic detector at the output plane; the resulting spectral class scores are denoted by the vector $\mathbf{s} = [s_0, s_1, \ldots, s_9]$ (Fig. 1A). Since, in this work, the learning task assigned to the diffractive network is the optical

classification of the images of handwritten digits [Modified National Institute of Standards and Technology (MNIST) database] (*49*), after its training and design phase, for a given input/test image, it learns to channel relatively more power to the spectral component assigned to the correct class (e.g., digit "8" in Fig. 1A) compared to the other class scores; therefore, max(**s**) reveals the correct data class. As demonstrated in Fig. 1B, the same class score vector, **s**, can also be used as an input to a shallow ANN with two hidden layers to reconstruct an image of the input object, decoding the spectral encoding performed by the broadband diffractive network.

On the basis of the system architecture shown in Fig. 1A, we trained broadband networks by taking the thickness of each pixel of a diffractive layer as a learnable variable (sampled at a lateral period of $\lambda_{min}/2$, where λ_{min} refers to the smallest wavelength of the illumination bandwidth) and accordingly defined a training loss ($\mathcal{L}_D$) for a given diffractive network design

$$\mathcal{L}_D = \mathcal{L}_I + \alpha \cdot \mathcal{L}_E + \beta \cdot \mathcal{L}_P \quad (1)$$

where $\mathcal{L}_I$ and $\mathcal{L}_E$ refer to the loss terms related to the optical inference task (e.g., object classification) and the diffractive power efficiency at the output detector, respectively. The spatial purity loss, $\mathcal{L}_P$, on the other hand, has a rather unique aim of clearing the light intensity over a small region of interest surrounding the active area of the single-pixel detector to improve the robustness of the machine vision system for uncontrolled lateral displacements of the detector position with respect to the optical axis (see the Supplementary Materials for detailed definitions of $\mathcal{L}_I$, $\mathcal{L}_E$, and $\mathcal{L}_P$). The hyperparameters, α and β, control the balance between the three major design factors, represented by these training loss terms.

To exemplify the performance of this design framework, using 10 class-specific wavelengths uniformly distributed between λ_{min} = 1.0 mm and λ_{max} = 1.45 mm, a three-layer diffractive optical network trained with α = β = 0 achieves >96% blind testing accuracy for spectrally encoded optical classification of handwritten digits (see Table 1, fourth row). Fine tuning of the hyperparameters, α and β, yields broadband diffractive network designs that provide improved diffractive power efficiency at the output detector and partial insensitivity to misalignments without excessively sacrificing the inference accuracy. For example, using α = 0.03 and β = 0.1, we achieve 95.05% blind testing accuracy for spectrally encoded optical classification of handwritten digits with ~1% inference accuracy drop compared to the diffractive model trained with α = β = 0 while at the same time achieving approximately eightfold higher diffractive power efficiency at the output detector (see Table 1). Figure 2B illustrates the resulting layer thickness distributions of this diffractive network trained with α = 0.03 and β = 0.1, setting a well-engineered example of the balance among inference accuracy, diffractive power efficiency at the output detector, and misalignment resilience of the diffractive network.

Next, we fabricated these diffractive layers shown in Fig. 2B (trained with α = 0.03 and β = 0.1 to achieve 95.05% blind testing accuracy) together with 50 handwritten digits (five per digit) randomly selected from the correctly classified blind testing samples using 3D printing (see Fig. 2A for the resulting diffractive network). Figure 2C also shows the THz-TDS setup with a plasmonic photoconductive detector that we used for the experimental validation of our machine vision framework (also see Materials and Methods). In this setup, the pulsed light emerging from a plasmonic photoconductive terahertz source is collimated and directed toward a square aperture with an area of 1 cm^2 (Fig. 2D), which serves as an entrance pupil to illuminate an unknown input object to be classified. As shown in Fig. 2D, we do not have any optical components or modulation layers between the illumination aperture and the object plane, indicating that there is no direct mapping between the spatial coordinates of the object plane and the spectral components of the illumination beam. On the basis of this experimental setup, the comparison between the power spectrum numerically generated using our trained forward model (dashed line) and its experimentally measured counterpart (straight line) for three fabricated digits, as examples, is illustrated in Fig. 3A, providing a decent match between the two and also revealing the correct class inference in each case through max(**s**). Despite 3D fabrication errors, possible misalignments, and other sources

Table 1. Numerical blind testing accuracies of different diffractive network models and their integration with decoder image reconstruction ANNs. The diffractive optical networks presented in the first three rows were trained with different (α, β) pairs for experimental validation, resulting in different diffractive power efficiencies at the output detector, while the model in the fourth and fifth row was trained with α = β = 0. The mean diffractive power efficiencies (η) of the diffractive network models were calculated at the output detector, considering the whole testing dataset, represented with the corresponding standard deviations (see the Supplementary Materials for details). MAE, mean absolute error; BerHu, reversed Huber; SCE, softmax cross-entropy.

Diffractive network	Diffractive power efficiency at the output detector: η (%)	Testing accuracy: max(s) (%)	Testing accuracy: max(s′) (%)
10 wavelengths, α = 0.4, β = 0.2 (Fig. 5); **s** = [s_0, s_1, …, s_9]	0.966 ± 0.465	84.02	MAE: 84.03 MAE + SCE: 91.29 BerHu + SCE: 91.06
10 wavelengths, α = 0.08, β = 0.2 (fig. S4); **s** = [s_0, s_1, …, s_9]	0.125 ± 0.065	93.28	MAE: 91.31 MAE + SCE: 94.27 BerHu + SCE: 94.02
10 wavelengths, α = 0.03, β = 0.1 (Fig. 3 and fig. S3); **s** = [s_0, s_1, …, s_9]	0.048 ± 0.027	95.05	MAE: 93.40 MAE + SCE: 95.32 BerHu + SCE: 95.37
10 wavelengths, α = β = 0 (fig. S5); **s** = [s_0, s_1, …, s_9]	0.006 ± 0.004	96.07	MAE: 94.58 MAE + SCE: 96.26 BerHu + SCE: 96.30
20 wavelengths (differential), α = β = 0 (fig. S8); $\mathbf{s_D}$ = [s_{0+}, s_{0-}, s_{1+}, s_{1-}, …, s_{9+}, s_{9-}]; **s** = Δ**s** = [Δs_0, Δs_1, …, Δs_9]	0.004 ± 0.002	96.82	MAE: 90.15 MAE + SCE: 96.81 BerHu + SCE: 96.64

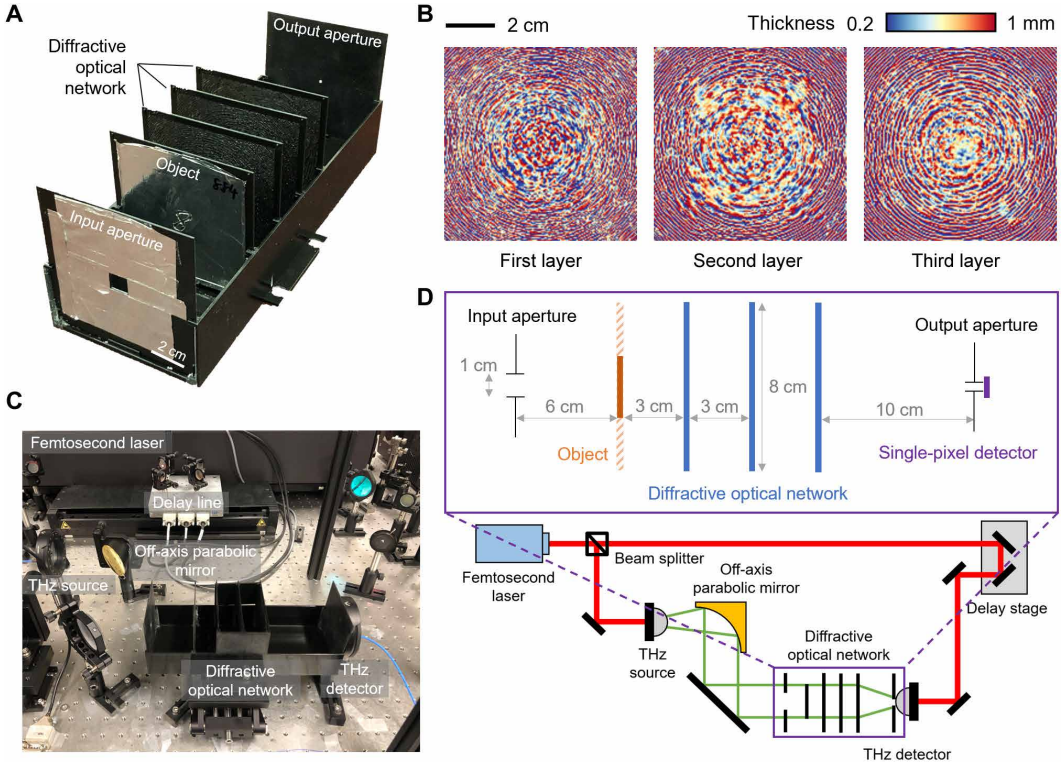

Fig. 2. Experimental setup. (**A**) A 3D printed diffractive network. (**B**) Learned thickness profiles of the three diffractive layers in (A). (**C**) Photograph of the experimental setup. (**D**) Top: Physical layout of the diffractive optical network setup; zoomed-in version of the bottom part. The object is a binary handwritten digit (from MNIST data), where the opaque regions are coated with aluminum to block the light transmission. Bottom: Schematic of the THz-TDS setup. Red lines depict the optical path of the femtosecond pulses generated by a Ti:sapphire laser operating at 780-nm wavelength. Green lines indicate the optical path of the terahertz pulse (peak frequency, ~500 GHz and observable bandwidth, ~5 THz), which is modulated by the 3D printed diffractive neural network to spectrally encode the task-specific spatial information of the objects. Photo credit (A and C): Jingxi Li, UCLA.

of error in our setup, the match between the experimental and numerical testing of our diffractive network design was found to be 88% using 50 handwritten digits that were 3D printed (see Fig. 3B).

For the same 3D printed diffractive model (Fig. 2, A and B), we also trained a shallow, fully connected ANN with two hidden layers to reconstruct images of the unknown input objects based on the detected **s**. The training of this decoder ANN is based on the knowledge of (i) the class scores ($\mathbf{s} = [s_0, s_1, \ldots, s_9]$), resulting from the trained diffractive network model, and (ii) the corresponding input object images. Without any fine tuning of the network parameters for possible deviations between our numerical forward model and the experimental setup, when this shallow ANN was blindly tested on our experimental measurements (**s**), the reconstructions of the images of the handwritten digits were successful, as illustrated in Fig. 1B (also see fig. S6), further validating the presented framework and the experimental robustness of our diffractive network model (see the Supplementary Materials for further details). It should be emphasized that this shallow ANN is trained to decode a highly compressed form of information that is spectrally encoded by a diffractive front-end and that it uses only 10 numbers (i.e., $s_0, s_1, \ldots, s_9$) at its input to reconstruct an image that has >780 pixels. Stated differently, this ANN performs a form of task-specific image decompression, the task being the reconstruction of the images of handwritten digits based on spectrally encoded inputs (**s**). In addition to performing task-specific image reconstruction, the presented machine vision framework can possibly be extended for the design of a general-purpose single-pixel imaging system based on spectral encoding; although, here, in this work, we focused on the reconstruction of the classified object images (i.e., handwritten digits).

In addition to the diffractive network shown in Fig. 2 that achieved a numerical blind testing accuracy of 95.05%, we also 3D-fabricated and experimentally tested two additional diffractive network models to further evaluate the match between our numerical models and their experimental/physical counterparts. By using different (α, β) pairs for the loss function defined in Eq. 1, the balance between the optical inference accuracy and the two practical design merits, i.e., the diffractive power efficiency at the output detector and the insensitivity to misalignments, is shifted in these two new diffractive designs in favor of experimental robustness. Performance comparisons of these diffractive network models are summarized in Table 1 and Fig. 3C; for example, using $\alpha = 0.4$ and $\beta = 0.2$, the blind testing accuracy attained by the same three-layer diffractive network architecture decreased to 84.02% for the handwritten digit classification task, while the diffractive power efficiency at the output detector increased by a factor of ~160 and the match between our experimental and numerical testing results increased to 96%. These results, as summarized in Fig. 3C and Table 1, further demonstrate the trade-off between the inference accuracy and the diffraction efficiency and experimental robustness of our diffractive network models.

To provide a mitigation strategy for this trade-off, next, we introduced a collaboration framework between the diffractive network

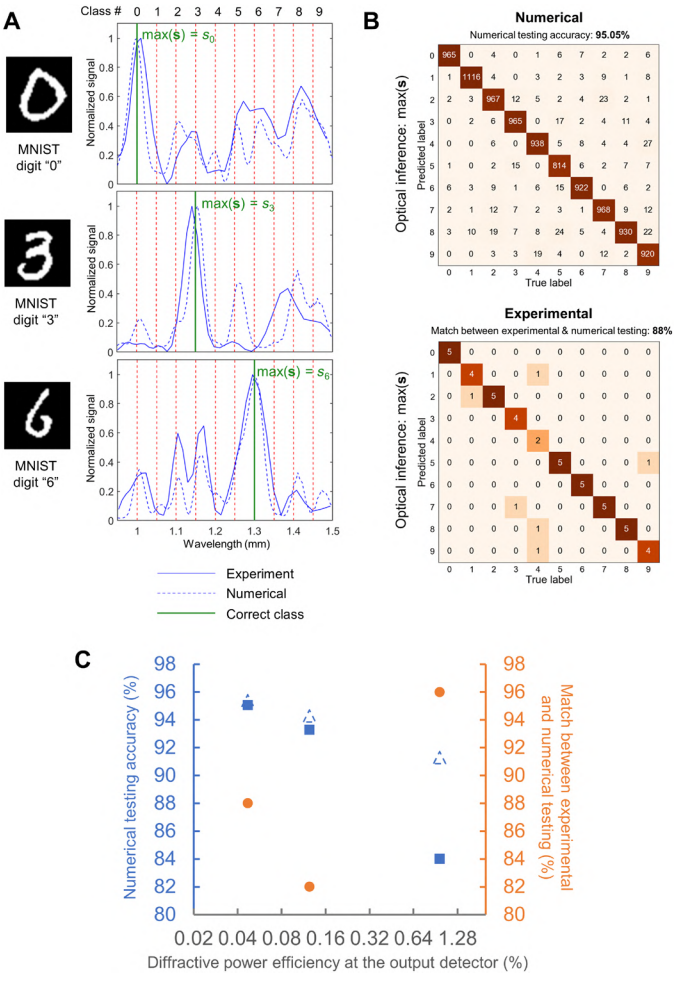

Fig. 3. Spectrally encoded optical classification of handwritten digits with a single detector. (**A**) Experimentally measured (blue solid line) and the numerically computed (blue dashed line) output power spectra for optical classification of three different handwritten digits, shown as examples. The object class is determined by the maximum of the spectral class scores, **s**, defined over a set of discrete wavelengths, each representing a digit. (**B**) Top: Confusion matrix summarizing the numerical classification performance of the diffractive optical network that attains a classification accuracy of 95.05% over 10,000 handwritten digits in the blind testing set. Bottom: Confusion matrix for the experimental results obtained by 3D printing of 50 handwritten digits randomly selected from the numerically successful classification samples in the blind testing set. An 88% match between the experimentally inferred and the numerically computed object classes is observed. (**C**) Comparison of three different diffractive networks that were trained, fabricated, and experimentally tested in terms of (i) their numerical blind testing accuracies (blue solid squares), (ii) the match between experimentally measured and numerically predicted object classes (orange solid circles), and (iii) the inference accuracy achieved by feeding the decoder ANN's reconstructed images back to the diffractive network as new inputs (blue dashed triangles).

and its corresponding reconstruction ANN. This collaboration is based on the fact that our decoder ANN can faithfully reconstruct the images of the input objects using the spectral encoding present in **s**, even if the optical classification is incorrect, pointing to a wrong class through max(**s**). We observed that by feeding the decoder ANN's reconstructed images back to the diffractive network as new inputs, we can have it correct its initial wrong inference (see Fig. 4 and fig. S2). Through this collaboration between the diffractive network and its decoder ANN, we improved the overall inference accuracy of a given diffractive network model as summarized in Fig. 3C and Table 1. For example, for the same, highly efficient diffractive network model that was trained using $\alpha = 0.4$ and $\beta = 0.2$, the blind testing accuracy for handwritten digit classification increased from 84.02 to 91.29% (see Figs. 3C and 5B), demonstrating a substantial improvement through the collaboration between the decoder ANN and the broadband diffractive network. A close examination of Fig. 5 and the provided confusion matrices reveal that the decoder ANN, through its image reconstruction, helped correct 870 misclassifications of the diffractive network, resulting in an overall gain/improvement of 7.27% in the blind inference performance of the optical network. Similar analyses for the other diffractive network models are also presented in figs. S3 to S5.

In this collaboration between the diffractive network and its corresponding shallow decoder, the training loss function of the latter (ANN) was coupled to the classification performance of the diffractive network. In other words, in addition to a structural loss function ($\mathcal{L}_S$) that is needed for a high-fidelity image reconstruction, we also added a second loss term that penalized the ANN by a certain weight if its reconstructed image cannot be correctly classified by the diffractive network (see the Supplementary Materials). This ensures that the collaboration between the optical encoder and its corresponding decoder ANN is constructive, i.e., the overall classification accuracy is improved through the feedback of the reconstructed images onto the diffractive network as new inputs. On the basis of this collaboration scheme, the general loss function of the decoder ANN can be expressed as

$$\mathcal{L}_{\text{Recon}} = \gamma \cdot \mathcal{L}_S(\mathbf{O}_{\text{recon}}, \mathbf{O}_{\text{input}}) + (1-\gamma) \cdot \mathcal{L}_I \quad (2)$$

where $\mathcal{L}_S$ refers to a structural loss term, e.g., mean absolute error (MAE) or reversed Huber ("BerHu") loss (*50*, *51*), computed through pixel-wise comparison of the reconstructed image ($\mathbf{O}_{\text{recon}}$) with the ground truth object image ($\mathbf{O}_{\text{input}}$) (see the Supplementary Materials for details). The second term in Eq. 2, $\mathcal{L}_I$, refers to the same loss function used in the training of the diffractive network (front-end) as in Eq. 1, except this time, it is computed over the new class scores, $\mathbf{s}'$, obtained by feeding the reconstructed image, $\mathbf{O}_{\text{recon}}$, back to the same diffractive network (see Fig. 5 and fig. S1). Equation 2 is only concerned with the training of the image reconstruction ANN, and therefore, the parameters of the decoder ANN are updated through standard error backpropagation, while the diffractive network model is preserved.

Table 1 summarizes the performance comparison of different loss functions used to train the decoder ANN and their impact on the improvement of the classification performance of the diffractive network. Compared to the case when $\gamma = 1$, which refers to independent training of the reconstruction ANN without taking into account $\mathcal{L}_I$, we see substantial improvements in the inference accuracy of the diffractive network through max($\mathbf{s}'$) when the ANN has been penalized during its training (with, e.g., $\gamma = 0.95$) if its reconstructed images cannot be correctly classified by the diffractive network (refer to the Supplementary Materials for details). Stated differently, the use of the $\mathcal{L}_I$ term in Eq. 2 for the training of the decoder ANN tailors the image reconstruction space to generate object features that are more favorable for the diffractive optical classification while also retaining its reconstruction fidelity to the ground truth object, $\mathbf{O}_{\text{input}}$, by the courtesy of the structural loss term, $\mathcal{L}_S$, in Eq. 2.

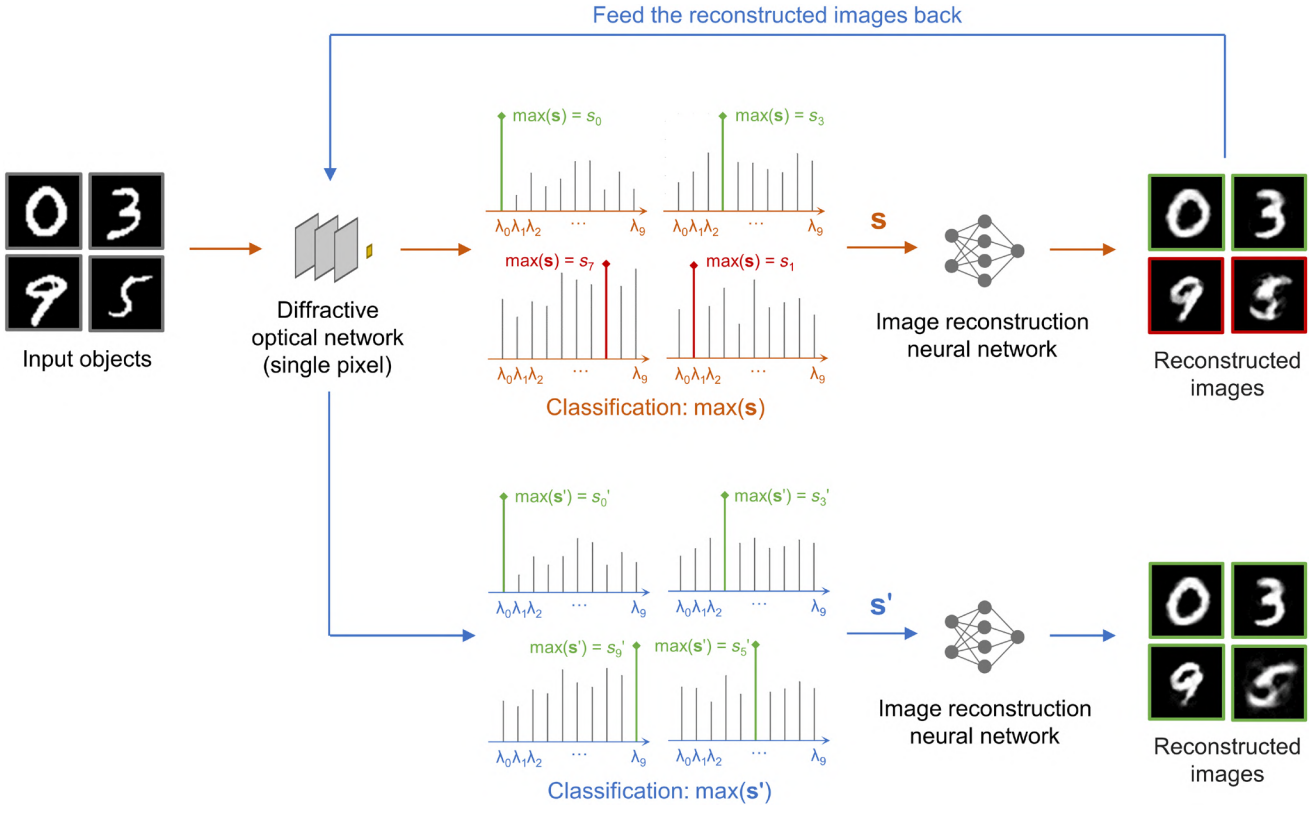

Fig. 4. Illustration of the coupling between the image reconstruction ANN and the diffractive network. Four MNIST images of handwritten digits are used here for illustration of the concept. Two of the four samples, "0" and "3", are correctly classified by the diffractive network based on max(**s**) (top green lines), while the other two, "9" and "5", are misclassified as "7" and "1", respectively (top red lines). Using the same class scores (**s**) at the output detector of the diffractive network, a shallow decoder ANN digitally reconstructs the images of the input objects. Next, these images are cycled back to the diffractive optical network as new input images, and the new spectral class scores **s′** are inferred accordingly, where all of the four digits are correctly classified through max(**s′**) (bottom green lines). Last, these new spectral class scores **s′** are used to reconstruct the objects again using the same image reconstruction ANN. The blind testing accuracy of this diffractive network for handwritten digit classification increased from 84.02 to 91.29% using this feedback loop (see Figs. 3C and 5B). This image reconstruction decoder ANN was trained using the MAE loss and softmax cross-entropy loss (see Eq. 2).

The performance of the presented spectral encoding–based machine vision framework can be further improved using a differential class encoding strategy (*45*). For this aim, we explored the use of two different wavelengths to encode each class score: Instead of using 10 discrete wavelengths to represent a spectral class score vector, $\mathbf{s} = [s_0, s_1, \ldots, s_9]$, we considered encoding the spatial information of an object into 20 different wavelengths ($s_{0+}, s_{0-}, s_{1+}, s_{1-}, \ldots, s_{9+}, s_{9-}$) that are paired in groups of two to differentially represent each spectral class score, i.e., $\Delta s_c = \frac{s_{c,+} - s_{c,-}}{s_{c,+} + s_{c,-}}$. In this differential spectral encoding strategy, the trained diffractive network makes an inference based on max(**Δs**) resulting from the spectral output at the single-pixel detector. With this spectrally encoded differential classification scheme, we numerically attained 96.82% optical classification accuracy for handwritten digits (see Table 1 and fig. S8).

As an alternative to the shallow decoder ANN with two hidden layers used earlier, we also explored the use of a much deeper ANN architecture (*52*) as the image reconstruction network in our spectrally encoded machine vision framework. For this, the output of the two hidden layer fully connected network (with an input of **s**) is further processed by a U-Net–like deep convolutional ANN with skip connections and a total of >1.4 million trainable parameters to reconstruct the images of handwritten digits using **s**. We found out that the collaboration of the diffractive networks with this deeper ANN architecture yielded only marginal improvements over the classification accuracies presented in Table 1. For example, when the diffractive optical network design shown in Fig. 2B ($\alpha = 0.03, \beta = 0.1$) was paired with this deep decoder ANN (through the feedback depicted in Fig. 4), the blind classification accuracy increased to 95.52% compared to the 95.37% provided by the shallow decoder ANN with two hidden layers. As another example, for the diffractive optical network trained with $\alpha = 0.4$ and $\beta = 0.2$, the collaboration with the deep convolutional ANN provides a classification accuracy of 91.49%, which is a minor improvement with respect to the 91.29% accuracy produced through the shallow ANN, falling short to justify the disadvantages of using a deeper ANN-based decoder architecture in terms of its slower inference speed and more power consumption per image reconstruction.

The function of the decoder ANN, up to this point, has been to reconstruct the images of the unknown input objects based on the encoding present in the spectral class scores, **s**. Therefore, note that, in these earlier results, the classification is performed optically, through the spectral output of the diffractive network, and the function of the decoder ANN is not to infer a data class but rather to reconstruct the image of the object that is classified by the diffractive optical network using max(**s**). As an alternative strategy, we also explored making use of the decoder ANN for a different task: to

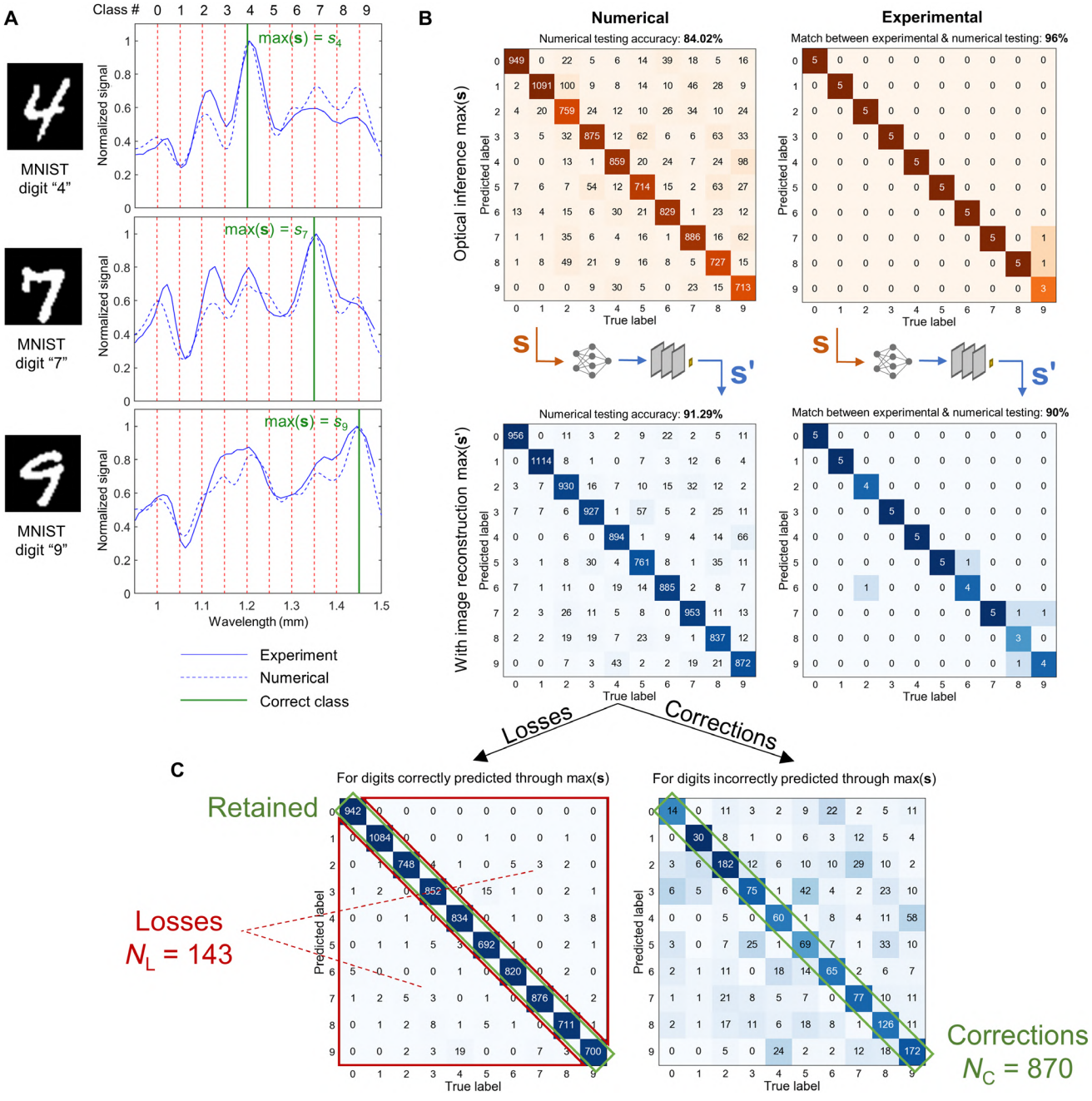

Fig. 5. Blind testing performance of an efficient diffractive network and its coupling with a corresponding decoder ANN. (**A**) Experimentally measured (blue solid line) and the numerically computed (blue dashed line) output power spectra for optical classification of three different handwritten digits, shown as examples. (**B**) Top left: Confusion matrix summarizing the numerical classification performance of the diffractive network that attains a classification accuracy of 84.02% over 10,000 handwritten digits in the blind testing set. Top right: Confusion matrix for the experimental results obtained by 3D printing 50 handwritten digits randomly selected from the numerically successful classification samples in the blind testing set. A 96% match between the experimentally inferred and the numerically computed object classes is observed. Bottom left: Confusion matrix provided by max(**s'**), computed by feeding the reconstructed images back to the diffractive network. A blind testing accuracy of 91.29% is achieved, demonstrating a substantial classification accuracy improvement of 7.27% (also see Fig. 4). Bottom right: Confusion matrix for the experimental results using the same 50 digits. (**C**) Left: Same as the bottom left matrix in (B) but solely for the digits that are correctly predicted by the optical network. Its diagonal entries can be interpreted as the digits that are retained to be correctly predicted, while its off-diagonal entries represent the "losses" after the image reconstruction and feedback process. Right: Same as the left one but solely for the digits that are incorrectly classified by the optical network. Its diagonal entries indicate the optical classification "corrections" after the image reconstruction and feedback process. The number $N_C - N_L = 727$ is the classification accuracy "gain" achieved through max(**s'**), corresponding to a 7.27% increase in the numerical testing accuracy of the diffractive model (also see Fig. 3C).

directly classify the objects on the basis of the spectral encoding (**s**) provided by the diffractive network. In this case, the decoder ANN is solely focused on improving the classification performance with respect to the optical inference results that are achieved using max(**s**). For example, on the basis of the spectral class scores encoded by the diffractive optical networks that achieved 95.05 and 96.07% blind testing accuracy for handwritten digit classification using max(**s**), a fully connected, shallow classification ANN with two hidden layers improved the classification accuracy to 95.74 and 96.50%, respectively. Compared to the accuracy values presented in Table 1, these numbers indicate that a slightly better classification performance is possible, provided that the image reconstruction is not essential for the target application, and can be replaced with a shallow classification decoder ANN that takes **s** as its input.

In the earlier machine vision systems that we have presented so far, the diffractive optical network and the corresponding back-end ANN have been separately trained, i.e., after the training of the diffractive network for optical image classification, the back-end ANN was trained on the basis of the spectral encoding of the converged diffractive network model, yielding either the reconstruction ANN or the classification ANN, as discussed earlier. As an alternative strategy, such hybrid systems can also be jointly trained through the error backpropagation between the electronic ANN and the diffractive optical front-end (*44*, *53*). Here, we demonstrated this opportunity using the MNIST dataset and jointly trained a diffractive network with an image reconstruction ANN at the back-end; in the next paragraphs, the same approach will also be extended to jointly train a diffractive network with a classification ANN at the back-end, covering a different dataset [Extended MNIST (EMNIST)] (*54*). In our joint training of hybrid network systems composed of a diffractive network and a reconstruction ANN, we used a linear superposition of two different loss functions to optimize both the optical classification accuracy and the image reconstruction fidelity; see eq. S22 and table S2. Through this linear superposition, we explored the impact of different relative weights of these loss functions on (i) the image classification accuracy of the diffractive network and (ii) the quality of the image reconstruction performed by the back-end ANN. For this goal, we changed the relative weight (ξ) of the optical classification loss term to shift the attention of the hybrid design between these two tasks. For instance, when the weight of the optical classification loss is set to be zero ($\xi = 0$), the entire hybrid system becomes a computational single-pixel imager that ignores the optical classification accuracy and focuses solely on the image reconstruction quality; as confirmed in figs. S11 and S12 and table S2, this choice ($\xi = 0$) results in a substantial degradation of the optical image classification accuracy with a considerable gain in the image reconstruction fidelity, as expected. By using different relative weights, one can achieve a sweet spot in the joint training of the hybrid network system, where both the optical image classification accuracy and the ANN image reconstruction fidelity are very good (see, e.g., $\xi = 0.5$ in table S2 and figs. S11 and S12).

We also investigated the inference performance of these hybrid systems in terms of the number of wavelengths that are simultaneously processed through the diffractive network. For this, we jointly trained hybrid systems that assign a group of wavelengths to each data class; inference of an object class is then based on the maximum average power accumulated in these selected spectral bands, where each band represents one data class (see the Supplementary Materials for further details). Our results, summarized in table S2, reveal that assigning, e.g., five distinct wavelengths to each data class (i.e., a total of 50 wavelengths for 10 data classes), achieved a similar optical classification accuracy compared to their counterparts that encoded the objects' spatial information using fewer wavelengths. This indicates that the diffractive networks can be designed to simultaneously process a larger number of wavelengths to successfully encode the spatial information of the input field of view into spectral features.

To further explore the capabilities of the presented single-pixel spectroscopic machine vision framework for more challenging image classification tasks beyond handwritten digits, we used the EMNIST dataset (*54*), containing 26 object classes, corresponding to handwritten capital letters (see fig. S13). For this EMNIST image dataset, we trained nondifferential and differential diffractive classification networks, encoding the information of the object data classes into the output power of 26 and 52 distinct wavelengths, respectively. Furthermore, to better highlight the benefits of the collaboration between the optical and electronic networks, we also jointly trained hybrid network systems that use a shallow classification ANN (with two hidden layers) described earlier to extract the object class from the spectral encoding performed by the diffractive optical front-end, through a single-pixel detector, same as before. Table S1 summarizes our results on this 26-class handwritten capital letter image dataset. First, a comparison between the all-optical diffractive classification networks and the jointly trained hybrid network systems highlights the importance of the collaboration between the optical and electronic networks: The jointly trained hybrid systems (where a diffractive network is followed by a classification encoder ANN) can achieve higher object classification accuracies (see table S1). For example, a jointly trained hybrid network using 52 encoding wavelengths that are processed through three diffractive layers and a shallow decoder ANN achieved a classification accuracy of 87.68% for EMNIST test dataset, which is >2% higher compared to the inference accuracy attained solely by an optical diffractive network design based on differential spectral encoding using the same 52 wavelengths (table S1). The results presented in table S1 further reveal that both the jointly trained hybrid systems and the optical diffractive classification systems that use 52 distinct wavelengths to encode the spatial information of the objects achieve higher classification accuracies compared to their counterparts that are designed to process 26 wavelengths.

DISCUSSION

Although Eq. 1 tries to find a balance among the optical inference accuracy, detector photon efficiency, and resilience to possible detector misalignment, there are other sources of experimental errors that contribute to the physical implementations of trained diffractive networks. First, because of the multilayer layout of these diffractive networks, any interlayer misalignments might have contributed to some of the errors that we observed during the experiments. In addition, our optical forward model does not take into account multiple reflections that occur through the diffractive layers. These are relatively weaker effects that can be mitigated by, e.g., time gating of the detector output and/or using antireflection coatings that are widely used in the fabrication of conventional optical components. Moreover, measurement errors that might have taken place during the characterization of the dispersion of the diffractive-layer material can cause our numerical models to slightly deviate from their

physical implementations. Furthermore, 3D fabrication errors stemming from printing overflow and cross-talk between diffractive features on the layers can also contribute to some of the differences observed between our numerical and experimental results. Figure S14 illustrates a comparison between one of the 3D printed diffractive layers and its numerical design, exemplifying some of these fabrication-related imperfections that are experimentally observed.

The negative effects of some of these experimental errors outlined above can be mitigated by modeling undesired physical system variations over random variables that are incorporated as part of the optical forward model used in the deep learning–based training of the diffractive network (*53*, *55*). Thereby, the evolution of the parameter space of the underlying diffractive layers can be regulated to preserve their collective inference accuracy despite, e.g., misalignments. We explored the impact of this idea for building misalignment resilience in jointly trained, hybrid MNIST classification networks formed by a three-layer spectral encoder diffractive front-end and a shallow classification ANN (with two hidden layers). On the basis of this test bed, we modeled the lateral misalignments of the spectral encoder diffractive layers by defining two independent, uniformly distributed random variables per layer, $D_x^l \sim U(-\Delta_x^l, \Delta_x^l)$ and $D_y^l \sim U(-\Delta_y^l, \Delta_y^l)$, representing the displacement of layer l with respect to its ideal location in x and y directions, respectively. The hyperparameters, Δ_x^l and Δ_y^l, determine the range of the positioning error along the corresponding axis. During the training (tr) phase, D_x^l and D_y^l were randomly updated at each iteration, uniformly taking random values from the range set by $\Delta_x^l = \Delta_y^l = \Delta_{tr}^l$. Such a training strategy (which we term as vaccination) introduces new perturbed diffractive layer locations at each update step and, as a result, guides the evolution of the spectral encoder diffractive layers to a solution that is more resilient against misalignments, allowing the diffractive networks to maintain their optical inference accuracy over larger margins of physical misalignments.

To demonstrate the impact of the outlined training strategy, we quantified the blind inference accuracies achieved by various vaccinated and nonvaccinated diffractive single-pixel machine vision systems under a series of misalignments, as shown in fig. S15. In fig. S15A, we report that the vaccination results when only the middle diffractive layer ($l = 2$) is misaligned from its ideal location, meaning that the centers of the first and third diffractive layers coincide with the optical axis, i.e., $\Delta_x^1 = \Delta_y^1 = \Delta_{test}^1 = \Delta_{tr}^1 = 0.0$ and $\Delta_x^3 = \Delta_y^3 = \Delta_{test}^3 = \Delta_{tr}^3 = 0.0$. In fig. S15B, on the other hand, all three diffractive layers experience random lateral shifts along both x and y axes during the testing phase. In our analyses, we also investigated the effect of the single-pixel detector size on the misalignment resilience of these hybrid neural network systems and accordingly trained vaccinated and nonvaccinated single-pixel spectral encoder diffractive systems, each with a detector active area of 2 mm by 2 mm, 4 mm by 4 mm, and 8 mm by 8 mm. As depicted in fig. S15 (A and B), almost independent of the active area of the single-pixel detector, the classification accuracy of the nonvaccinated hybrid networks (blue) are rather sensitive to mechanical misalignments of the diffractive network layers. The vaccinated networks, on the other hand, can maintain their blind inference accuracy over a wider range of misalignments, which is confirmed in both panels A and B of fig. S15. Furthermore, a comparison of the classification accuracies provided by the vaccinated hybrid network systems reveals that the design with a larger active area (8 mm by 8 mm) single-pixel detector achieves better resilience over misalignments (see fig. S15B).

Without loss of generality, in this work, we used a three-layer diffractive network architecture to encode the spatial features of the object field of view into the output power spectrum for single-pixel machine vision. Note that if the material absorption of the diffractive layers is lower and/or the signal-to-noise ratio of the single-pixel detector is increased, then the optical inference accuracy of the presented network designs could be further improved by, e.g., increasing the number of diffractive layers or the number of learnable features (i.e., neurons) within the optical network (*44*, *56*). Compared to using wider diffractive layers, increasing the number of diffractive layers offers a more practical method to enhance the information processing capacity of diffractive networks since training higher–numerical aperture diffractive systems through image data is, in general, relatively harder (*56*). Despite their improved generalization capability, such deeper diffractive systems composed of larger numbers of diffractive layers would partially suffer from increased material absorption and surface backreflections. However, note that the optical power efficiency of a broadband network also depends on the size of the output detector. For example, the relatively lower power efficiency numbers reported in Table 1 are by and large due to the small size of the output detector used in these designs ($2 \times \lambda_{min}$) and can be substantially improved by using a detector with a much larger active area (*57*).

In conclusion, we demonstrated an optical machine vision system composed of trained diffractive layers to encode the spatial information of objects into the power spectrum of diffracted light, which is used to perform optical classification of unknown objects with a single-pixel spectroscopic detector. We also showed that shallow, low-complexity ANNs can be used as decoders to reconstruct images of the input objects based on the spectrally encoded class scores, demonstrating task-specific image decompression. Although we used terahertz pulses to experimentally validate our spectrally encoded machine vision framework, it can be broadly adopted for various applications covering other parts of the electromagnetic spectrum. In addition to object recognition, this machine vision concept can also be extended to perform other learning tasks such as scene segmentation and multilabel classification, as well as to design single-pixel or few-pixel, low-latency superresolution imaging systems by harnessing the spectral encoding provided by diffractive networks coupled with shallow decoder ANNs. We also envision that dispersion-engineered material systems such as metamaterials will open up a new design space for enhancing the inference and generalization performance of spectral encoding through diffractive optical networks. Last, the methods presented in this work would create new 3D imaging and sensing modalities that are integrated with optical inference and spectral encoding capabilities of broadband diffractive networks and can be merged with some of the existing spectroscopic measurement techniques, such as OCT, FTIR, and others, to find various new applications in biomedical imaging, analytical chemistry, material science, and other fields.

MATERIALS AND METHODS

THz-TDS setup

The schematic diagram of the THz-TDS setup is shown in Fig. 2D. We used a Ti:sapphire laser (Coherent Mira-HP) in a mode-locked operation mode to generate femtosecond optical pulses at a center wavelength of 780 nm. The laser beam was first split in two parts. One part of the beam illuminated the terahertz source, a plasmonic

photoconductive nanoantenna array (*48*), to generate terahertz pulses. The other part of the laser beam passed through an optical delay line and illuminated the terahertz detector, which was another plasmonic photoconductive nanoantenna array offering high-sensitivity and broadband operation (*47*). The generated terahertz radiation was collimated and guided to the terahertz detector using an off-axis parabolic mirror. The output signal as a function of the delay line position, which provides the temporal profile of the detected terahertz pulses, was amplified using a current preamplifier (Femto DHPCA-100) and detected with a lock-in amplifier (Zurich Instruments MFLI). For each measurement, 10 time-domain traces were captured over 5 s and averaged. The acquired time-domain signal has a temporal span of 400 ps, and its power spectrum was obtained through a Fourier transform. Overall, the THz-TDS system offers signal-to-noise ratio levels of >90 dB and observable bandwidths exceeding 5 THz.

The 3D printed diffractive optical network was placed between the terahertz source and the detector. It consisted of an input aperture, an input object, three diffractive layers, and an output aperture, as shown in Fig. 2D, with their dimensions and spacing annotated. Upon their training in a computer, the diffractive optical networks were fabricated using a 3D printer (Objet30 Pro, Stratasys Ltd.) with an ultraviolet curable material (VeroBlackPlus RGD875, Stratasys Ltd.). A 1 cm–by–1 cm square aperture was positioned at the input plane serving as an entrance pupil for the subsequent optical system. The terahertz detector has an integrated Si lens in the form of a hemisphere directly attached to the backside of the chip. This Si lens was modeled as an achromatic flat Si slab with a thickness of 0.5 cm and a refractive index of 3.4 in our optical forward model. During the experiments, a 2 mm–by–2 mm output aperture was placed at the output plane, right before the terahertz detector, to shrink the effective area of the Si lens, ensuring that the uniform slab model assumed during the training forward model accurately translates into our experimental setup. The input and output apertures and the 3D printed objects were coated with aluminum to block terahertz radiation outside the transparent openings and object features. Furthermore, a 3D printed holder (Fig. 2A) was designed to support and align all of the components of the diffractive setup.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/13/eabd7690/DC1

REFERENCES AND NOTES

1. J. B. Pendry, Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
2. E. Cubukcu, K. Aydin, E. Ozbay, S. Foteinopoulou, C. M. Soukoulis, Negative refraction by photonic crystals. *Nature* **423**, 604–605 (2003).
3. N. Fang, H. Lee, C. Sun, X. Zhang, Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308**, 534–537 (2005).
4. Z. Jacob, L. V. Alekseyev, E. Narimanov, Optical hyperlens: Far-field imaging beyond the diffraction limit. *Opt. Express* **14**, 8247–8256 (2006).
5. N. Engheta, Circuits with light at nanoscales: Optical nanocircuits inspired by metamaterials. *Science* **317**, 1698–1702 (2007).
6. Z. Liu, H. Lee, Y. Xiong, C. Sun, X. Zhang, Far-field optical hyperlens magnifying sub-diffraction-limited objects. *Science* **315**, 1686–1686 (2007).
7. K. F. MacDonald, Z. L. Sámson, M. I. Stockman, N. I. Zheludev, Ultrafast active plasmonics. *Nat. Photonics* **3**, 55–58 (2009).
8. D. Lin, P. Fan, E. Hasman, M. L. Brongersma, Dielectric gradient metasurface optical elements. *Science* **345**, 298–302 (2014).
9. N. Yu, F. Capasso, Flat optics with designer metasurfaces. *Nat. Mater.* **13**, 139–150 (2014).
10. A. I. Kuznetsov, A. E. Miroshnichenko, M. L. Brongersma, Y. S. Kivshar, B. Luk'yanchuk, Optically resonant dielectric nanostructures. *Science* **354**, aag2472 (2016).
11. S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, A. A. G. Requicha, Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides. *Nat. Mater.* **2**, 229–232 (2003).
12. A. Alù, N. Engheta, Achieving transparency with plasmonic and metamaterial coatings. *Phys. Rev. E* **72**, 016623 (2005).
13. D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, D. R. Smith, Metamaterial electromagnetic cloak at microwave frequencies. *Science* **314**, 977–980 (2006).
14. J. B. Pendry, D. Schurig, D. R. Smith, Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
15. W. Cai, U. K. Chettiar, A. V. Kildishev, V. M. Shalaev, Optical cloaking with metamaterials. *Nat. Photonics* **1**, 224–227 (2007).
16. J. Valentine, J. Li, T. Zentgraf, G. Bartal, X. Zhang, An optical cloak made of dielectrics. *Nat. Mater.* **8**, 568–571 (2009).
17. E. E. Narimanov, A. V. Kildishev, Optical black hole: Broadband omnidirectional light absorber. *Appl. Phys. Lett.* **95**, 041106 (2009).
18. R. F. Oulton, V. J. Sorger, T. Zentgraf, R.-M. Ma, C. Gladden, L. Dai, G. Bartal, X. Zhang, Plasmon lasers at deep subwavelength scale. *Nature* **461**, 629–632 (2009).
19. Y. Zhao, M. A. Belkin, A. Alù, Twisted optical metamaterials for planarized ultrathin broadband circular polarizers. *Nat. Commun.* **3**, 870 (2012).
20. C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, W. J. Padilla, Terahertz compressive imaging with metamaterial spatial light modulators. *Nat. Photonics* **8**, 605–609 (2014).
21. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, A. Ozcan, All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
22. N. Mohammadi Estakhri, B. Edwards, N. Engheta, Inverse-designed metastructures that solve equations. *Science* **363**, 1333–1338 (2019).
23. T. W. Hughes, I. A. D. Williamson, M. Minkov, S. Fan, Wave physics as an analog recurrent neural network. *Sci. Adv.* **5**, eaay6946 (2019).
24. D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
25. K. H. Wagner, in *Frontiers in Optics 2017* (Optical Society of America, 2017), p. FW2C.1; www.osapublishing.org/abstract.cfm?uri=FiO-2017-FW2C.1.
26. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, M. Soljačić, Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
27. T. F. de Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, P. R. Prucnal, Progress in neuromorphic photonics. *Nanophotonics* **6**, 577–599 (2017).
28. B. J. Shastri, A. N. Tait, T. F. de Lima, M. A. Nahmias, H.-T. Peng, P. R. Prucnal, Principles of neuromorphic photonics. arXiv:1801.00016 [cs.ET] (2018).
29. J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, D. Brunner, Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* **5**, 756–760 (2018).
30. E. Khoram, A. Chen, D. Liu, L. Ying, Q. Wang, M. Yuan, Z. Yu, Nanophotonic media for artificial neural inference. *Photonics Res.* **7**, 823–827 (2019).
31. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, W. H. P. Pernice, All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
32. L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, T. Mueller, Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
33. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558 (1982).
34. N. H. Farhat, D. Psaltis, A. Prata, E. Paek, Optical implementation of the Hopfield model. *Appl. Optics* **24**, 1469–1475 (1985).
35. K. Wagner, D. Psaltis, Multilayer optical learning networks. *Appl. Optics* **26**, 5061–5076 (1987).
36. D. Psaltis, A. Sideris, A. A. Yamamura, A multilayered neural network controller. *IEEE Control Syst. Mag.* **8**, 17–21 (1988).
37. D. Psaltis, D. Brady, X.-G. Gu, S. Lin, Holography in artificial neural networks. *Nature* **343**, 325–330 (1990).
38. A. V. Krishnamoorthy, G. Yayla, S. C. Esener, in *Proceedings of the IJCNN-91-Seattle International Joint Conference on Neural Networks* (IEEE, 1991), vol. 1, pp. 527–534.
39. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, G. Wetzstein, Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* **8**, 12324 (2018).
40. Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, S. Du, All-optical neural network with nonlinear activation functions. *Optica* **6**, 1132–1137 (2019).

41. L. Lu, L. Zhu, Q. Zhang, B. Zhu, Q. Yao, M. Yu, H. Niu, M. Dong, G. Zhong, Z. Zeng, Miniaturized diffraction grating design and processing for deep neural network. *IEEE Photonics Technol. Lett.* **31**, 1952–1955 (2019).
42. P. del Hougne, M. F. Imani, A. V. Diebold, R. Horstmeyer, D. R. Smith, Learned integrated sensing pipeline: Reconfigurable metasurface transceivers as trainable physical layer in an artificial neural network. *Adv. Sci.* **7**, 1901913 (2020).
43. Y. Luo, D. Mengu, N. T. Yardimci, Y. Rivenson, M. Veli, M. Jarrahi, A. Ozcan, Design of task-specific optical systems using broadband diffractive neural networks. *Light Sci. Appl.* **8**, 112 (2019).
44. D. Mengu, Y. Luo, Y. Rivenson, A. Ozcan, Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 1–14 (2020).
45. J. Li, D. Mengu, Y. Luo, Y. Rivenson, A. Ozcan, Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *Adv. Photonics* **1**, 046001 (2019).
46. C. Qian, X. Lin, X. Lin, J. Xu, Y. Sun, E. Li, B. Zhang, H. Chen, Performing optical logic operations by a diffractive neural network. *Light Sci. Appl.* **9**, 59 (2020).
47. N. T. Yardimci, M. Jarrahi, High sensitivity terahertz detection through large-area plasmonic nano-antenna arrays. *Sci. Rep.* **7**, 42667 (2017).
48. N. T. Yardimci, S.-H. Yang, C. W. Berry, M. Jarrahi, High-power terahertz generation using large-area plasmonic photoconductive emitters. *IEEE Trans. Terahertz Sci. Technol.* **5**, 223–229 (2015).
49. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
50. A. B. Owen, in *Contemporary Mathematics*, J. S. Verducci, X. Shen, J. Lafferty, Eds. (American Mathematical Society, 2007), vol. 443, pp. 59–71; www.ams.org/conm/443/.
51. I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, N. Navab, in *2016 Fourth International Conference on 3D Vision (3DV)* (IEEE, 2016), pp. 239–248.
52. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi, Eds. (Lecture Notes in Computer Science, Springer International Publishing, 2015), pp. 234–241.
53. D. Mengu, Y. Zhao, N. T. Yardimci, Y. Rivenson, M. Jarrahi, A. Ozcan, Misalignment resilient diffractive optical networks. *Nanophotonics* **9**, 4207–4219 (2020).
54. G. Cohen, S. Afshar, J. Tapson, A. van Schaik, EMNIST: An extension of MNIST to handwritten letters. arXiv:1702.05373 [cs.CV] (2017); http://arxiv.org/abs/1702.05373.
55. D. Mengu, Y. Rivenson, A. Ozcan, Scale-, shift-, and rotation-invariant diffractive optical networks. *ACS Photonics* **8**, 324–334 (2021).
56. O. Kulce, D. Mengu, Y. Rivenson, A. Ozcan, All-optical information processing capacity of diffractive surfaces. *Light Sci. Appl.* **10**, 25 (2020).
57. M. Veli, D. Mengu, N. T. Yardimci, Y. Luo, J. Li, Y. Rivenson, M. Jarrahi, A. Ozcan, Terahertz pulse shaping using diffractive surfaces. *Nat. Commun.* **12**, 37 (2020).
58. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG] (2014).

Acknowledgments

Funding: The Ozcan Research Group at UCLA acknowledges the support of Fujikura, Japan. We also acknowledge the support of the Burroughs Wellcome Fund. **Author contributions:** A.O., J.L., D.M., and Y.R. conceived the research. J.L., N.T.Y., and X.L. conducted the experiments. J.L., N.T.Y., X.L., and D.M. processed the data. All authors contributed to the preparation of the manuscript. A.O. supervised the research. **Competing interests:** A.O., J.L., D.M., and Y.R. are inventors on a patent application related to this work filed by University of California, Los Angeles (no. 63/022469, filed 9 May 2020). The authors declare that they have no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 10 July 2020
Accepted 10 February 2021
Published 26 March 2021
10.1126/sciadv.abd7690

Citation: J. Li, D. Mengu, N. T. Yardimci, Y. Luo, X. Li, M. Veli, Y. Rivenson, M. Jarrahi, A. Ozcan, Spectrally encoded single-pixel machine vision using diffractive networks. *Sci. Adv.* **7**, eabd7690 (2021).

Science Advances

Spectrally encoded single-pixel machine vision using diffractive networks

Jingxi Li, Deniz Mengu, Nezih T. Yardimci, Yi Luo, Xurong Li, Muhammed Veli, Yair Rivenson, Mona Jarrahi and Aydogan Ozcan

Sci Adv **7** (13), eabd7690.
DOI: 10.1126/sciadv.abd7690

Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/7/13/eabd7690/DC1

Supplementary Materials for

Spectrally encoded single-pixel machine vision using diffractive networks

Jingxi Li, Deniz Mengu, Nezih T. Yardimci, Yi Luo, Xurong Li, Muhammed Veli,
Yair Rivenson, Mona Jarrahi, Aydogan Ozcan*

*Corresponding author. Email: ozcan@ucla.edu

Published 26 March 2021, *Sci. Adv.* **7**, eabd7690 (2021)
DOI: 10.1126/sciadv.abd7690

This PDF file includes:

Supplementary Methods

Forward model of the diffractive optical network and its training. A diffractive optical network is, in general, composed of successive diffractive layers (transmissive and/or reflective) that collectively modulate the incoming object waves. According to our forward model used in this work, the diffractive layers are assumed to be thin optical modulation elements, where the i^{th} feature on the l^{th} layer at a spatial location (x_i, y_i, z_i) represents a wavelength (λ) dependent complex-valued transmission coefficient, t^l, given by:

$$t^l(x_i, y_i, z_i, \lambda) = a^l(x_i, y_i, z_i, \lambda)\exp\left(j\phi^l(x_i, y_i, z_i, \lambda)\right) \quad (S1),$$

where a and ϕ denote the amplitude and phase coefficients, respectively.

The diffractive layers are connected to each other by free-space propagation, which is modeled through the Rayleigh-Sommerfeld diffraction equation *(21, 44)*:

$$w_i^l(x, y, z, \lambda) = \frac{z - z_i}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right) \quad (S2),$$

where $w_i^l(x, y, z, \lambda)$ is the complex-valued field on the i^{th} pixel of the l^{th} layer at (x, y, z) with a wavelength of λ, which can be viewed as a secondary wave generated from the source at (x_i, y_i, z_i); and $r = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$ and $j = \sqrt{-1}$. For the l^{th} layer ($l \geqslant 1$, treating the input plane as the 0^{th} layer), the modulated optical field u^l at location (x_i, y_i, z_i) is given by

$$u^l(x_i, y_i, z_i, \lambda) = t^l(x_i, y_i, z_i, \lambda) \cdot \sum_{k \in I} u^{l-1}(x_k, y_k, z_k, \lambda) \cdot w_k^{l-1}(x_i, y_i, z_i, \lambda) \quad (S3),$$

where I denotes all the pixels on the previous diffractive layer.

We used 0.5 mm as the smallest feature size of the diffractive layers, which is mainly restricted by the resolution of our 3D-printer. To model the Rayleigh–Sommerfeld diffraction integral more accurately over a wide range of illumination wavelengths, the diffractive space was sampled with a step size of 0.25 mm so that the x and y coordinate system in the simulation window was oversampled by two times with respect to the smallest feature size. In other words, in the sampling space a 2×2 binning was performed to form an individual feature of the diffractive layers, and thus all these 4 elements share the same physical thickness, which is a learnable parameter. The printed thickness value, h, of each pixel of a diffractive layer is composed of two parts, h_m and h_{base}, as follows:

$$h = q(h_m) + h_{\text{base}} \quad (S4),$$

where h_m denotes the learnable thickness parameters of each diffractive feature and is confined between $h_{\text{min}} = 0$ and $h_{\text{max}} = 0.75$ mm. The additional base thickness, h_{base}, is a constant, non-trainable value chosen as 0.2 mm to ensure robust 3D printing and avoid bending of the diffractive layers after fabrication. Quantization operator in Eq. (S4), i.e., $q(.)$, denotes a 12-level uniform quantization (0.0625 mm for each level). To achieve the constraint applied to h_m, an associated latent trainable variable h_p was defined using the following analytical form:

$$h_m = \frac{h_{\max}}{2} \cdot \left(\sin\left(h_p\right) + 1\right) \quad (S5).$$

Note that before the training starts, h_m of all the diffractive neurons are initialized as 0.375 mm, resulting in an initial h of 0.575 mm. Based on these definitions, the amplitude and phase components of the complex transmittance of i^{th} feature of layer l, i.e., $a^l(x_i, y_i, z_i, \lambda)$ and $\phi^l(x_i, y_i, z_i, \lambda)$, can be written as a function of the thickness of each individual neuron h_i and the incident wavelength λ:

$$a^l(x_i, y_i, z_i, \lambda) = \exp\left(-\frac{2\pi\kappa(\lambda)h_i^l}{\lambda}\right) \quad (S6),$$

$$\phi^l(x_i, y_i, z_i, \lambda) = (n(\lambda) - n_{air})\frac{2\pi h_i^l}{\lambda} \quad (S7),$$

where the wavelength dependent parameters $n(\lambda)$ and $\kappa(\lambda)$ are the refractive index and the extinction coefficient of the diffractive layer material corresponding to the real and imaginary parts of the complex-valued refractive index $\tilde{n}(\lambda)$, i.e., $\tilde{n}(\lambda) = n(\lambda) + j\kappa(\lambda)$ *(43)*. Both of these parameters for the 3D-printing material used in this work were experimentally measured over a broad spectral range (see Fig. S9).

Based on this outlined optical forward model, Fig. S10 exemplifies the projection of the spatial amplitude distributions onto the x-z plane created by a diffractive network model in response to two input digits, '4' and '7'.

Based on the diffractive network layout reported in Fig. 2d of the main text, the half diffraction cone angle that enables full connectivity between the diffractive features/neurons on two successive layers is found to be ~53°. This suggests that, for a lateral feature size of 0.5 mm, the smallest wavelength that can fully utilize all the free-space communication channel between two successive layers is ~0.8 mm. Therefore, smaller wavelengths within the illumination band has a slight disadvantage in terms of layer-to-layer optical connectivity within the diffractive network. This imbalance among different spectral components of a given illumination band can be resolved in different ways: (1) using a smaller diffractive feature/neuron size through a higher-resolution fabrication method, or (2) increasing the layer-to-layer distance in the diffractive network design.

Spectral class scores. Each spectral component contained in the incident broadband terahertz beam is assumed to be a plane wave with a Gaussian lateral distribution. The beam waist corresponding to different wavelength components was experimentally measured. Although, a flat spectral magnitude (equal weight for each spectral component) was assumed during the training of the diffractive optical networks, the pulsed terahertz source used in our setup contained a different spectral profile within our band of operation. To circumvent this mismatch and *calibrate* our diffractive system (which is a one-time effort), we measured the power spectrum of the pulsed terahertz source without any objects or diffractive layers serving as our experimental reference, $I^R_{\text{exp}}(\lambda)$. In addition, we propagated through free-space the corresponding wave of each spectral component containing equal power across the entire operation band from the plane of the input aperture all the way to the output plane, forming the numerical reference wave collected by the

detector aperture, i.e., $I_{\mathrm{tr}}^{R}(\lambda)$. Based on these spectral power distributions used for calibration, the experimentally measured power spectrum, $I_{\mathrm{exp}}(\lambda)$, that is optically created by a 3D-printed diffractive optical network is normalized as:

$$I_{\mathrm{exp,\,corrected}}(\lambda) = I_{\mathrm{exp}}(\lambda) \cdot \frac{I_{\mathrm{tr}}^{R}(\lambda)}{I_{\mathrm{exp}}^{R}(\lambda)} \quad (\mathrm{S8}),$$

which corrects the mismatch between the spectral profiles assumed in the training phase and the one provided by the broadband terahertz illumination source. *In fact, this is an important practical advantage of our framework since our diffractive models can work with different forms of broadband radiation, following this calibration/normalization routine outlined above*. Figs. 3 and 5 in the main text as well as Figs. S3 and S4 illustrate the experimental spectral curves $I_{\mathrm{exp,\,corrected}}(\lambda)$ defined by Eq. (S8).

In the main text, there are two types of diffractive optical networks presented. With the number of wavelengths that we would like to encode the object information denoted by M and the number of data classes denoted by C, in the first type we assign a single wavelength to each data class, thus we take $M = C$ (e.g., $C = 10$ for MNIST data). For differential diffractive networks, on the other hand, each data class is represented by a pair of spectral components, i.e., $M = 2C$. As the dataset of handwritten digits has 10 classes, during the training of the standard diffractive optical networks, we selected 10 discrete wavelengths, each representing one digit. These wavelengths were distributed between λ_{min} = 1.00 mm and λ_{max} = 1.45 mm with 0.05 mm spacing; for the EMNIST image dataset this wavelength range was changed to be 0.825 mm to 1.45 mm with 0.025 mm spacing. For the differential diffractive optical network designs used for classifying the MNIST handwritten digits, 20 wavelengths were uniformly distributed between λ_{min} = 0.65 mm and λ_{max} = 1.6 mm; for differential designs involving EMNIST image dataset, 52 wavelengths were used, uniformly distributed between λ_{min} = 0.755 mm and λ_{max} = 1.52 mm. The first C spectral components (s_0, s_1, … , s_{C-1}) are assigned to be positive signals ($s_{0,+}$, $s_{1,+}$, … , $s_{C-1,+}$) and the subsequent C spectral components (s_C, s_{C+1}, … , s_{2C-1}) are assigned to be negative signals ($s_{0,-}$, $s_{1,-}$, … , $s_{C-1,-}$). Based on this, the differential spectral class score $\Delta s_{\boldsymbol{c}}$ for class *c* is defined as:

$$\Delta s_c = \frac{1}{T} \cdot \frac{s_{c,+} - s_{c,-}}{s_{c,+} + s_{c,-}} \quad (\mathrm{S9}),$$

where $s_{c,+}$ and $s_{c,-}$ denote the positive and negative spectral signals for the c^{th} class, respectively, and T is a non-learnable hyperparameter (also referred to as the 'temperature' hyperparameter in machine learning literature) used *only in the training phase* to improve the convergence speed and the accuracy of the final model; we empirically chose T = 0.1.

Image reconstruction neural network architecture. Our image reconstruction neural network is a 3-layer (with 2 hidden layers) fully-connected neural network, which receives an input of spectral class score vector ($\boldsymbol{s}$) and outputs a reconstructed image of the object. The 2 hidden layers have 100 and 400 neurons, respectively. The size of the 3D-printed objects used in our experiments is 2 cm × 2 cm and when they are sampled at 0.5 mm intervals, in the discrete space each input object corresponds to 40 × 40 pixels, hence the dimension of the output layer of our image

reconstruction network is 1600. Each fully connected layer of this image reconstruction ANN has the following structure:

$$\mathrm{z}_{k+1} = \mathrm{BN}\{\mathrm{LReLU}[\mathrm{FC}\{\mathrm{z}_k\}]\} \quad (\mathrm{S10}).$$

where z_k and z_{k+1} denotes the input and output of the k^{th} layer, respectively, FC denotes the fully connected layer, LReLU denotes leaky rectified linear unit, and BN is the batch normalization layer. In our architecture, LReLU is defined as:

$$\mathrm{LReLU}[x] = \begin{cases} x & \text{for x>0} \\ 0.2x & \text{otherwise} \end{cases} \quad (\mathrm{S11}).$$

For the batch normalization layer, BN, with a d-dimensional input $x = (x^{(1)}, \ldots, x^{(d)})$, each dimension of the input is first normalized (i.e., re-centered and re-scaled) using its mean μ_B and standard deviation σ_B calculated across the mini-batch B of size m, and then multiplied and shifted by the parameters $\gamma^{(k)}$ and $\beta^{(k)}$ respectively, which are both subsequently learnt during the optimization process:

$$\mathrm{BN}[x_i] = \gamma^{(k)} \cdot \frac{x_i^{(k)} - \mu_B^{(k)}}{\sqrt{{\sigma_B^{(k)}}^2 + \epsilon}} + \beta^{(k)} \quad (\mathrm{S12}).$$

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i, \quad \sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \quad (\mathrm{S13}).$$

where $k \in [1, d]$, $i \in [1, m]$ and ϵ is a small number added in the denominator for numerical stability.

Loss function for the training of spectral encoding diffractive networks. The total loss for training of diffractive optical networks, $\mathcal{L}_D$, is defined as

$$\mathcal{L}_D = \mathcal{L}_I + \alpha \cdot \mathcal{L}_E + \beta \cdot \mathcal{L}_P \quad (\mathrm{S14}),$$

where $\mathcal{L}_I$ stands for the optical inference loss, $\mathcal{L}_E$ denotes the output detector diffractive power efficiency-related loss and $\mathcal{L}_P$ denotes the spatial purity loss. The non-trainable hyperparameters, α and β, are relative weight coefficients for the corresponding loss terms. For different diffractive optical networks presented in the main text, the (α, β) pairs are set to be (0.4,0.2), (0.08,0.2), (0.03,0.1), (0,0) and (0,0) providing 84.02%, 93.28%, 95.05%, 96.07% and 96.82% optical inference accuracy, respectively (see Table 1 of the main text). For multi-class object classification, we defined $\mathcal{L}_I$ using softmax-cross-entropy (SCE) as follows:

$$\mathcal{L}_I = \mathrm{SCE}(\hat{\boldsymbol{s}}, \boldsymbol{g})$$

$$\mathcal{L}_I = -\sum_{c=1}^{C} g_c \cdot \log\left(\frac{\exp(\widehat{s_c})}{\sum_{c\prime=1}^{C}\exp(\widehat{s_{c\prime}})}\right) \quad (\mathrm{S15}),$$

where $\widehat{s_c}$, C and g_c denote the normalized spectral class score for the c^{th} class, the number of data classes, and the c^{th} entry of the ground truth label vector, respectively. For example, in the 10-wavelength diffractive optical network designs, $M = C = 10$. $\widehat{s_c}$ is calculated as:

$$\widehat{s_c} = \frac{1}{T'} \cdot \frac{s_c}{\max(\boldsymbol{s})} \tag{S16}$$

where T' is a non-learnable hyperparameter, which is used *only in the training phase* and empirically chosen as 0.1. For the differential diffractive optical network design, $\widehat{s_c}$ is equal to Δs_c defined in Eq. (S9).

The output detector diffractive power efficiency-related loss term $\mathcal{L}_E$ in Eq. (S14) is defined as:

$$\mathcal{L}_E = \begin{cases} -\log\left(\dfrac{\eta}{\eta_{\text{th}}}\right), & \text{if } \eta < \eta_{\text{th}} \\ 0, & \text{if } \eta \geq \eta_{\text{th}} \end{cases} \tag{S17},$$

where ***η denotes the diffractive power efficiency at the output detector*** and η_{th} refers to the penalization threshold that was taken as 0.015 during the training phase. η is defined as:

$$\eta = \frac{I_{c_{\text{GT}}}}{I_{\text{in}_{\text{GT}}}} \tag{S18},$$

where $I_{c_{\text{GT}}}$ represents the power (calculated at the output detector aperture) for the spectral component corresponding to the ground truth class of the input object, and $I_{\text{in}_{\text{GT}}}$ represents the power of the same spectral component right after the input aperture, *before* the object and the diffractive network.

The spatial purity loss $\mathcal{L}_P$ is used to clear the optical power over a small region of interest, 1 cm × 1 cm surrounding the active area of the single-pixel detector, for the purpose of decreasing the sensitivity of the diffractive optical network to potential misalignment of the detector in the transverse plane with respect to the optical axis. $\mathcal{L}_P$ is calculated using:

$$\mathcal{L}_P = -\sum_{c=1}^{\text{C}} \log\left(\frac{I_{\text{detector},\, c}}{I_{\text{peripheral},\, c} + I_{\text{detector},\, c}}\right) \tag{S19},$$

where $I_{\text{detector},\, c}$ and $I_{\text{peripheral},\, c}$ denote the optical power of the c^{th} spectral component collected by the active area of the output detector and within a 1 cm × 1 cm periphery around the output detector aperture, respectively.

Loss function for the training of image reconstruction (decoder) networks. Total loss of an electronic image reconstruction network, $\mathcal{L}_{\text{Recon}}$, is defined as:

$$\mathcal{L}_{\text{Recon}} = \gamma \cdot \mathcal{L}_S\left(\boldsymbol{O}_{\text{recon}}, \boldsymbol{O}_{\text{input}}\right) + (1 - \gamma) \cdot \mathcal{L}_I{}' \tag{S20},$$

where $\mathcal{L}_S$ stands for the pixel-wise *structural loss* between the reconstructed image of the object $\boldsymbol{O}_{\text{recon}}$ and the ground truth object structure $\boldsymbol{O}_{\text{input}}$. $\mathcal{L}_I{}'$ is the same loss function as $\mathcal{L}_I$ defined in

Eq. (S15); except, instead of $\hat{\boldsymbol{s}}$, it computes the loss $\mathrm{SCE}(\widehat{\boldsymbol{s}'}, \boldsymbol{g})$ using $\widehat{\boldsymbol{s}'}$ and ground truth label vector $\boldsymbol{g}$. Here, $\widehat{\boldsymbol{s}'}$ denotes the new class scores computed by cycling $\boldsymbol{O}_{\mathrm{recon}}$ back to the object plane of the diffractive optical network model at hand and numerically propagating it through the optical forward model as depicted in Fig. 4 of the main text. The hyperparameter, γ, is a coefficient that controls the ratio of two loss terms. In the training phase, γ was empirically chosen as $\gamma = 0.95$. Two kinds of image structural loss terms, i.e., Mean Absolute Error (MAE) loss and reversed Huber (or "BerHu") loss, are used. Reversed Huber loss between 2D images a and b is defined as *(50, 51)*:

$$BerHu(\boldsymbol{a}, \boldsymbol{b}) = \sum_{\substack{x,y \\ |a(x,y)-b(x,y)|\leq q}} |a(x,y) - b(x,y)| + \sum_{\substack{x,y \\ |a(x,y)-b(x,y)|>q}} \frac{[a(x,y) - b(x,y)]^2 + q^2}{2q} \quad (S21),$$

where q is a hyperparameter that is empirically set as 20% of the standard deviation of the normalized input ground truth image. Examples of the reconstructed images resulting from the decoder ANN models trained using these different loss terms are shown in Figs. S6 and S7, where the loss curves of these models are also reported.

Joint-training of a diffractive optical network and an image reconstruction ANN.

Figures S11 and S12 illustrate some test examples for the jointly-trained, hybrid machine vision systems that are based on the presented diffractive spectral encoding framework. These hybrid systems consist of a diffractive optical front-end and an image reconstruction ANN at the back-end that were trained simultaneously, communicating with each other through error back-propagation. Similar to the machine vision systems presented in the main text, we have an image reconstruction ANN constituting the electronic part of the hybrid system and there is a 3-layer diffractive optical front-end, encoding the spatial object information into the power spectrum of the light collected by the single-pixel detector at the output plane. For the training of these hybrid systems, we used the following loss function,

$$\mathcal{L}_H = \xi \cdot \mathcal{L}_I + \mathcal{L}_S(\boldsymbol{O}_{\mathrm{recon}}, \boldsymbol{O}_{\mathrm{input}}) \quad (S22),$$

where, $\mathcal{L}_I$ and $\mathcal{L}_S$ represent the all-optical class inference loss in Eq. (S15) and the image reconstruction loss in Eq. (S21), respectively. The constant multiplicative coefficient, ξ, determines the relative contribution of the all-optical classification loss to the total loss, $\mathcal{L}_H$. For instance, when ξ is set to be zero, the diffractive front-end entirely ignores the classification accuracy and, thus, the hybrid machine vision system acts purely as a single-pixel computational imaging platform. On the other hand, for $\xi > 0$, the diffractive optical network evolution takes both the all-optical classification and the subsequent image recovery tasks into account due to the linear superposition of two losses in Eq. (S22), which enables the joint-training of the presented hybrid machine vision systems as depicted in Figs. S11 and S12.

Supplementary Table S2 summarizes the optical classification accuracies of different hybrid networks designed around the presented diffractive spectral encoding framework. We can categorize all the hybrid network systems investigated here into three main categories based on

the number of wavelengths that are simultaneously processed by the diffractive front-end. The diffractive network systems in the first category encode the entire spatial information of the sample field-of-view into the power levels of 10 discrete wavelengths (see the first 4 rows of Supplementary Table S2), similar to the diffractive networks depicted in Figs. 3, 5, S3, S4 and S5. The main difference of the results reported in Supplementary Table S2 compared to those earlier systems (Figs. 3, 5, S3, S4 and S5) is that these new diffractive networks were trained jointly with the image reconstruction ANN at the back-end, and therefore, in addition to the classification accuracy, they also try to help the electronic ANN to achieve the task of image reconstruction. Therefore, compared to the diffractive networks that were trained solely towards the task of image classification, these networks achieve slightly lower classification accuracies since they are also penalized during their training if the image reconstruction ANN fails (see Table 1 and Supplementary Table 2).

In the second category, the diffractive optical networks use 20, i.e., $M = 2C$, wavelengths and the classification decision is made based on the normalized differential signal between each pair of wavelengths representing a data class. However, the input to the image reconstruction ANN is not the differential class scores ($\Delta\boldsymbol{s}$), but instead, it is the entire power spectrum vector, $\boldsymbol{s}$, of length 20. The third category that we examined in Supplementary Table S2 controls/processes 50 unique wavelengths, i.e., $M = 5C$. Therefore, in this latter case, the input of the reconstruction ANN, $\boldsymbol{s}$, is a vector of length 50. Since for the MNIST dataset $C = 10$, the forward training model groups the power coefficients into C different subgroups (each with 5 wavelengths) and averages the detected power for each subgroup to construct, $\boldsymbol{s^{\mu}}$. Then, the class scores are determined based on Eq. (S16), by replacing s_c with s_c^{μ}. In Supplementary Table S2, as another variant of jointly-trained hybrid networks that utilize $M = 5C$ wavelengths, we also utilized multiplicative, learnable weights, $\boldsymbol{w}$, for all $M = 5C$ wavelengths, enabling our forward model to learn the relative contributions of each wavelength as this scheme promotes weighted averaging of the spectral subgroups.

Supplementary Table S2 summarizes the results of this comparative analysis, corresponding to these three categories of jointly-trained hybrid network systems as detailed above. During the training of these hybrid networks, the optical forward models of the first and second categories of diffractive front-ends target the same wavelength ranges, (1.00-1.45) mm and (0.65-1.6) mm, respectively, same as the nondifferential and differential all-optical classification systems presented in Table 1. For the third category, i.e., hybrid network systems that process $M = 5C$ wavelengths, we used a wavelength range of (0.45-1.50) mm with a step size of 0.021 mm. This corresponds to a data class spacing of 0.105 mm (5×0.021 mm) between the sub-bands representing successive/neighboring data classes.

Vaccination of single-pixel diffractive machine vision systems against misalignments

The deep learning-based training of the spectral encoder diffractive networks presented throughout the main text and the Supplementary Materials, assumes ideal physical implementation conditions, including e.g., perfect opto-mechanical alignment and fabrication of the diffractive network. This is, in general, not the case for physically fabricated platforms, which contributes to the mismatch

between the numerically predicted and the experimentally measured output power spectra for a given input object (see e.g., Supplementary Figure S3).

Spectral encoding diffractive optical networks and related hybrid network systems can adapt to undesired physical implementation errors such as misalignments, provided that these random error sources are taken into account as part of their forward training model. In this work, we extended the methods presented in *(53)* to our machine vision framework and tested their efficacy for the vaccination of single-pixel broadband diffractive optical networks against lateral misalignments of diffractive layers. Towards this end, we have chosen the jointly-trained hybrid classification systems as our testbed (with 10 unique wavelengths, each representing one MNIST data class). Each one of these hybrid network systems consists of a 3-layer diffractive network (see Fig. 2 of the main text) that is jointly-trained with a *shallow classification* ANN (2 hidden layers). The results of this analysis are presented in Supplementary Fig. S15 and discussed in the main text, Discussion section.

Training-related details. Both the diffractive optical networks and the corresponding decoder ANNs used in this work were simulated and trained using Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.). We selected Adam *(58)* optimizer during the training of all the models, and its parameters were taken as the default values in TensorFlow and kept identical in each model. The learning rate was set as 0.001. The MNIST handwritten digit image data are divided into three parts: training, validation and testing, which contain 55,000, 5,000 and 10,000 images, respectively. The handwritten capital letter image data are extracted from the EMNIST-balanced dataset *(54)* and divided into training, validation and testing sets, which contain 57,200, 5,200 and 10,400 images, respectively. During the training of the diffractive models, the feeding sequence of the image data was randomized, and therefore if the same training process is independently run for multiple times, slightly different diffractive designs will be achieved (see Supplementary Fig. S16 for an example). In this article, all the diffractive optical networks were trained for 50 epochs and the best models were selected based on the classification performance on the validation data set. All the image reconstruction ANNs were trained for 20 epochs. In Fig. S1, we present two different training schemes for image reconstruction ANNs. If there is no feedback cycle, i.e., $\gamma = 1$ in Eq. (S20), the remaining loss factor is the structural loss, $\mathcal{L}_S\left(\boldsymbol{O}_{\text{recon}}, \boldsymbol{O}_{\text{input}}\right)$. In this case, the best ANN model was selected based on the minimum loss value over the validation data set. If there was an image feedback cycle, i.e., $\gamma < 1$ in Eq. (S20), the best ANN model was selected based on the classification performance provided by $\widehat{\boldsymbol{s}'}$ over the validation set.

For the training of our models, we used a desktop computer with a TITAN RTX graphical processing unit (GPU, Nvidia Inc.) and Intel® Core™ i9-9820X central processing unit (CPU, Intel Inc.) and 128 GB of RAM, running Windows 10 operating system (Microsoft Inc.). For the diffractive optical front-end design involving $M = C = 10$, the batch size was set to be 4 and 5 for the diffractive network and the associated image reconstruction networks, respectively. However, for the differential design of the diffractive optical front-end with $M = 2C = 20$, the batch size was set to be 2 and 5 during the training of the diffractive network and the associated image reconstruction network, respectively. The main limiting factor on these batch size selections is the GPU memory of our computer. The typical training time of a diffractive optical network with $C =$

10 is ~80 hours. The typical training time of an image reconstruction decoder ANN with and without the image feedback/collaboration loop is ~20 hours and ~2 hours, respectively.

Supplementary Figures

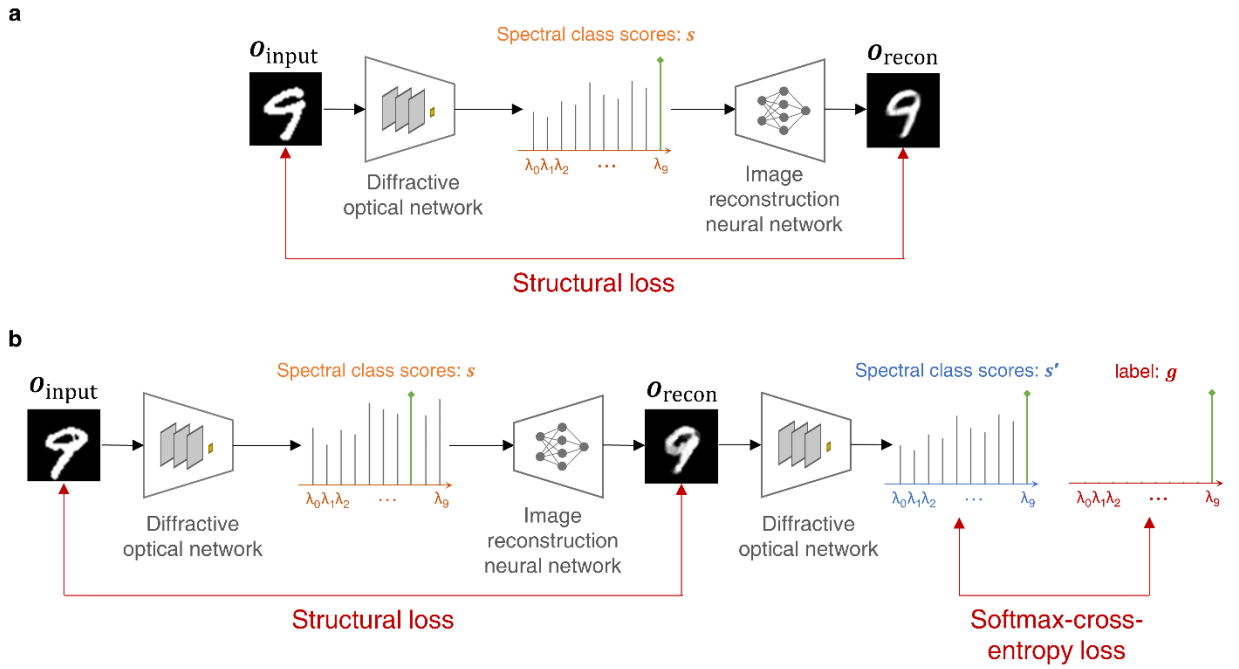

Fig. S1. Different strategies for training an image reconstruction ANN to decode spectral encoding. a, Training strategy for image reconstruction ANN based on a structural loss function that pixel-wise compares the reconstructed image, $\boldsymbol{O}_{\text{recon}}$, with the ground truth $\boldsymbol{O}_{\text{input}}$. **b**, Application of the image feedback mechanism used for tailoring the image reconstruction space of the decoder ANN in order to collaborate with the corresponding diffractive optical network and improve its optical image classification.

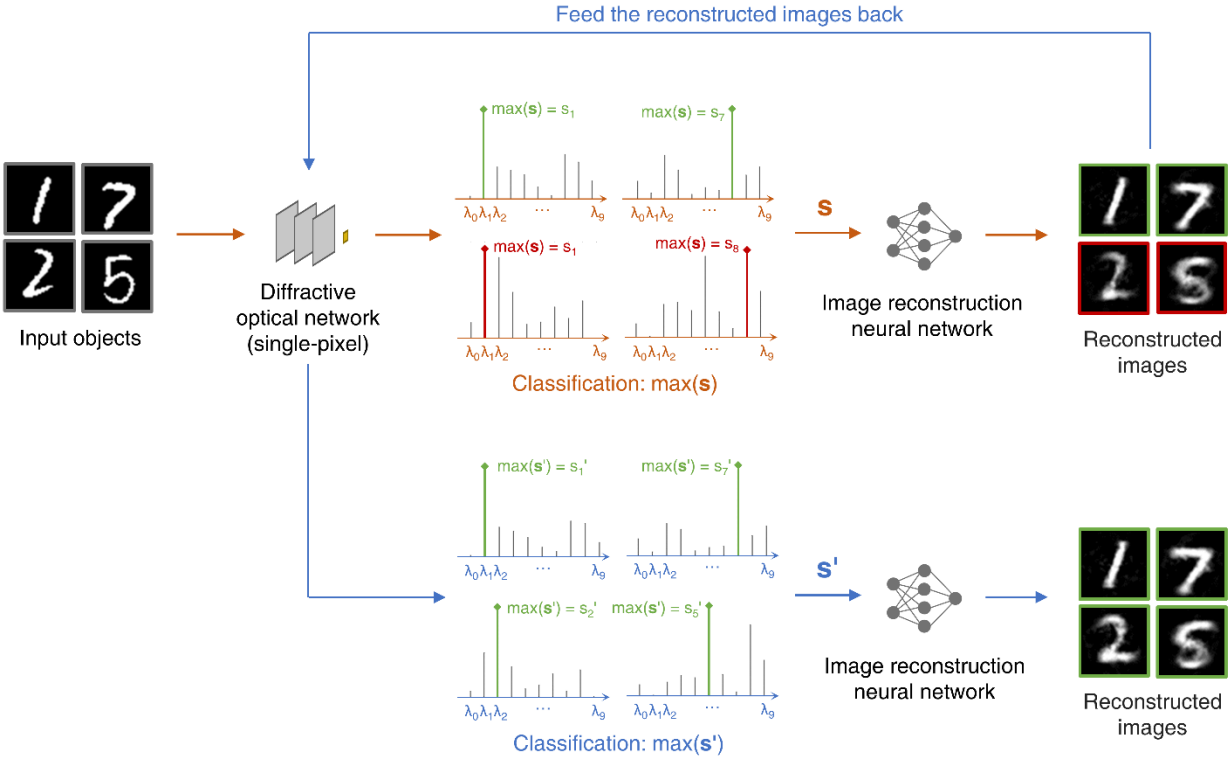

Fig. S2. Illustration of the collaboration between the image reconstruction decoder ANN and the spectral encoding diffractive optical network. Same as in Fig. 4 of main text, except for the diffractive network design reported in the 3rd row of Table 1, main text. The blind testing accuracy of this diffractive optical network for handwritten digit classification increased from 95.05% to 95.37% using this image feedback loop (see Figs. 3c, 5b and Table 1 of the main text). The image reconstruction ANN was trained using the BerHu loss defined in Eq. (S21) and softmax-cross-entropy loss reported in Eq. (S15).

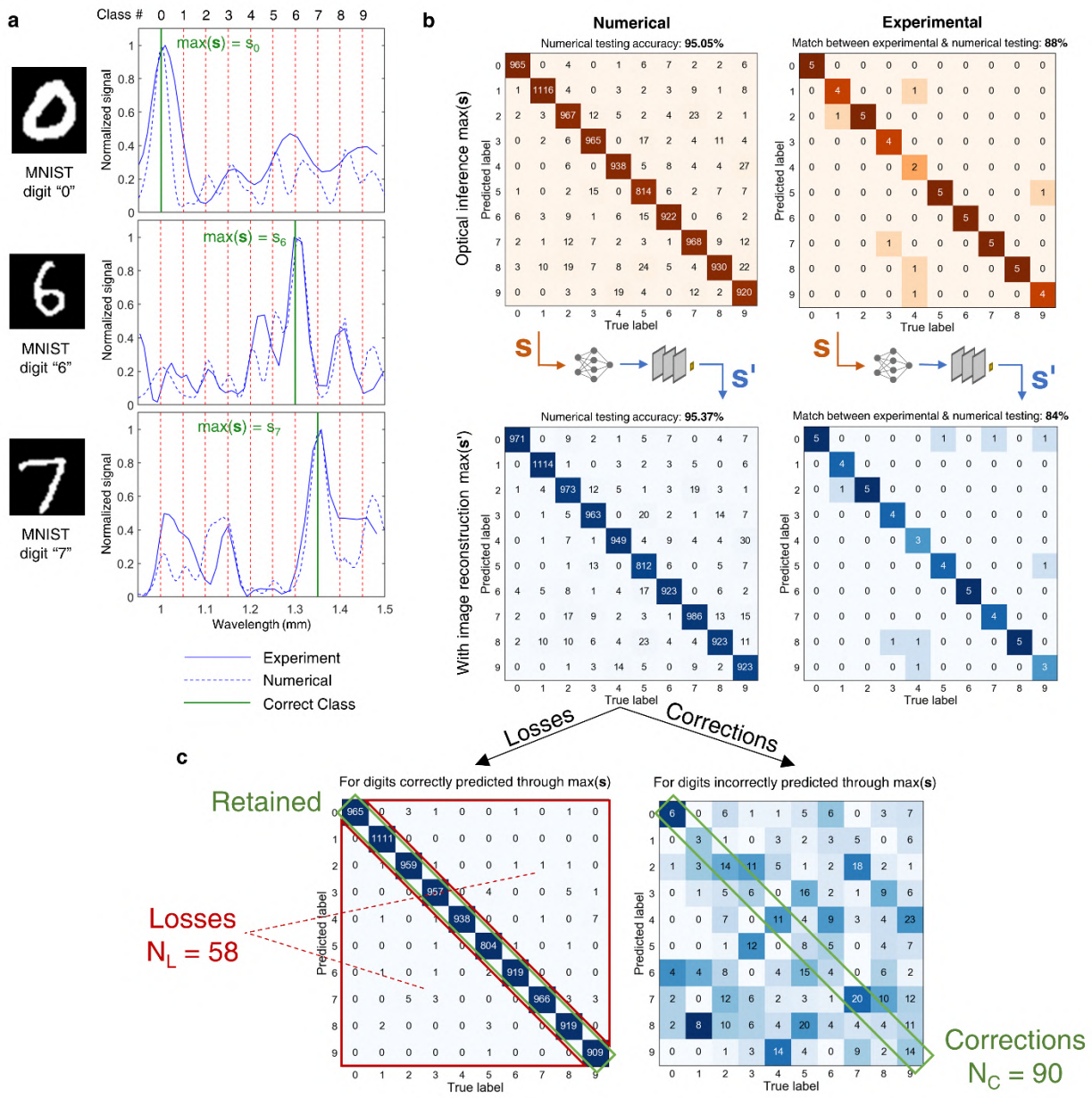

Fig. S3. Blind testing performance of a diffractive optical network using spectral encoding and its coupling with a corresponding decoder ANN. a, Experimentally measured (blue-solid line) and the numerically computed (blue-dashed line) output power spectra for optical classification of three different handwritten digits, shown as examples. **b**, **Top Left:** confusion matrix summarizing the numerical classification performance of the diffractive optical network that attains a classification accuracy of 95.05% over 10,000 handwritten digits in the blind testing image set. **Top Right:** confusion matrix for the experimental results obtained by 3D-printing 50 handwritten digits randomly selected from the numerically successful classification samples in the blind testing set. An 88% match between experimentally inferred and numerically computed object classes is observed. **Bottom Left:** confusion matrix provided by max($\boldsymbol{s}'$) computed by feeding the reconstructed images back to the diffractive network. A blind testing accuracy of 95.37% is achieved, demonstrating a classification accuracy improvement of 0.32%. **Bottom Right:** confusion matrix for the experimental results using the same 50 digits. **c**, **Left:** same as the bottom

left matrix in (**b**), but solely for the digits that are correctly predicted by the optical network. Its diagonal entries can be interpreted as the digits that are retained to be correctly predicted, while its off-diagonal entries represent the "losses" after the image reconstruction and feedback process. **Right:** same as the left one, but solely for the digits that are incorrectly classified by the optical network. Its diagonal entries indicate the classification "corrections" after the image reconstruction and feedback process. The number N_C - N_L = 32 is the classification accuracy "gain" achieved through $\max(\boldsymbol{s}')$, corresponding to a 0.32% increase in the numerical testing accuracy of the diffractive model (also see Fig. 3c of the main text).

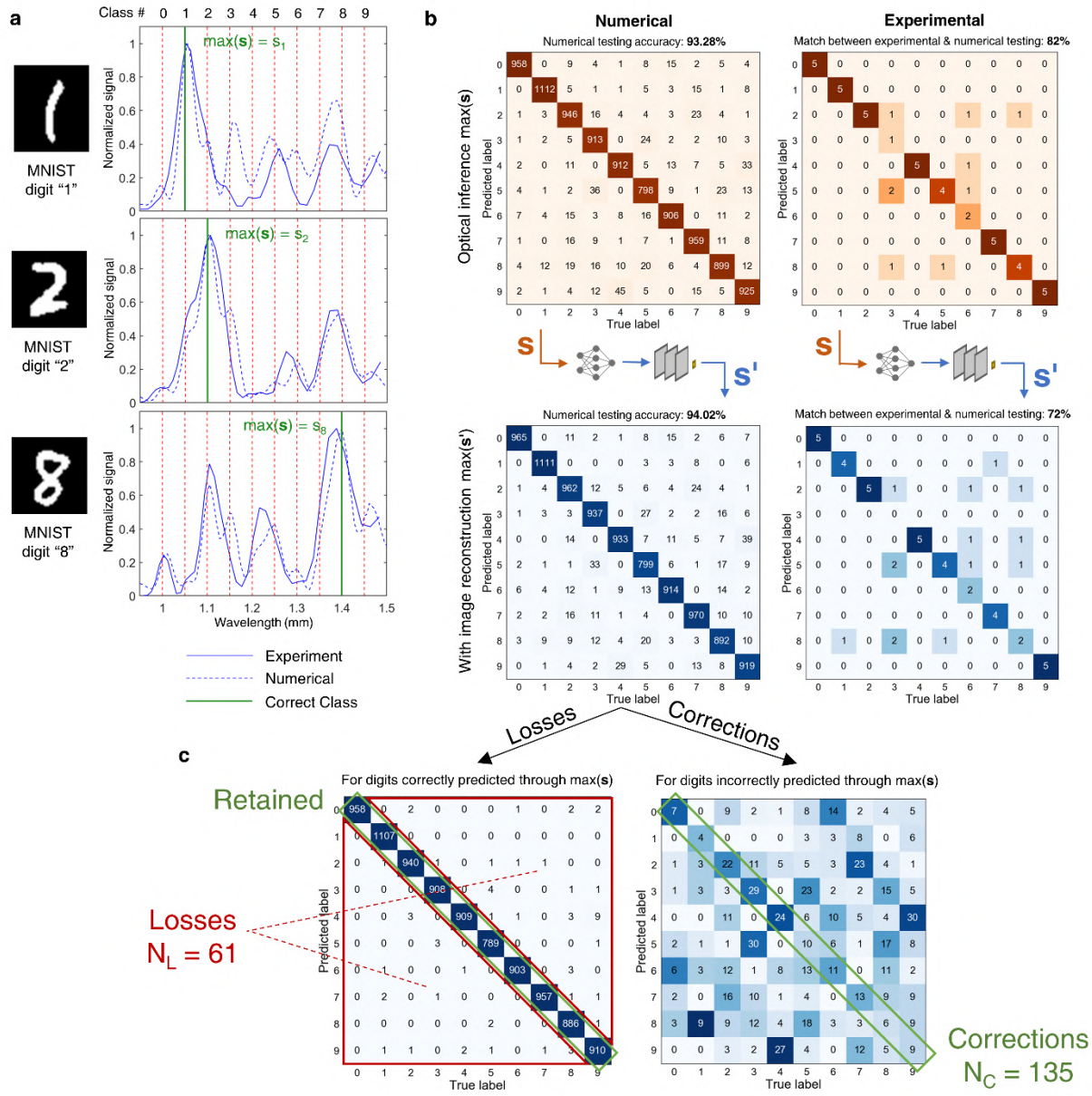

Fig. S4. Blind testing performance of a diffractive optical network using spectral encoding and its coupling with a corresponding decoder ANN. Same as in Fig. S3, except for the diffractive network design reported in the 2nd row of Table 1, main text. The number N_C - N_L = 74 is the classification accuracy "gain" achieved through max($\boldsymbol{s}'$), corresponding to a 0.74% increase in the numerical testing accuracy of the diffractive model (also see Fig. 3c of the main text).

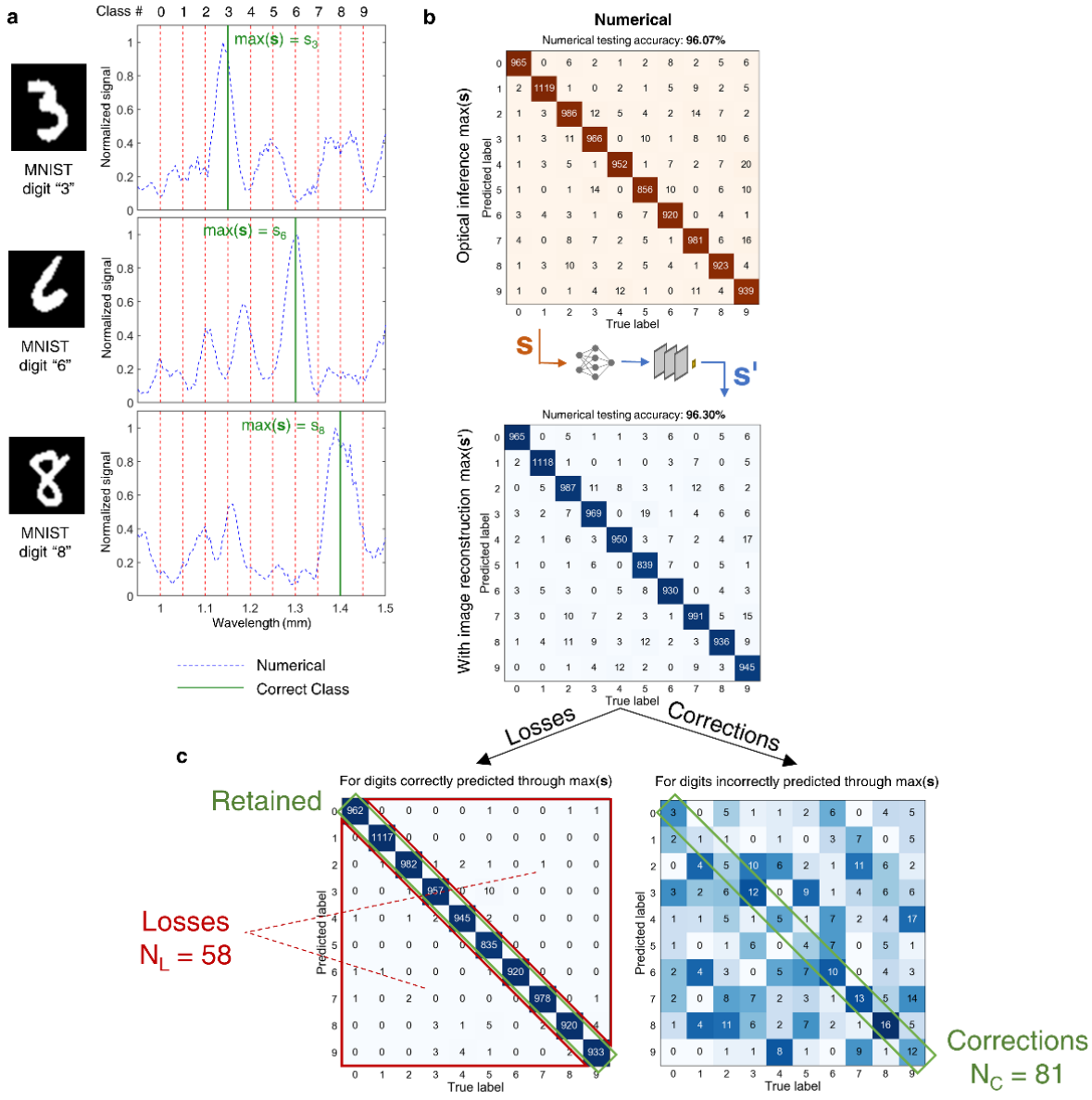

Fig. S5. Blind testing performance of a diffractive optical network using spectral encoding and its coupling with a corresponding decoder ANN. Same as in Fig. S3, except for the diffractive network design reported in the 4th row of Table 1, main text. The number N_C - N_L = 23 is the classification accuracy "gain" achieved through max($\boldsymbol{s}'$), corresponding to a 0.23% increase in the numerical testing accuracy of the diffractive model.

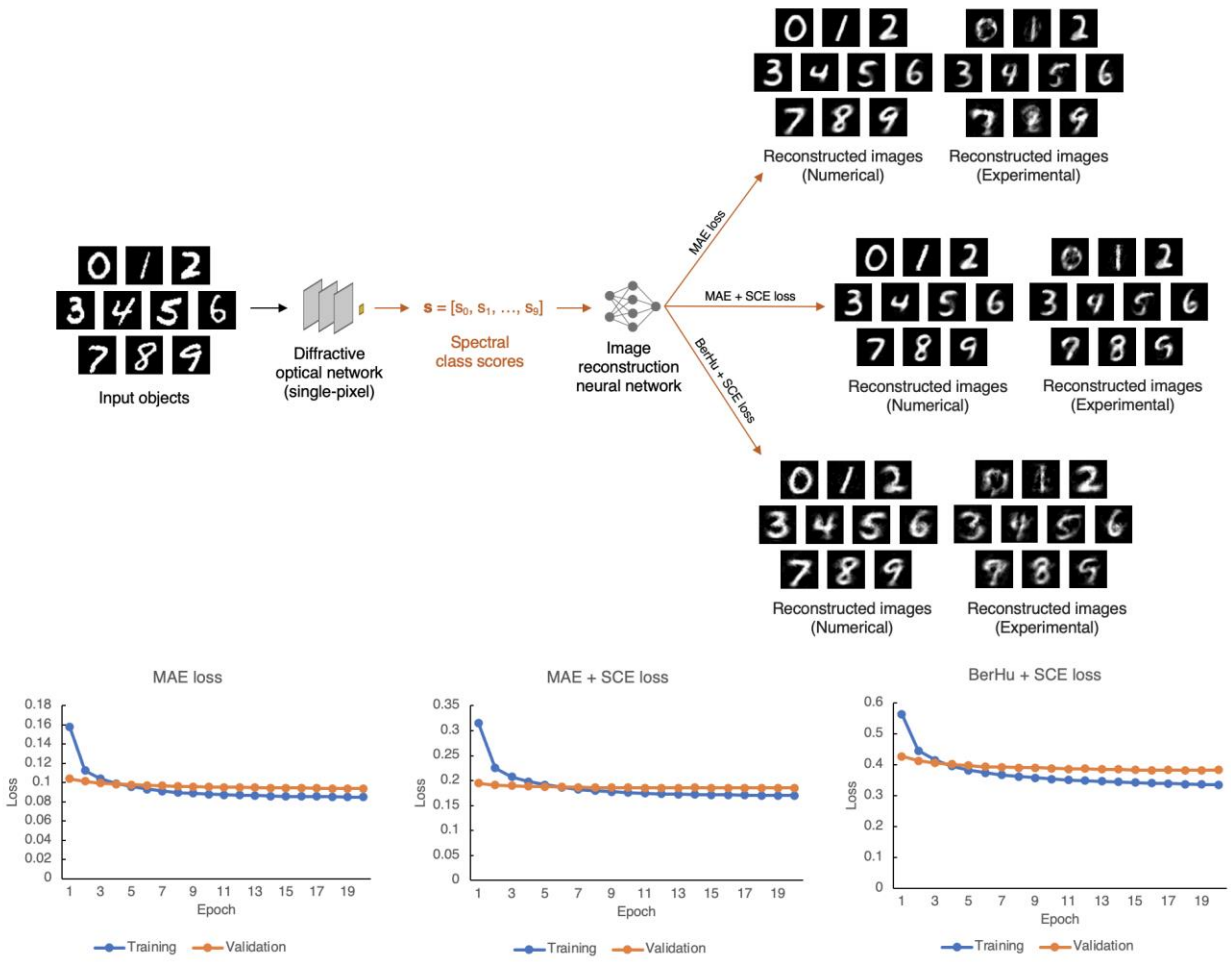

Fig. S6. Image reconstruction results using decoder ANNs trained with different loss functions. (Top) The images of the input objects reconstructed from both numerically predicted and experimentally measured spectral class scores, $\boldsymbol{s}$, are shown. The diffractive optical network that provides a blind testing accuracy of 95.05% is used here (3rd row of Table 1, main text). Each image of a handwritten digit is composed of >780 pixels, and the shallow image reconstruction ANN with 2-hidden layers receives an input vector size of 10 ($\boldsymbol{s} = [s_0, s_1, \ldots, s_9]$) to successfully decode the task-specific spectral encoding of the diffractive optical network. (Bottom) The training and validation loss curves of the decoder ANNs trained with different loss functions.

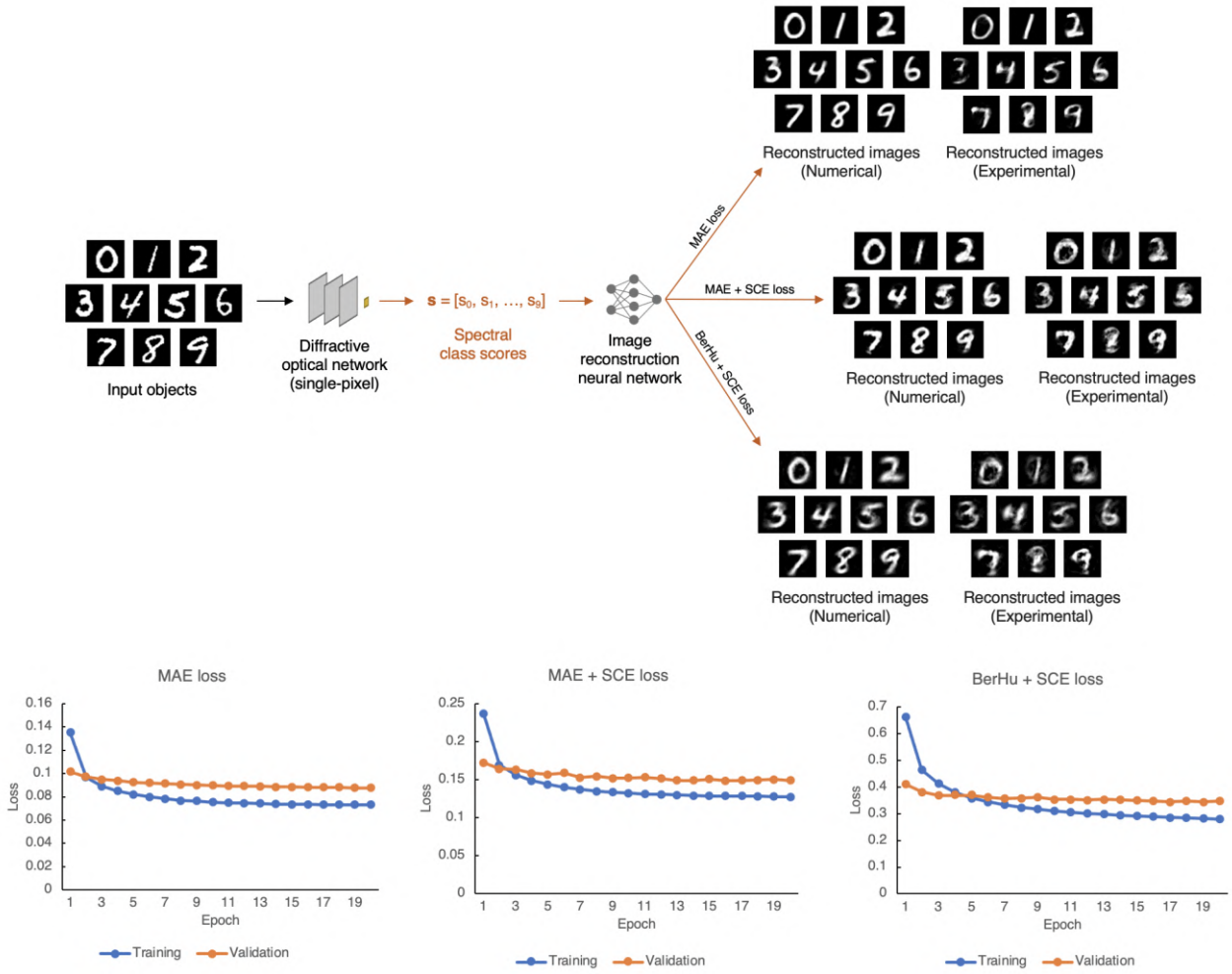

Fig. S7. Image reconstruction results using decoder ANNs trained with different loss functions. (Top) Same as the top panel in Fig. S6, except that the diffractive optical network that provides a blind testing accuracy of 84.02% is used here (1st row of Table 1, main text). Each image of a handwritten digit is composed of >780 pixels, and the shallow image reconstruction ANN with 2-hidden layers receives an input vector size of 10 ($\boldsymbol{s} = [s_0, s_1, \ldots, s_9]$) to successfully decode the task-specific spectral encoding of the diffractive optical network. (Bottom) The training and validation loss curves of the decoder ANNs trained with different loss functions.

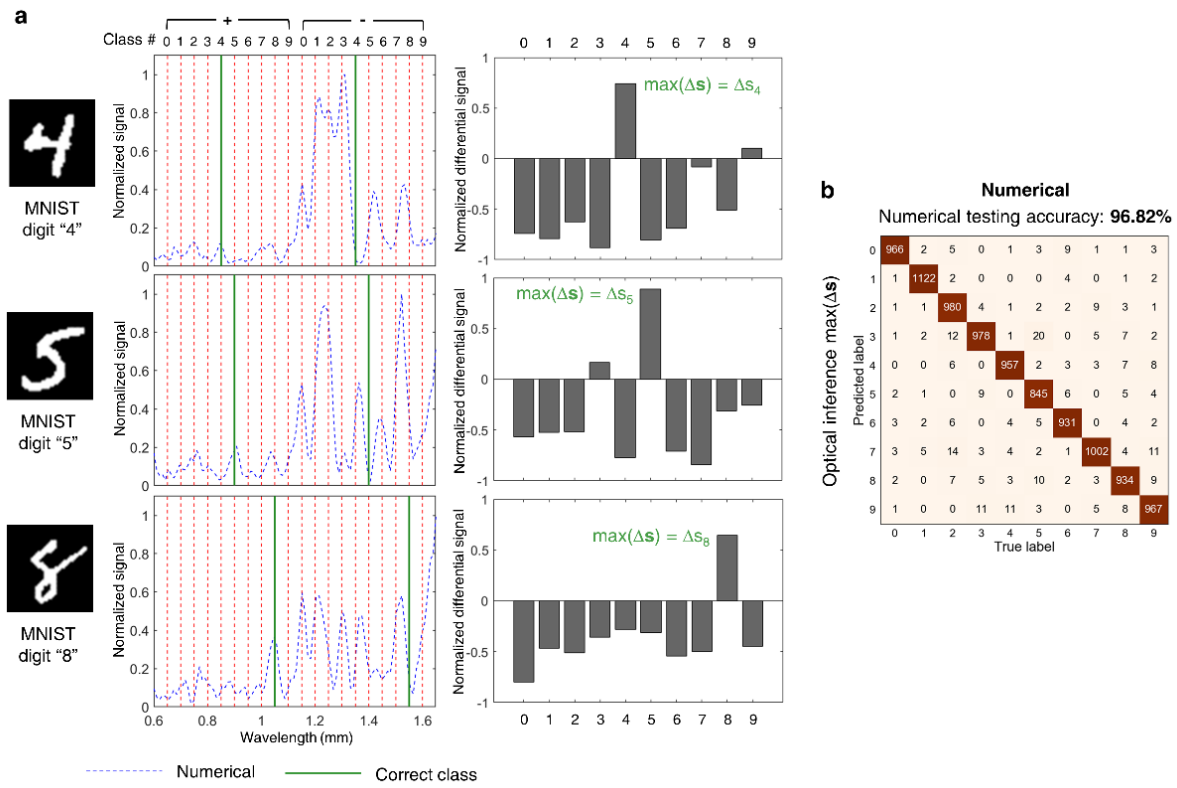

Fig. S8. Blind testing performance of a broadband diffractive network using *differential spectral encoding* of data classes. a, The numerically computed (blue-dashed lines) output power spectra for optical classification of three different handwritten digits (shown as examples) and their normalized differential class scores, $\Delta \boldsymbol{s}$. **b,** Confusion matrix summarizing the numerical classification performance of the diffractive optical network using differential spectral encoding strategy (max(**Δs**)) that attains a classification accuracy of 96.82% over 10,000 handwritten digits in the blind testing set.

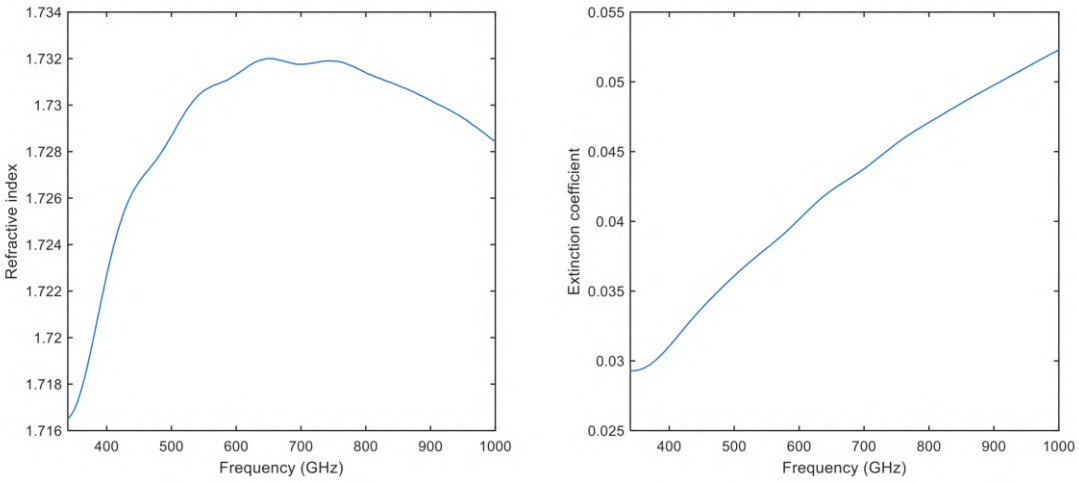

Fig. S9. Dispersion curves of the polymer material used for 3D fabrication of our diffractive optical network models. The refractive index (left) and the extinction coefficient (right) of VeroBlackPlus RGD875 printing material were extracted from the real and imaginary parts of the complex refractive index, respectively, and used for $n(\lambda)$ and $\kappa(\lambda)$ in Eqs. (S6) and (S7).

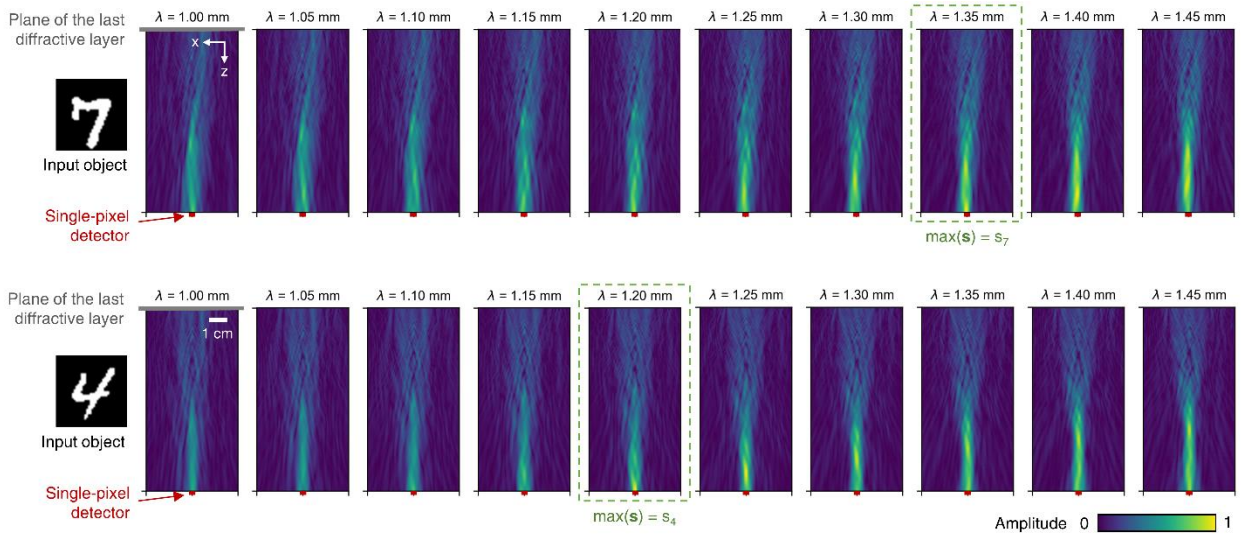

Fig. S10. The projection of the spatial amplitude distributions created by 2 input objects (handwritten digits) on the x-z plane (y = 0, where the detector is located) for 10 different wavelengths that encode the spectral class scores, *s*. The z range is from the plane of the last diffractive layer (grey lines) to the plane of the detector (small red rectangles), and the diffractive model used here is the one that has an 84.02% numerical testing accuracy (see Fig. 5 and Fig. S7). The spectral class scores (***s***) can be viewed as the integral of the light intensity at the detector region in each case and the maximum one predicts the correct data class (green dashed box for each row).

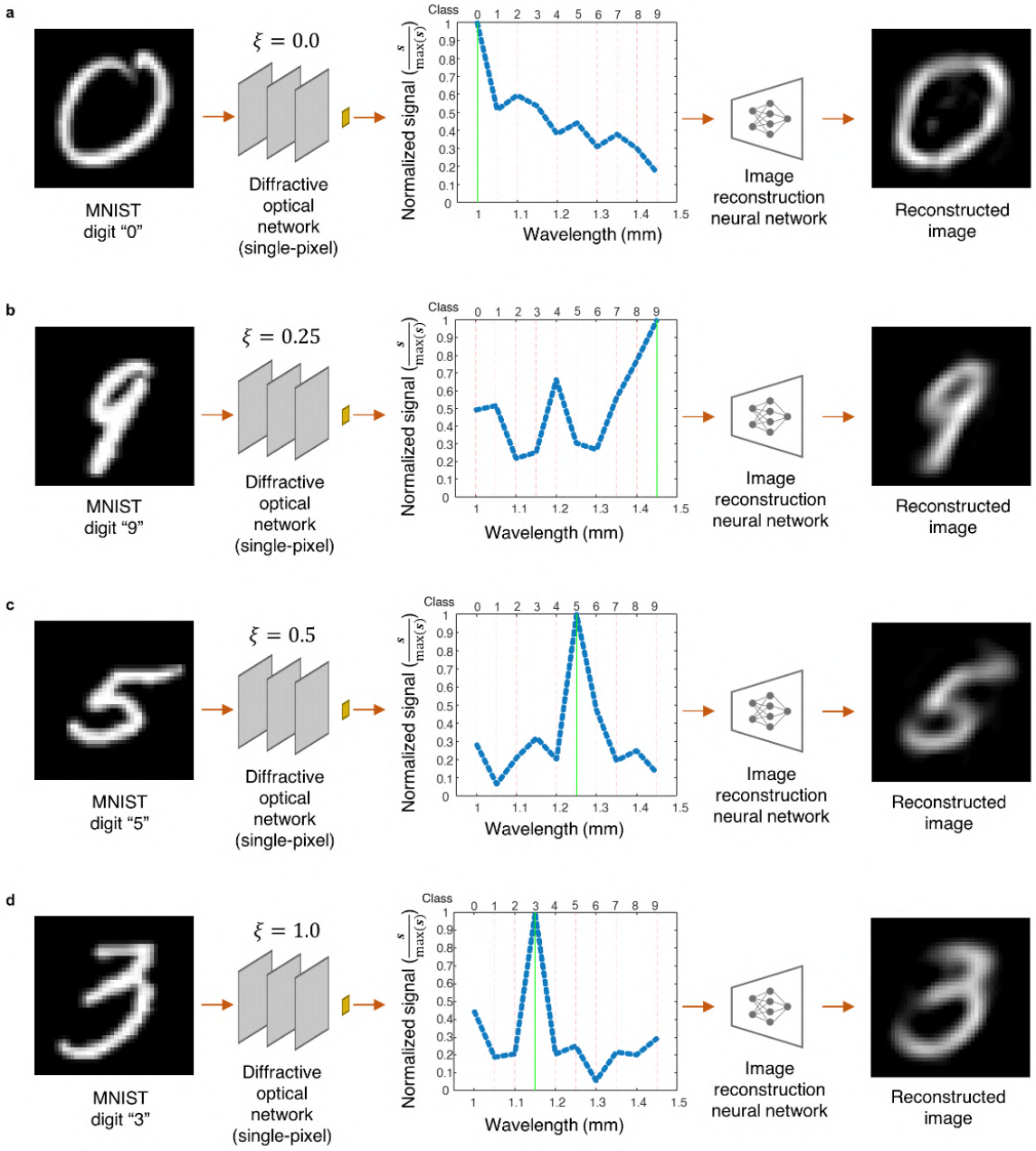

Fig. S11. Jointly-trained hybrid machine vision systems for all-optical image classification and ANN-based image reconstruction using spectral encoding of data classes through a single-pixel. a, $\xi = 0$. **b,** $\xi = 0.25$. **c,** $\xi = 0.5$. **d,** $\xi = 1.0$. The input object is an amplitude-encoded MNIST digit and it propagates through the diffractive spectral encoder; its data class is blindly inferred by *max(s)* at the output of a single-pixel spectroscopic detector. The collected power coefficients of 10 wavelengths are further processed by the image reconstruction ANN to blindly recover the image of the input object. Also see Supplementary Table S2.

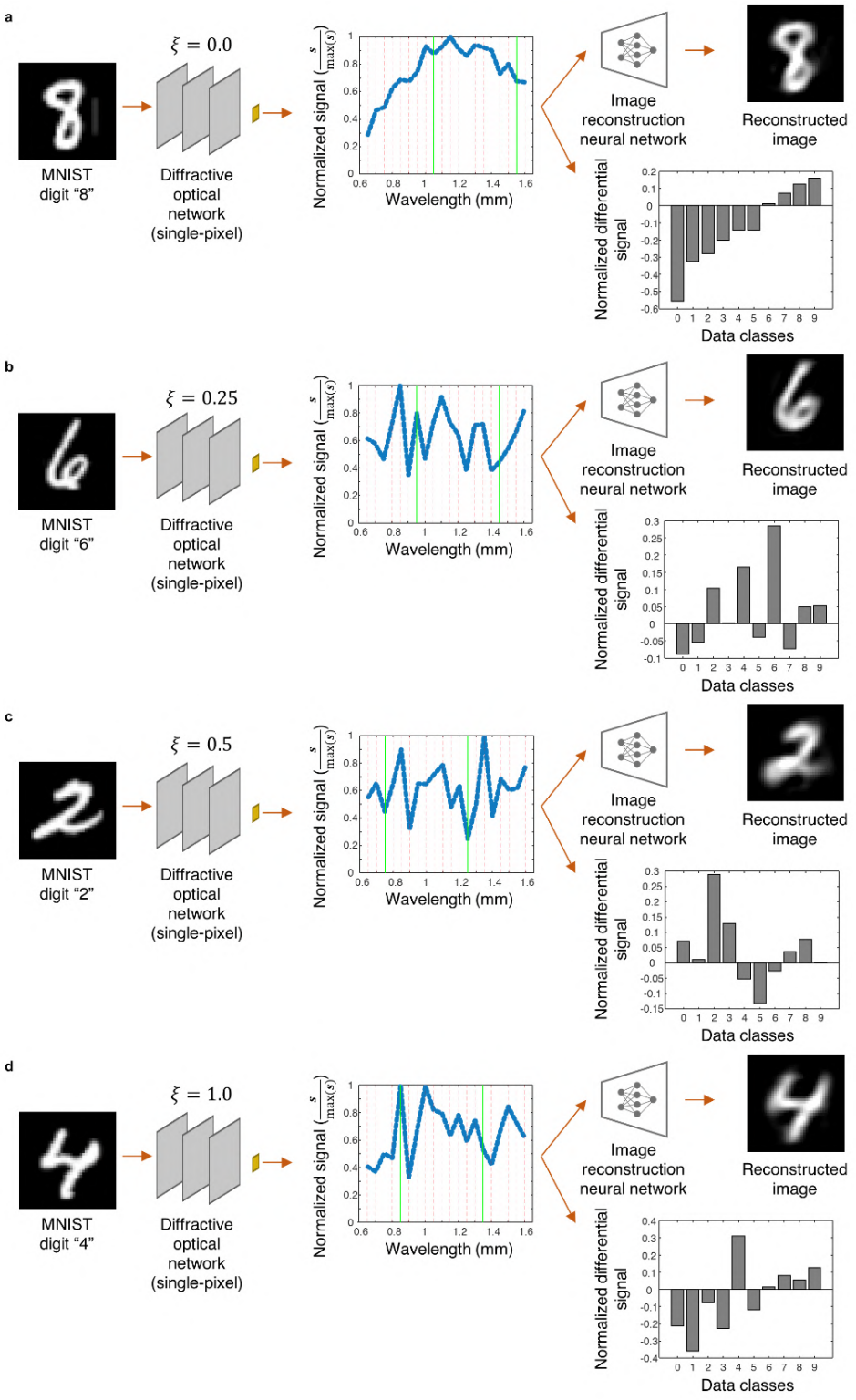

Fig. S12. Jointly-trained hybrid machine vision systems for all-optical image classification and ANN-based image reconstruction using *differential* spectral encoding of data classes through a single-pixel. a, $\xi = 0$. **b,** $\xi = 0.25$. **c,** $\xi = 0.5$. **d,** $\xi = 1.0$. The input object is an amplitude-encoded MNIST digit and it propagates through the diffractive spectral encoder; its data

class is blindly inferred by a normalized differential signal, $max(\boldsymbol{\Delta s})$, at the output of a single-pixel spectroscopic detector. The collected power coefficients of $M = 20$ wavelengths are further processed by the image reconstruction ANN to blindly recover the image of the input object. For $\xi = 0$ shown in (a), the jointly-trained model completely focused on the image reconstruction task, and ignored the optical classification loss, which resulted in a poor classification accuracy, unlike the other cases shown here. Also see Supplementary Table S2.

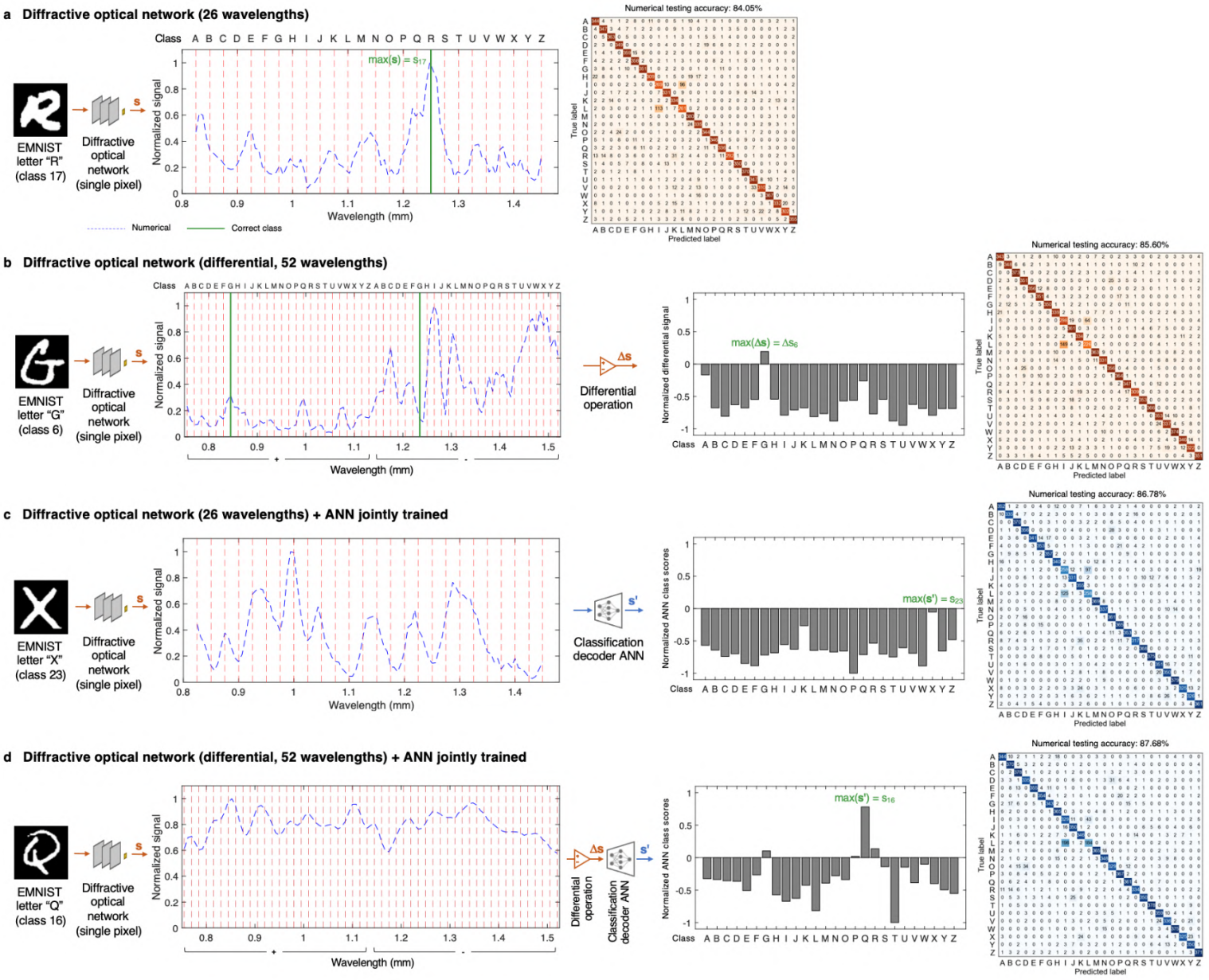

Fig. S13. EMNIST image classification results. a, the numerically computed output power spectrum for optical classification of one handwritten capital letter (shown as an example) using a trained standard diffractive optical network which performs space-to-spectrum encoding based on 26 discrete wavelengths, achieving a blind testing accuracy of 84.05%. **b**, same as in (a), but a differential diffractive network that performs differential class encoding based on 52 discrete wavelengths is used, which achieved a classification accuracy of 85.60%. The wavelengths are divided into two groups, one containing 26 wavelengths for positive ("+") spectral signals while the other one for negative ("−") spectral signals. The computed differential class scores of the sample object, $\Delta \boldsymbol{s}$, are normalized and shown in the bar chart. **c**, same as in (a), but a back-end decoder ANN is introduced and jointly-trained with the front-end diffractive network, which helped to achieve a classification accuracy of 86.78%. The bar chart shows the normalized final class scores, $\boldsymbol{s}'$, given by the output of the decoder ANN. **d**, same as (c), but the standard diffractive

network is replaced with a differential diffractive network design, and a classification accuracy of 87.68% is achieved.

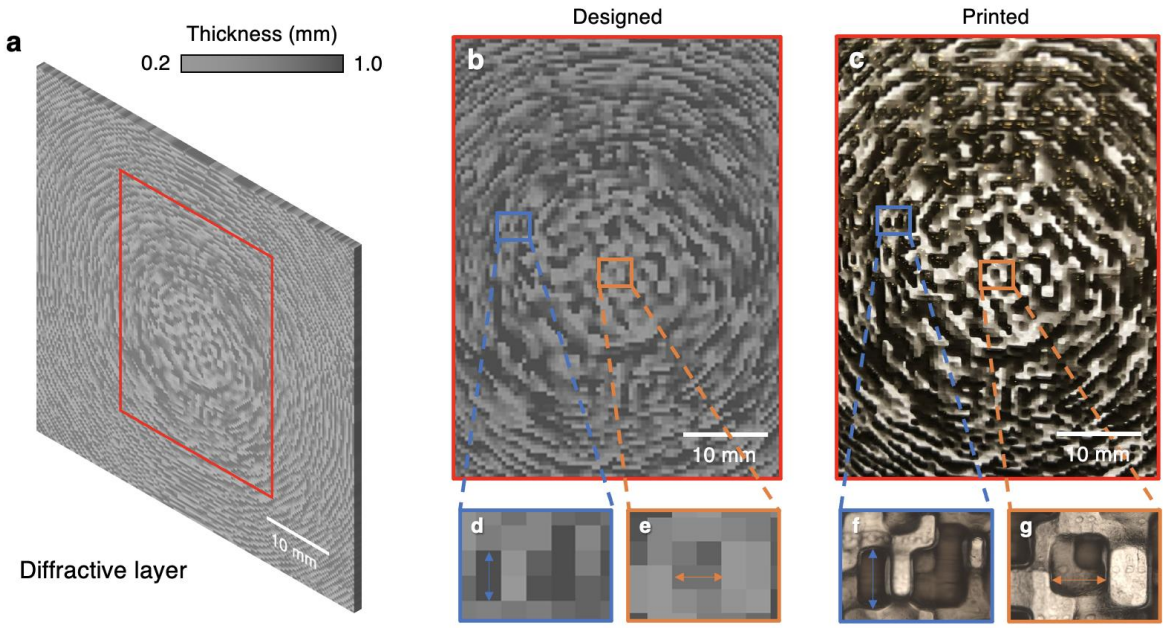

Fig. S14. Comparison between a 3D-printed diffractive layer and its numerical design. a, The grayscale coded thickness profile of the 1st diffractive layer used in the diffractive network model that attains a 95.05% testing accuracy (same as the one shown in Fig. 2b, but coded with a different color map). **b**, The zoomed-in image of the red-boxed area in (a). **c**, The photograph of the 3D-printed diffractive layer zoomed-in to the same region as (b). **d-g**, The designed thickness profile (d, e) and the microscopic details (f, g) of the example local regions (blue and orange boxes) that share the same locations on the diffractive layer across (b) and (c).

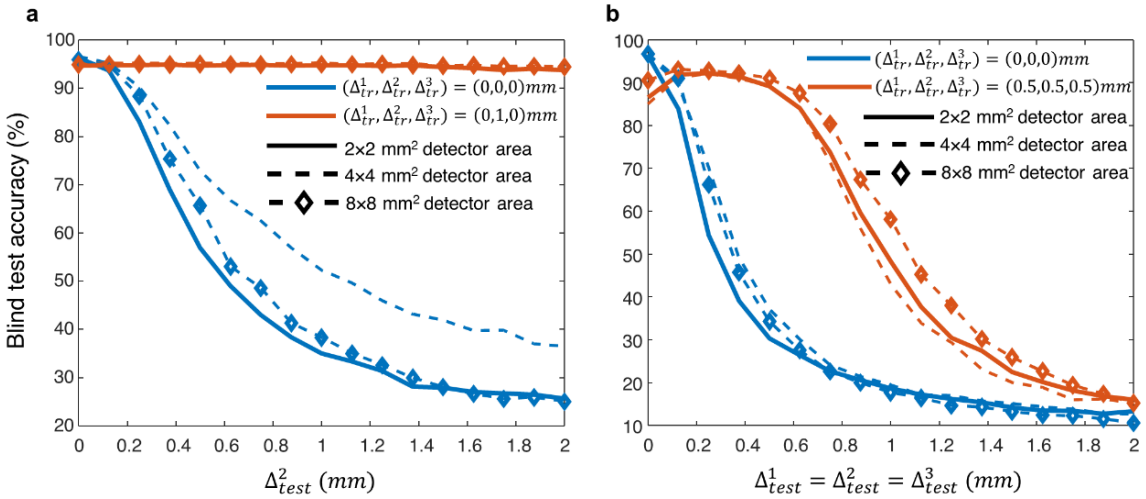

Fig. S15. Misalignment resilient, jointly-trained single-pixel hybrid machine vision systems classifying handwritten digits (MNIST) using spectral encoding of data classes through 3 diffractive layers. a, The blind image classification accuracies provided by the misalignment-free designs, i.e., $(\Delta^1_x, \Delta^2_x, \Delta^3_x) = (\Delta^1_y, \Delta^2_y, \Delta^3_y) = (\Delta^1_{tr}, \Delta^2_{tr}, \Delta^3_{tr}) = (0, 0, 0)$, and the vaccinated designs, i.e., $(\Delta^1_x, \Delta^2_x, \Delta^3_x) = (\Delta^1_y, \Delta^2_y, \Delta^3_y) = (\Delta^1_{tr}, \Delta^2_{tr}, \Delta^3_{tr}) = (0, 1\ mm, 0)$, are shown. These designs are tested under random lateral misalignments of *only the middle (2nd) diffractive layer*. Each data point reports the average blind classification accuracy computed over 1000 different misalignment configurations that were created by randomly moving *the middle diffractive layer* of each hybrid network model to a new location within the range $(-\Delta^2_{test}, \Delta^2_{test})$ on both *x* and *y* axes. **b,** Same as (a), except the vaccination strategy is extended to include the lateral misalignments of all 3 diffractive layers during the training, i.e., $(\Delta^1_x, \Delta^2_x, \Delta^3_x) = (\Delta^1_y, \Delta^2_y, \Delta^3_y) = (\Delta^1_{tr}, \Delta^2_{tr}, \Delta^3_{tr}) = (0.5\ mm, 0.5\ mm, 0.5\ mm)$. During blind testing, random lateral misalignment errors were introduced in the positions of *all three diffractive layers*. Single-pixel detectors with larger active areas provide more resilience to misalignments, as expected. Each data point reports the average blind classification accuracy computed over 1000 different misalignment configurations that were created by randomly moving *all three diffractive layers* of each hybrid network model to three separate locations within the range $(-\Delta_{test}, \Delta_{test})$ on both *x* and *y* axes, at their corresponding diffractive planes. Therefore, in each implementation of misalignment, a given diffractive layer is randomly moved a new lateral position, separate from the other diffractive layers that are also misaligned randomly. $\Delta^1_{test} = \Delta^2_{test} = \Delta^3_{test} = \Delta_{test}$.

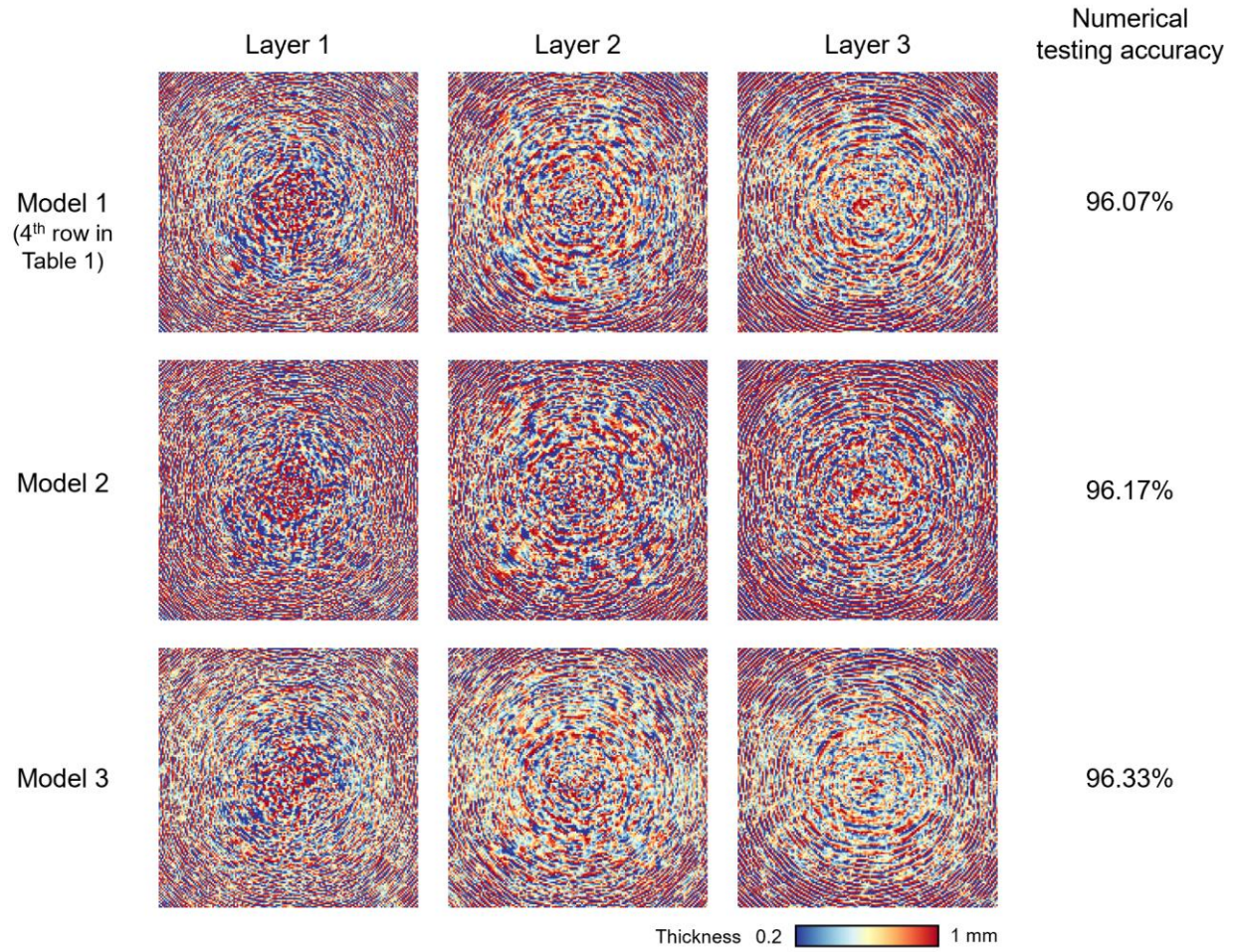

Fig. S16. Using the same network architecture and hyperparameters but different data feeding sequence (randomized), three different diffractive network models are trained, and the learned thickness profiles of the resulting diffractive layers along with their numerical testing (MNIST classification) accuracies are shown. The first row (Model 1) represents the diffractive model that was presented in the main text (4^{th} row in Table 1) and achieved 96.07% blind testing accuracy, while the other two rows represent the diffractive models that achieve classification accuracies of 96.17% (Model 2) and 96.33% (Model 3), respectively. Mean classification accuracy: 96.19%, standard deviation: 0.13%.

Supplementary Tables

Diffractive network	Testing accuracy $\max(s)$ or $\max(s')$ (%)
26 wavelengths $\mathbf{s} = [s_0, s_1, \ldots, s_{25}]$ (1st row in Fig. S13)	84.05
52 wavelengths (differential) $\mathbf{s_D} = [s_{0+}, s_{0-}, s_{1+}, s_{1-}, \ldots, s_{25+}, s_{25-}]$ $\mathbf{s} = \mathbf{\Delta s} = [\Delta s_0, \Delta s_1, \ldots, \Delta s_{25}]$ (2nd row in Fig. S13)	85.60
26 wavelengths (jointly-trained with ANN) $\mathbf{s} = [s_0, s_1, \ldots, s_{25}]$ (3rd row in Fig. S13)	86.78
52 wavelengths (differential, jointly-trained with ANN) $\mathbf{s_D} = [s_{0+}, s_{0-}, s_{1+}, s_{1-}, \ldots, s_{25+}, s_{25-}]$ $\mathbf{s} = \mathbf{\Delta s} = [\Delta s_0, \Delta s_1, \ldots, \Delta s_{25}]$ (4th row in Fig. S13)	87.68

Table S1. Blind testing accuracies for EMNIST handwritten capital letter classification. Also see Fig. S13.

Diffractive network	ξ (Eq. S22)	Blind testing accuracy $\max(\boldsymbol{s})$ (%)
10 wavelengths $\mathbf{s} = [s_0, s_1, \ldots, s_9]$ ((a) in Fig. S11)	0.0	10.72
10 wavelengths $\mathbf{s} = [s_0, s_1, \ldots, s_9]$ ((b) in Fig. S11)	0.25	94.94
10 wavelengths $\mathbf{s} = [s_0, s_1, \ldots, s_9]$ ((c) in Fig. S11)	0.5	95.66
10 wavelengths $\mathbf{s} = [s_0, s_1, \ldots, s_9]$ ((d) in Fig. S11)	1.0	96.01
20 wavelengths (differential) $\mathbf{s_D} = [s_{0+}, s_{0-}, s_{1+}, s_{1-}, \ldots, s_{9+}, s_{9-}]$ $\mathbf{s} = \boldsymbol{\Delta s} = [\Delta s_0, \Delta s_1, \ldots, \Delta s_9]$ ((a) in Fig. S12)	0.0	8.88
20 wavelengths (differential) $\mathbf{s_D} = [s_{0+}, s_{0-}, s_{1+}, s_{1-}, \ldots, s_{9+}, s_{9-}]$ $\mathbf{s} = \boldsymbol{\Delta s} = [\Delta s_0, \Delta s_1, \ldots, \Delta s_9]$ ((b) in Fig. S12)	0.25	95.17
20 wavelengths (differential) $\mathbf{s_D} = [s_{0+}, s_{0-}, s_{1+}, s_{1-}, \ldots, s_{9+}, s_{9-}]$ $\mathbf{s} = \boldsymbol{\Delta s} = [\Delta s_0, \Delta s_1, \ldots, \Delta s_9]$ ((c) in Fig. S12)	0.5	95.83
20 wavelengths (differential) $\mathbf{s_D} = [s_{0+}, s_{0-}, s_{1+}, s_{1-}, \ldots, s_{9+}, s_{9-}]$ $\mathbf{s} = \boldsymbol{\Delta s} = [\Delta s_0, \Delta s_1, \ldots, \Delta s_9]$ ((d) in Fig. S12)	1.0	96.04
50 wavelengths (averaging) $\mathbf{s_D} = [s_0^1, s_0^2, s_0^3, s_0^4, s_0^5, \ldots, s_9^1, s_9^2, s_9^3, s_9^4, s_9^5]$ $\mathbf{s} = [s_0, s_1, \ldots, s_9]$	0.5	95.86
50 wavelengths (learnable weighted averaging) $\mathbf{s_D} = [s_0^1, s_0^2, s_0^3, s_0^4, s_0^5, \ldots, s_9^1, s_9^2, s_9^3, s_9^4, s_9^5]$ $\mathbf{s} = [s_0, s_1, \ldots, s_9]$	0.5	95.22

Table S2. Blind testing accuracies of jointly-trained hybrid machine vision systems for MNIST image dataset. Image classification is performed by the corresponding diffractive network's output, $\max(\boldsymbol{s})$, and a decoder ANN is jointly-trained for image reconstruction using the spectral encoding of data classes through a single-pixel detector. Also see Figs. S11 and S12.

REFERENCE AND NOTES

1. J. B. Pendry, Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).

2. E. Cubukcu, K. Aydin, E. Ozbay, S. Foteinopoulou, C. M. Soukoulis, Negative refraction by photonic crystals. *Nature* **423**, 604–605 (2003).

3. N. Fang, H. Lee, C. Sun, X. Zhang, Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308**, 534–537 (2005).

4. Z. Jacob, L. V. Alekseyev, E. Narimanov, Optical hyperlens: Far-field imaging beyond the diffraction limit. *Opt. Express* **14**, 8247–8256 (2006).

5. N. Engheta, Circuits with light at nanoscales: Optical nanocircuits inspired by metamaterials. *Science* **317**, 1698–1702 (2007).

6. Z. Liu, H. Lee, Y. Xiong, C. Sun, X. Zhang, Far-field optical hyperlens magnifying sub-diffraction-limited objects. *Science* **315**, 1686–1686 (2007).

7. K. F. MacDonald, Z. L. Sámson, M. I. Stockman, N. I. Zheludev, Ultrafast active plasmonics. *Nat. Photonics* **3**, 55–58 (2009).

8. D. Lin, P. Fan, E. Hasman, M. L. Brongersma, Dielectric gradient metasurface optical elements. *Science* **345**, 298–302 (2014).

9. N. Yu, F. Capasso, Flat optics with designer metasurfaces. *Nat. Mater.* **13**, 139–150 (2014).

10. A. I. Kuznetsov, A. E. Miroshnichenko, M. L. Brongersma, Y. S. Kivshar, B. Luk'yanchuk, Optically resonant dielectric nanostructures. *Science* **354**, aag2472 (2016).

11. S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, A. A. G. Requicha, Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides. *Nat. Mater.* **2**, 229–232 (2003).

12. A. Alù, N. Engheta, Achieving transparency with plasmonic and metamaterial coatings. *Phys. Rev. E* **72**, 016623 (2005).

13. D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, D. R. Smith, Metamaterial electromagnetic cloak at microwave frequencies. *Science* **314**, 977–980 (2006).

14. J. B. Pendry, D. Schurig, D. R. Smith, Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).

15. W. Cai, U. K. Chettiar, A. V. Kildishev, V. M. Shalaev, Optical cloaking with metamaterials. *Nat. Photonics* **1**, 224–227 (2007).

16. J. Valentine, J. Li, T. Zentgraf, G. Bartal, X. Zhang, An optical cloak made of dielectrics. *Nat. Mater.* **8**, 568–571 (2009).

17. E. E. Narimanov, A. V. Kildishev, Optical black hole: Broadband omnidirectional light absorber. *Appl. Phys. Lett.* **95**, 041106 (2009).

18. R. F. Oulton, V. J. Sorger, T. Zentgraf, R.-M. Ma, C. Gladden, L. Dai, G. Bartal, X. Zhang, Plasmon lasers at deep subwavelength scale. *Nature* **461**, 629–632 (2009).

19. Y. Zhao, M. A. Belkin, A. Alù, Twisted optical metamaterials for planarized ultrathin broadband circular polarizers. *Nat. Commun.* **3**, 870 (2012).

20. C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, W. J. Padilla, Terahertz compressive imaging with metamaterial spatial light modulators. *Nat. Photonics* **8**, 605–609 (2014).

21. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, A. Ozcan, All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).

22. N. Mohammadi Estakhri, B. Edwards, N. Engheta, Inverse-designed metastructures that solve equations. *Science* **363**, 1333–1338 (2019).

23. T. W. Hughes, I. A. D. Williamson, M. Minkov, S. Fan, Wave physics as an analog recurrent neural network. *Sci. Adv.* **5**, eaay6946 (2019).

24. D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).

25. K. H. Wagner, in *Frontiers in Optics 2017* (Optical Society of America, 2017), p. FW2C.1; www.osapublishing.org/abstract.cfm?uri=FiO-2017-FW2C.1.

26. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, M. Soljačić, Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).

27. T. F. de Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, P. R. Prucnal, Progress in neuromorphic photonics. *Nanophotonics* **6**, 577–599 (2017).

28. B. J. Shastri, A. N. Tait, T. F. de Lima, M. A. Nahmias, H.-T. Peng, P. R. Prucnal, Principles of neuromorphic photonics. arXiv:1801.00016 [cs.ET] (2018).

29. J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, D. Brunner, Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* **5**, 756–760 (2018).

30. E. Khoram, A. Chen, D. Liu, L. Ying, Q. Wang, M. Yuan, Z. Yu, Nanophotonic media for artificial neural inference. *Photonics Res.* **7**, 823–827 (2019).

31. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, W. H. P. Pernice, All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).

32. L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, T. Mueller, Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).

33. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554–2558 (1982).

34. N. H. Farhat, D. Psaltis, A. Prata, E. Paek, Optical implementation of the Hopfield model. *Appl. Optics* **24**, 1469–1475 (1985).

35. K. Wagner, D. Psaltis, Multilayer optical learning networks. *Appl. Optics* **26**, 5061–5076 (1987).

36. D. Psaltis, A. Sideris, A. A. Yamamura, A multilayered neural network controller. *IEEE Control Syst. Mag.* **8**, 17–21 (1988).

37. D. Psaltis, D. Brady, X.-G. Gu, S. Lin, Holography in artificial neural networks. *Nature* **343**, 325–330 (1990).

38. A. V. Krishnamoorthy, G. Yayla, S. C. Esener, in *Proceedings of the IJCNN-91-Seattle International Joint Conference on Neural Networks* (IEEE, 1991), vol. 1, pp. 527–534.

39. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, G. Wetzstein, Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* **8**, 12324 (2018).

40. Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, S. Du, All-optical neural network with nonlinear activation functions. *Optica* **6**, 1132–1137 (2019).

41. L. Lu, L. Zhu, Q. Zhang, B. Zhu, Q. Yao, M. Yu, H. Niu, M. Dong, G. Zhong, Z. Zeng, Miniaturized diffraction grating design and processing for deep neural network. *IEEE Photonics Technol. Lett.* **31**, 1952–1955 (2019).

42. P. del Hougne, M. F. Imani, A. V. Diebold, R. Horstmeyer, D. R. Smith, Learned integrated sensing pipeline: Reconfigurable metasurface transceivers as trainable physical layer in an artificial neural network. *Adv. Sci.* **7**, 1901913 (2020).

43. Y. Luo, D. Mengu, N. T. Yardimci, Y. Rivenson, M. Veli, M. Jarrahi, A. Ozcan, Design of task-specific optical systems using broadband diffractive neural networks. *Light Sci. Appl.* **8**, 112 (2019).

44. D. Mengu, Y. Luo, Y. Rivenson, A. Ozcan, Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 1–14 (2020).

45. J. Li, D. Mengu, Y. Luo, Y. Rivenson, A. Ozcan, Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *Adv. Photonics* **1**, 046001 (2019).

46. C. Qian, X. Lin, X. Lin, J. Xu, Y. Sun, E. Li, B. Zhang, H. Chen, Performing optical logic operations by a diffractive neural network. *Light Sci. Appl.* **9**, 59 (2020).

47. N. T. Yardimci, M. Jarrahi, High sensitivity terahertz detection through large-area plasmonic nano-antenna arrays. *Sci. Rep.* **7**, 42667 (2017).

48. N. T. Yardimci, S.-H. Yang, C. W. Berry, M. Jarrahi, High-power terahertz generation using large-area plasmonic photoconductive emitters. *IEEE Trans. Terahertz Sci. Technol.* **5**, 223–229 (2015).

49. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).

50. A. B. Owen, in *Contemporary Mathematics*, J. S. Verducci, X. Shen, J. Lafferty, Eds. (American Mathematical Society, 2007), vol. 443, pp. 59–71; www.ams.org/conm/443/.

51. I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, N. Navab, in *2016 Fourth International Conference on 3D Vision (3DV)* (IEEE, 2016), pp. 239–248.

52. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi, Eds. (Lecture Notes in Computer Science, Springer International Publishing, 2015), pp. 234–241.

53. D. Mengu, Y. Zhao, N. T. Yardimci, Y. Rivenson, M. Jarrahi, A. Ozcan, Misalignment resilient diffractive optical networks. *Nanophotonics* **9**, 4207–4219 (2020).

54. G. Cohen, S. Afshar, J. Tapson, A. van Schaik, EMNIST: An extension of MNIST to handwritten letters. arXiv:1702.05373 [cs.CV] (2017); http://arxiv.org/abs/1702.05373.

55. D. Mengu, Y. Rivenson, A. Ozcan, Scale-, shift-, and rotation-invariant diffractive optical networks. *ACS Photonics* **8**, 324–334 (2021).

56. O. Kulce, D. Mengu, Y. Rivenson, A. Ozcan, All-optical information processing capacity of diffractive surfaces. *Light Sci. Appl.* **10**, 25 (2020).

57. M. Veli, D. Mengu, N. T. Yardimci, Y. Luo, J. Li, Y. Rivenson, M. Jarrahi, A. Ozcan, Terahertz pulse shaping using diffractive surfaces. *Nat. Commun.* **12**, 37 (2020).

58. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG] (2014).