Scale-, Shift-, and Rotation-Invariant Diffractive Optical Networks

Deniz Mengu, Yair Rivenson, and Aydogan Ozcan*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Recent research efforts in optical computing have gravitated toward developing optical neural networks that aim to benefit from the processing speed and parallelism of optics/photonics in machine learning applications. Among these endeavors, Diffractive Deep Neural Networks (D²NNs) harness light-matter interaction over a series of trainable surfaces, designed using deep learning, to compute a desired statistical inference task as the light waves propagate from the input plane to the output field-of-view. Although earlier studies have demonstrated the generalization capability of diffractive optical networks to unseen data, achieving, e.g., >98% image classification accuracy for handwritten digits, these previous designs are in general sensitive to the spatial scaling, translation, and rotation of the input objects. Here, we demonstrate a new training strategy for diffractive networks that introduces input object translation, rotation, and/or scaling during the training phase as uniformly distributed random variables to build resilience in their blind inference performance against such object transformations. This training strategy successfully guides the evolution of the diffractive optical network design toward a solution that is scale-, shift-, and rotation-invariant, which is especially important and useful for dynamic machine vision applications in, e.g., autonomous cars, in vivo imaging of biomedical specimen, among others.

KEYWORDS: optical computing, diffractive networks, deep learning, optical neural networks
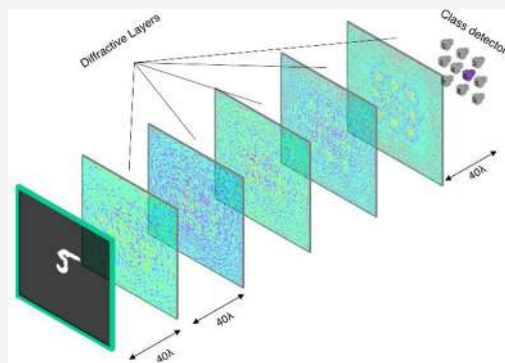
Motivated by the success of deep learning[1,2] in various applications,[3−16] optical neural networks have gained an important momentum in recent years. Although optical neural networks and related optical computing hardware are relatively at an earlier stage in terms of their inference and generalization capabilities, when compared to the state-of-the-art electronic deep neural networks and the underlying digital processors, optics/photonics technologies might potentially bring significant advantages for machine learning systems in terms of their power efficiency, parallelism, and computational speed.[17−30] Among different physical architectures used for the design of optical neural networks,[17,22−25,27,31,32] Diffractive Deep Neural Networks (D²NNs)[31,33−39] utilize the diffraction of light through engineered surfaces/layers to form an optical network that is based on light-matter interaction and free-space propagation of light. D²NNs offer a unique optical machine learning framework that formulates a given learning task as a black-box function approximation problem, parametrized through the trainable physical features of matter that control the phase and/or amplitude of light. One of the most convenient methods to devise a D²NN is to employ multiple transmissive and/or reflective diffractive surfaces/layers that collectively form an optical network between an input and output field-of-view. During the training stage, the complex-valued transmission/reflection coefficients of the layers of a D²NN are designed for a given statistical (or deterministic) task/goal, where each diffractive feature (i.e., neuron) of a

given layer is iteratively adjusted during the training phase using, e.g., the error back-propagation method.[30,40,41] After this training and design phase, the resulting diffractive layers/ surfaces are physically fabricated using, e.g., 3D printing or lithography, to form a passive optical network that performs inference as the input light diffracts from the input plane to the output. Alternatively, the final diffractive layer models can also be implemented by using various types of spatial light modulators (SLMs) to bring reconfigurability and data adaptability to the diffractive network, at the expense of, e.g., increased power consumption of the system.

Since the initial experimental demonstration of image classification using D²NNs that are composed of 3D-printed diffractive layers,[31,36] the optical inference capacity of diffractive optical networks has been significantly improved based on, e.g., differential detection scheme, class-specific designs, and ensemble-learning techniques.[33,34] Owing to these systematic advances in diffractive optical networks and training methods, recent studies have reported classification accuracies of >98%, >90%, and >62% for the data sets of handwritten
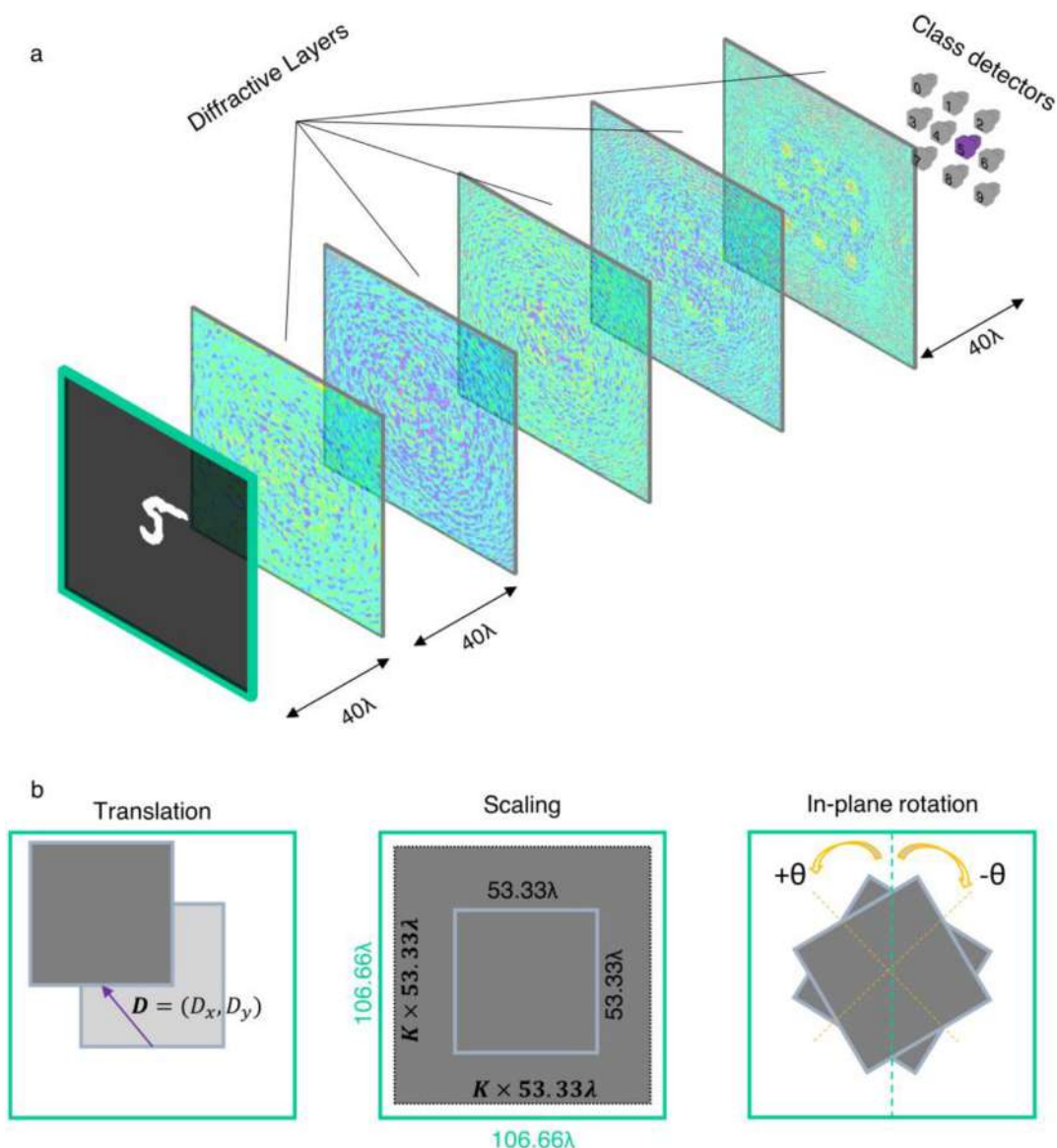
Figure 1. (a) The layout of the diffractive optical networks trained and tested in this study. (b) The object transformations modeled during the training and testing of the diffractive optical networks presented in this work.

digits (MNIST), fashion products (Fashion-MNIST), and CIFAR-10 images, respectively.[33,34] Beyond classification tasks, diffractive networks were also shown to serve as trainable optical front-ends, forming hybrid (optical-electronic) machine learning systems.[35] Replacing the conventional imaging-optics in machine vision systems with diffractive optical networks has been shown to offer unique opportunities to lower the computational complexity and burden on back-end electronic neural networks as well as to mitigate the inference accuracy loss due to pixel-pitch limited, low-resolution imaging systems.[35] Furthermore, in a recent study, diffractive optical networks have been trained to encode the spatial information on input objects into the power spectrum of the diffracted broadband light, enabling object classification and image reconstruction using only a single-pixel spectroscopic detector at the output plane, demonstrating an unconventional, task-specific, and resource-efficient machine vision platform.[36] This extension of diffractive optical networks and the related training models to conduct inference based on broadband light sources could also be important for processing colorful objects

or images at multiple spectral bands, covering, e.g., red, green, and blue channels in the visible part of the spectrum.

In all of these existing diffractive optical network designs, the inference accuracies are in general sensitive to object transformations such as, e.g., lateral translation, rotation, and/or scaling of the input objects that are frequently encountered in various machine vision applications. In this work, we quantify the sensitivity of diffractive optical networks to these uncertainties associated with the lateral position, scale and in-plane orientation/rotation angle of the input objects (see Figure 1). Furthermore, we demonstrate a D^2NN design scheme that formulates these object transformations through random variables used during the deep learning-based training phase of the diffractive layers. In this manner, the evolution of the layers of a diffractive optical network can adapt to random translation, scaling and rotation of the input objects and, hence, the blind inference capacity of the optical network can be maintained despite these input object uncertainties. The presented training strategy will enable diffractive optical networks to find applications in machine vision systems that
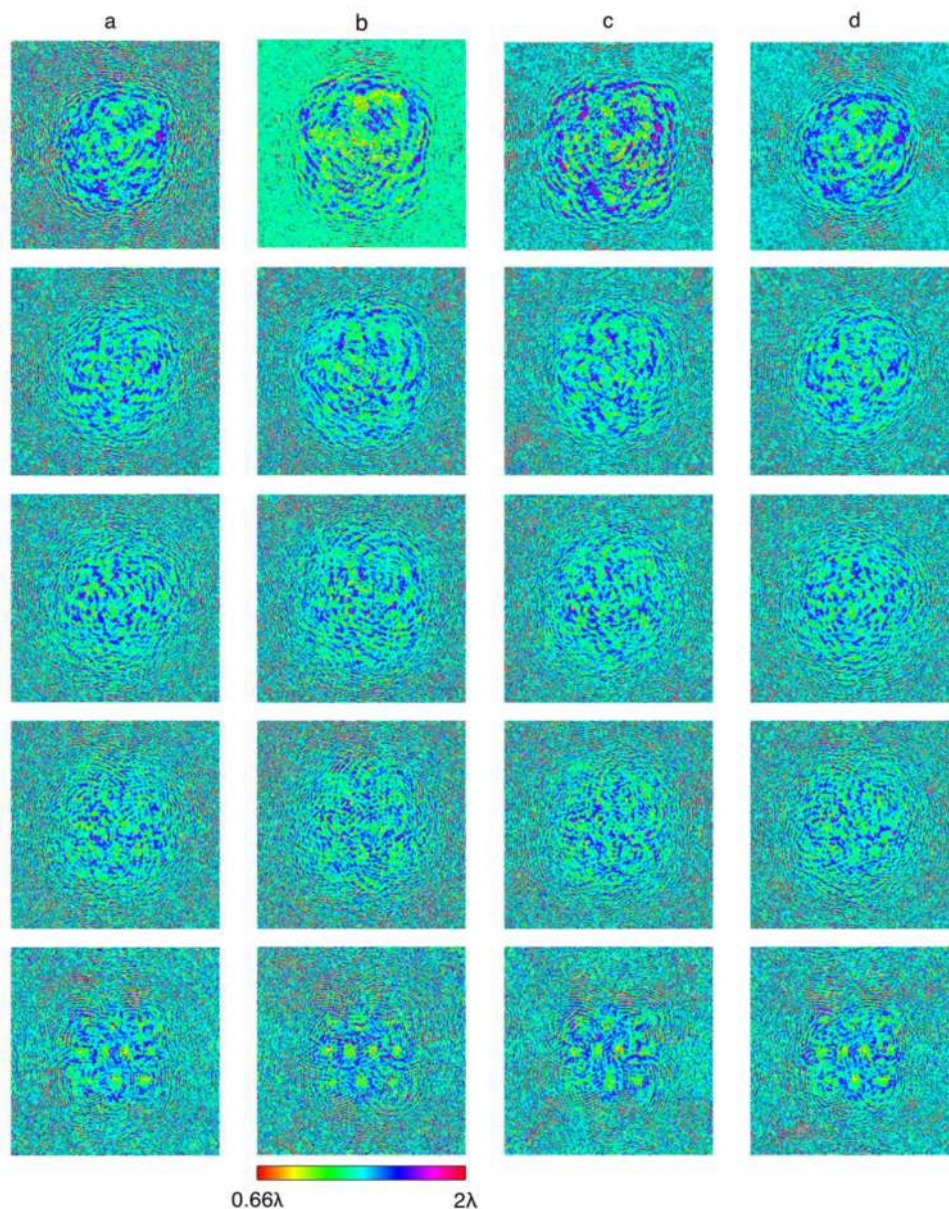
Figure 2. Thickness profiles of the designed diffractive layers constituting (a) the standard design ($\Delta_{tr} = \zeta_{tr} = \theta_{tr} = 0$); (b) the shift-invariant design trained with $\Delta_{tr} = 8.48\lambda$ (purple curve shown in Figure 3); (c) the scale-invariant design trained with $\zeta_{tr} = 0.4$ (purple curve shown in Figure 5); (d) the rotation-invariant design trained with $\theta_{tr} = 20°$ (purple curve shown in Figure 6).

require low-latency as well as memory- and power-efficient inference engines for monitoring dynamic events. Beyond diffractive networks, the outlined training scheme can be utilized in other optical machine learning platforms as well as in deep learning-based inverse design problems to create robust solutions that can sustain their target performance against undesired/uncontrolled input field transformations.

RESULTS AND DISCUSSION

In a standard D²NN-based optical image classifier,[31,33−35,42] the number of optoelectronic detectors positioned at the output plane is equal to the number of classes in the target data set, and each detector uniquely represents one data class (see Figure 1a). The final class decision is based on the *max* operation over the collected optical signals by these class detectors. According to the diffractive network layout illustrated in Figure 1a, the input objects (e.g., handwritten

MNIST digits) lie within a predefined field-of-view (FOV) of $53.33\lambda \times 53.33\lambda$, where λ denotes the wavelength of the illumination light. The center of the FOV coincides with the optical axis passing through the center of the diffractive layers. The size of each diffractive layer is chosen to be $106.66\lambda \times 106.66\lambda$, i.e., exactly 2× the size of the input FOV on each lateral axis. The smallest diffractive feature size on each D²NN layer is set to be $\sim 0.53\lambda$; i.e., there are 200×200 trainable features on each diffractive layer of a given D²NN design. At the output plane, each detector is assumed to cover an area of $6.36\lambda \times 6.36\lambda$ and they are located within an output FOV of $53.33\lambda \times 53.33\lambda$—matching the input FOV size.

On the basis of these design parameters, a 5-layer diffractive optical network with phase-only modulation at each neuron achieves a blind testing accuracy of 97.64% for the classification of amplitude-encoded MNIST images illuminated with a uniform plane wave. Figure 2a illustrates the thickness
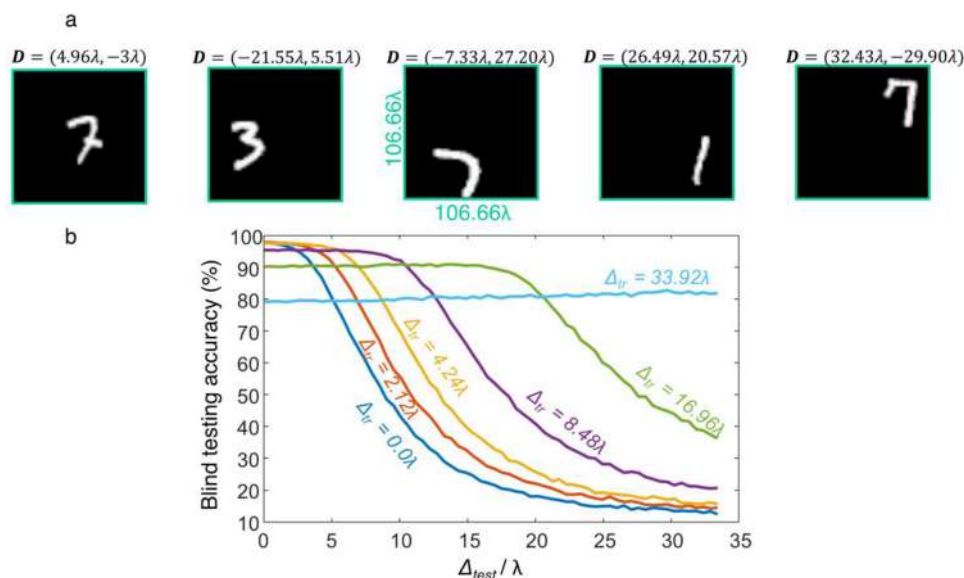
Figure 3. Shift-invariant diffractive optical networks. (a) Randomly shifted object samples from the MNIST test data set. Green frame around each object demonstrates the size of the diffractive layers ($106.66\lambda \times 106.66\lambda$). (b) The blind inference accuracies provided by six different diffractive network models trained with $\Delta_x = \Delta_y = \Delta_{tr}$, taken as 0.0λ (blue), 2.12λ (red), 4.24λ (yellow), 8.48λ (purple), 16.96λ (green), 33.92λ (light-blue) when they were tested under different levels random object shifts with the control parameter, $\Delta_x = \Delta_y = \Delta_{test}$, swept from 0.0λ to 33.92λ.

profiles of the resulting 5 diffractive layers, constituting this standard D^{2}NN design. To quantify the sensitivity of the blind inference accuracy of this D^{2}NN design against uncontrolled lateral object translations, we introduced an object displacement vector (Figure 1b), $\mathbf{D} = (D_x, D_y)$, that has two components, defined as independent, uniformly distributed random variables:

$$D_x \sim U(-\Delta_x, \Delta_x) \tag{1a}$$

$$D_y \sim U(-\Delta_y, \Delta_y) \tag{1b}$$

The standard diffractive network model (shown in Figure 2a) was trained (tr) with $\Delta_x = \Delta_y = \Delta_{tr} = 0$, and was then tested under different levels of input object position shifts by sweeping the values of $\Delta_x = \Delta_y = \Delta_{test}$ from 0 to 33.92λ with steps of 0.53λ. Stated differently, the final test accuracy corresponding to each Δ_{test} value reflects the image classification performance of the same diffractive network model that was tested with 10 000 different object positions randomly chosen within the range set by Δ_{test} (see Figure 3a for exemplary test objects). This analysis revealed that the blind inference accuracy of the standard D^{2}NN design ($\Delta_{tr} = 0$) which achieves 97.64% under $\Delta_{test} = 0$ quickly falls below 90% as the input objects starts to move within the range $\pm 3.5\lambda$ (blue curve in Figure 3, defined with $\Delta_{tr} = 0$). As the area covered by the possible object shifts is increased further, the inference accuracy of this native network model decreases rapidly (see Figure 3).

In this conventional design approach, the optical forward model of the diffractive network training assumes that the input objects inside the sample FOV are free-from any type of undesired geometrical variations, i.e., $\Delta_{tr} = 0$. Hence, the diffractive layers are *not* challenged to process optical waves coming from input objects at different spatial locations, possibly overfitting to the assumed FOV location. As a result, the inference performance of the resulting diffractive network model becomes dependent on the relative lateral location of

the input object with respect to the plane of the diffractive layers and the output detectors.

To mitigate this problem, we adopted a training strategy inspired by data augmentation techniques used in deep learning. According to this scheme, each training image sample in a batch is randomly shifted, based on a realization of the displacement vector ($\mathbf{D}$), and subsequently, the loss function is computed by propagating these randomly shifted object fields through the diffractive network (see the Methods for details). Using this training scheme, we designed 5 different diffractive network models based on different ranges of object displacement, i.e., $\Delta_x = \Delta_y = \Delta_{tr} = 2.12\lambda$, 4.24λ, 8.48λ, 16.96λ, and 33.92λ (see eq 1). Figure 3 illustrates the MNIST image classification accuracies provided by these 5 new diffractive network models as a function of Δ_{test}. Comparison between the diffractive network models trained with $\Delta_{tr} = 0$ (blue) and $\Delta_{tr} = 2.12\lambda$ (red) reveals that due to the data augmentation introduced by the small object shifts during the training, the latter can achieve an improved inference accuracy of 98.00% for MNIST digits under $\Delta_{test} = 0$. Furthermore, the diffractive network trained with $\Delta_{tr} = 2.12\lambda$ can maintain its classification performance when the input objects are randomly shifted within a certain lateral range (see the right shift of the red curve in Figure 3). Similarly, training a diffractive network model with $\Delta_{tr} = 4.24\lambda$ (yellow curve in Figure 3) also results in a better classification accuracy of 97.75% when compared to the 97.64% achieved by the standard model ($\Delta_{tr} = 0$) under $\Delta_{test} = 0$. In addition, this new diffractive model exhibits further resilience to random shifts of the objects within the input FOV, which is indicated by the stronger right shift of the yellow curve in Figure 3. For example, for $\Delta_{test} = 3.71\lambda$ in Figure 3, the input test objects are randomly shifted in x and y by an amount determined by $D_x \sim U(-3.71\lambda, 3.71\lambda)$ and $D_y \sim U(-3.71\lambda, 3.71\lambda)$, respectively, and this results in a classification accuracy of 97.07% for the new diffractive model ($\Delta_{tr} = 4.24\lambda$), whereas the inference accuracy of the standard model ($\Delta_{tr} = 0$) decreases to 89.88% under the same random lateral shifts of the input test objects.
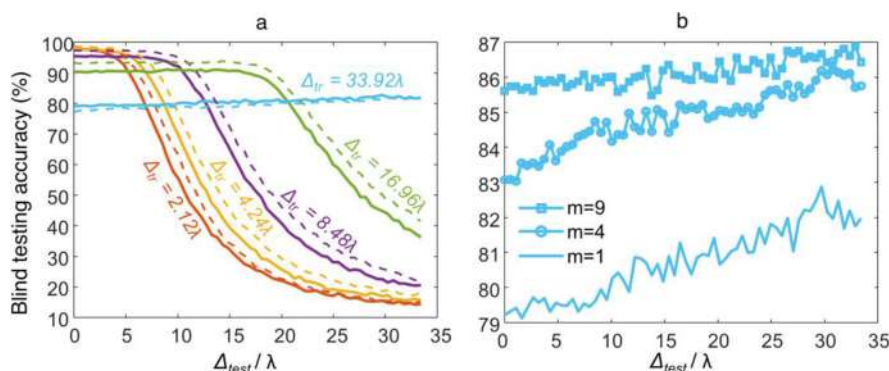
Figure 4. Different design strategies that can improve the performance of shift-invariant diffractive optical networks. (a) The comparison between the inference accuracies of standard (solid curves) and differential (dashed curves) diffractive optical networks trained using various Δ_{tr} values. (b) Blind testing classification accuracies of three nondifferential, 5-layer D^2NN designs that have $m \times 40K$ optical neurons per layer, with $m = 1$, 4, and 9. All these diffractive optical networks were trained using $\Delta_{tr} = 33.92\lambda$. The diffractive network designs with wider diffractive layers and more neurons per layer can generalize more effectively to random object translations.

Further increasing the range of the object location uncertainty, e.g., to $\Delta_{tr} = 8.48\lambda$ (purple curve in Figure 3), we start to observe a trade-off between the peak inference accuracy and the resilience of the diffractive network to random object shifts. For instance, the diffractive optical network trained with $\Delta_{tr} = 8.48\lambda$ can achieve a peak classification accuracy of 95.55%, which represents a ~2% accuracy compromise with respect to the native diffractive network model ($\Delta_{tr} = 0$) tested under $\Delta_{test} = 0$. However, using such a large object location uncertainty in the training phase also results in a rather flat accuracy curve over a much larger Δ_{test} range as shown in Figure 3; in other words, this design strategy expands the effective input object FOV that can be utilized for the desired machine learning task. For example, if the test objects were to freely move within the area defined by $\Delta_x = \Delta_y = \Delta_{test} = 6.89\lambda$, the diffractive network model trained with $\Delta_{tr} = 8.48\lambda$ (purple curve in Figure 3) brings a >30% inference accuracy advantage compared to the standard model (blue curve in Figure 3). The resulting layer thickness profiles for this diffractive optical network design trained with $\Delta_{tr} = 8.48\lambda$ are also shown in Figure 2b.

For the case where Δ_{tr} was set to be 16.96λ, the mean test classification accuracy over the range $0 < \Delta_{test} < \Delta_{tr}$ is observed to be 90.46% (see the green curve in Figure 3b). The relatively more pronounced performance trade-off in this case can be explained based on the increased input FOV. Stated differently, with larger Δ_{tr} values, the effective input FOV of the diffractive network is increased, and the dimensionality of the solution space[37] provided by a diffractive network design with a limited number of layers (and neurons) might not be sufficient to provide the desired solution when compared to a smaller input FOV diffractive network design. The use of wider diffractive layers (i.e., larger number of neurons per layer) can be a strategy to further boost the inference accuracy over larger Δ_{tr} values (or larger effective input FOVs), which will be further discussed and demonstrated later in our analysis below (see Figure 4b).

As an alternative design strategy, the detector plane configuration shown in Figure 1a can also be replaced with a differential detection scheme[33] to mitigate this relative drop in blind inference accuracy for designs with large Δ_{tr}. In this scheme, instead of assigning a single optoelectronic detector per class, we designate two detectors to each data class and represent the corresponding class scores based on the

normalized difference between the optical signals collected by each detector pair. Figure 4a illustrates a comparison between the blind classification accuracies of standard (solid curves) and differential (dashed curves) diffractive network designs, when they were trained with random lateral shifts of the input objects. For all of these designs, except the $\Delta_{tr} = 33.92\lambda$ case, the differential diffractive networks achieve higher classification accuracies throughout the entire testing range, showing their superior robustness and adaptability to input field variations compared to their nondifferential counterparts. For example, the peak inference accuracy (95.55%) achieved by the diffractive optical network trained with $\Delta_{tr} = 8.48\lambda$ (solid purple curve in Figure 4a) increases to 97.33% using the differential detection scheme (dashed purple curve in Figure 4a). As another example, for $\Delta_{tr} = 16.96\lambda$, the mean classification accuracy of the differential diffractive network over $0 < \Delta_{test} < \Delta_{tr}$ yields 93.38%, which is ~3% higher compared to the performance of its nondifferential counterpart for the same test range.

On the other hand, enlarging the uncertainty in the input object translation further, e.g., $\Delta_{tr} = 33.92\lambda$, starts to balance out the benefits of using differential detection at the output plane (see the solid and dashed blue curves in Figure 4, which closely follow each other). In fact, when Δ_x and Δ_y in eq 1 are large enough, such as $\Delta_{tr} = 33.92\lambda$, the effective input FOV increases considerably with respect to the size of the diffractive layers; as we discussed earlier, the use of wider diffractive layers with larger numbers of neurons per layer could be used to mitigate this and improve inference performance of D^2NN designs that are trained with relatively large Δ_{tr} values. To shed more light on this, using $\Delta_{tr} = 33.92\lambda$ we trained two additional diffractive optical network models with wider diffractive layers that cover $m = 4$ and $m = 9$ fold larger number of neurons per layer compared to the standard design ($m = 1$) that has 40K neurons per diffractive layer; stated differently, each diffractive layer of these two new designs contain $(2 \times 200) \times (2 \times 200) = 4 \times 40K$ and $(3 \times 200) \times (3 \times 200) = 9 \times 40K$ neurons per layer, covering 5 diffractive layers, same as the standard D^2NN design. The comparison of the blind classification accuracies of these 5-layer D^2NN designs with $m = 1$, 4, and 9, all trained with $\Delta_{tr} = 33.92\lambda$, reveals that an increase in the width of the diffractive layers not only increases the input numerical aperture (NA) of the diffractive network, but also significantly improves the
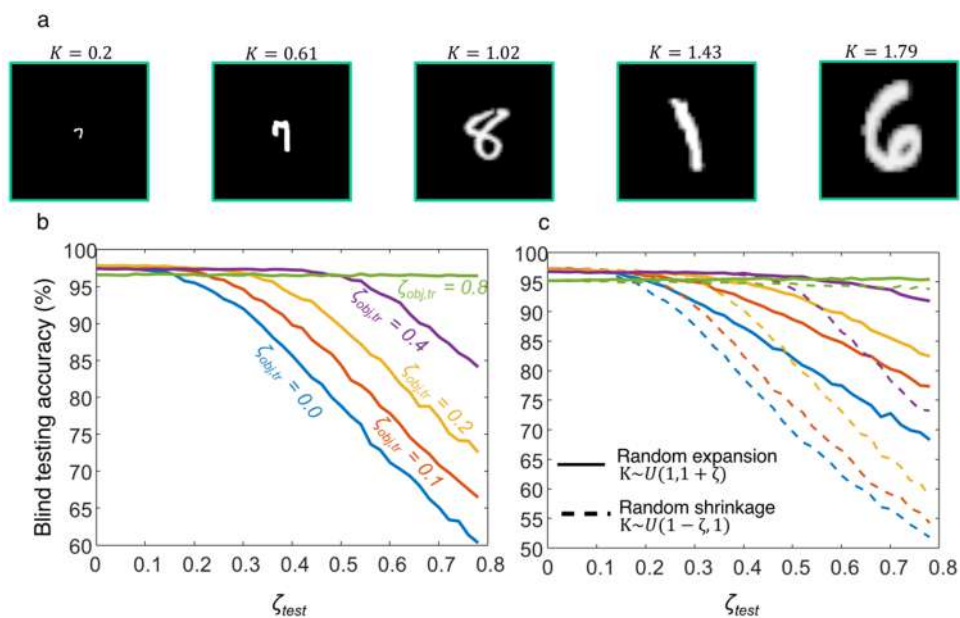
Figure 5. Scale-invariant diffractive optical networks. (a) Randomly scaled object examples from the MNIST test data set. Green frame around each object demonstrates the size of the diffractive layers. (b) The blind inference accuracies provided by five different D²NN models trained with ζ = ζ_{tr}, taken as 0.0 (blue), 0.1 (red), 0.2 (yellow), 0.4 (purple), and 0.8 (green); the resulting models were tested under different levels random object scaling with the parameter, ζ = ζ_{test}, swept from 0.0 to 0.8. (c) The classification performance of the diffractive networks in (b) for the case of expansion-only (solid curves) and shrinkage-only (dashed curves).

classification accuracies even under large Δ_{test} (see Figure 4b). For example, the D²NN design with Δ_{tr} = 33.92λ and m = 4 achieves classification accuracies of 83.08% and 85.76% for the testing conditions, Δ_{test} = 0.0λ and Δ_{test} = Δ_{tr} = 33.92λ, respectively. With the same Δ_{test} values, the diffractive network with m = 1, i.e., 40K neurons per layer can only achieve classification accuracies of 79.23% and 81.98%, respectively. The expansion of the diffractive layers to accommodate 9 × 40K neurons per layer (m = 9), further increases the mean classification accuracies over the entire Δ_{test} range, as illustrated in Figure 4b.

Next, we expanded the presented training approach to design diffractive optical network models that are resilient to the *scale* of the input objects. To this end, similar to eqs 1a and 1b, we defined a scaling parameter, $K \sim U(1 - \zeta, 1 + \zeta)$, randomly covering the scale range $(1 - \zeta, 1 + \zeta)$ determined by the hyperparameter, ζ. According to this formulation, for a given value of K, the physical size of the input object is scaled up ($K > 1$) or down ($K < 1$); see Figure 5a. On the basis of this formulation, in addition to the standard D²NN design with ζ_{tr} = 0, we trained 4 new diffractive network models with ζ_{tr} = 0.1, 0.2, 0.4, and 0.8. The resulting diffractive network models were then tested by sweeping ζ_{test} from 0 to 0.8 with steps of 0.02 and for each case, the classification accuracy on testing data attained by each diffractive model was computed (see Figure 5b). This analysis reveals that the resulting diffractive network designs are rather resilient to random scaling of the input objects, maintaining a competitive inference performance over a large range of object shrinkage or expansion (Figure 5b). Similar to the case shown in Figure 3, the relatively small values of ζ_{tr}, e.g., 0.1 (red curve in Figure 5b) or 0.2 (yellow curve in Figure 5b), effectively serve as data augmentation and the corresponding diffractive network models achieve higher peak inference accuracies of 97.84% (ζ_{tr} = 0.1) and 97.88% (ζ_{tr} = 0.2) compared to the 97.64% achieved by the standard

design (ζ_{tr} = 0). Furthermore, the comparison between the shift- and scale-invariant diffractive optical network models trained with Δ_{tr} = 16.96λ (green curve in Figure 3b) and ζ_{tr} = 0.8 (green curve in Figure 5b) is highly interesting since the effective FOVs induced by these two training parameters at the input/object plane are quite comparable, resulting in ~1.87× and 1.8× of the FOV of the standard design (Δ_{tr} = ζ_{tr} = 0), respectively. Despite these comparable effective FOVs at the input plane, the diffractive network trained against random scaling, ζ_{tr} = 0.8, achieves nearly ~6% higher inference accuracy compared to the shift-invariant design, Δ_{tr} = 16.96λ; in general, lateral random shifts of the input object with respect to a fixed diffractive network width seems to lower the inference accuracy of the diffractive models more than random object scaling, which indicates the physical importance of the trainable pixels/neurons within the central region of a diffractive network and the effectiveness of their optical communication with the neighboring layers. The mean classification accuracy provided by this scale-invariant diffractive optical network model (ζ_{tr} = 0.8) over the entire testing range, 0 < ζ_{test} < 0.8, is found to be 96.57% (Figure 5b), which is only ~1% lower than that of the standard diffractive design tested in the absence of random object scaling (ζ_{test} = 0).

To explore if there is a large performance gap between the classification accuracies attained for demagnified and magnified input objects, next we *separately* tested the diffractive optical network models in Figure 5b for the case of expansion-only, i.e., $K \sim U(1, 1 + \zeta)$ and shrinkage-only, i.e., $K \sim U(1 - \zeta, 1)$; see Figure 5c. A comparison of the solid (expansion-only) and the dashed (shrinkage-only) curves in Figure 5c reveals that, in general, diffractive networks' resilience toward object expansion and object shrinkage is similar. For instance, for the case of ζ_{tr} = 0.4 (purple curves in Figure 5c) the mean classification accuracy difference observed between the
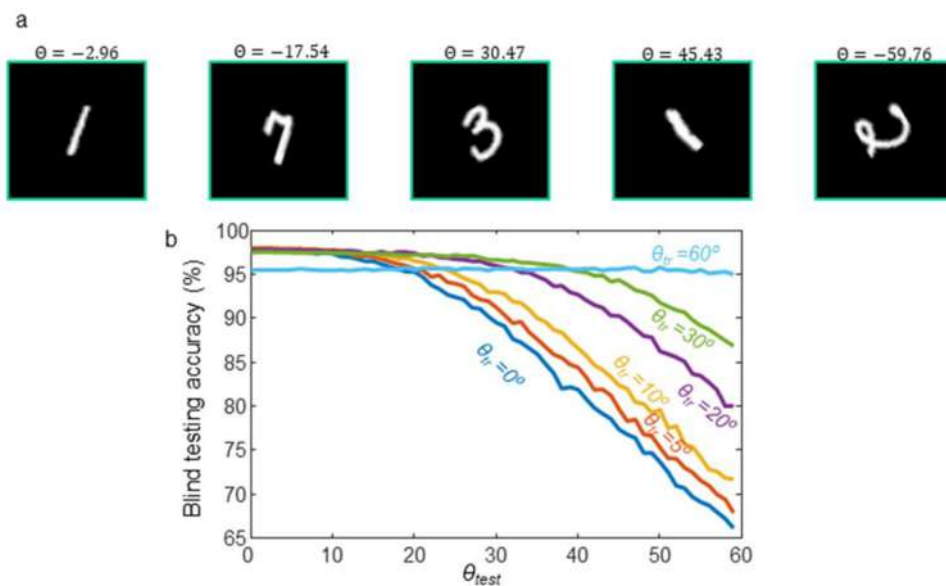
Figure 6. Rotation-invariant diffractive optical networks. (a) Randomly rotated object examples from the MNIST test data set. Green frame around each object demonstrates the size of the diffractive layers. (b) The blind inference accuracies provided by five different diffractive network models trained with $\theta = \theta_{tr}$, taken as $0°$ (blue), $5°$ (red), $10°$ (yellow), $20°$ (purple), $30°$ (green), and $60°$ (light-blue) when they were tested under different levels of random object rotations with the parameter, $\theta = \theta_{test}$, swept from $0°$ to $60°$, covering both clockwise and counterclockwise image rotations.

Table 1. Blind Inference Accuracy of the D²NN Models Trained against the Combinations of the Three Object Field Transformations Investigated in This Work: (Upper) Shift-Rotation, (Middle) Shiftscaling, (Lower) Rotation-Scaling

rotation / shift	$\theta_{tr} = \theta_{test} = 0°$	θ_{tr} and θ_{test} are iid $U(-10°, 10°)$	θ_{tr} and θ_{test} are iid $U(-20°, 20°)$
$\Delta_{tr} = \Delta_{test} = 0\lambda$	97.64	97.64	97.13
Δ_{tr} and Δ_{test} are iid $U(-2.12\lambda, 2.12\lambda)$	97.51	97.48	96.78
Δ_{tr} and Δ_{test} are iid $U(-8.48\lambda, 8.48\lambda)$	94.41	94.09	92.00

scaling / shift	$K_{tr} = K_{test} = 1$	K_{tr} and K_{test} are iid $U(0.6, 1.4)$	K_{tr} and K_{test} are iid $U(0.2, 1.8)$
$\Delta_{tr} = \Delta_{test} = 0\lambda$	97.64	97.29	96.50
Δ_{tr} and Δ_{test} are iid $U(-2.12\lambda, 2.12\lambda)$	97.51	96.79	95.76
Δ_{tr} and Δ_{test} are iid $U(-8.48\lambda, 8.48\lambda)$	94.41	91.71	89.20

rotation / scaling	$\theta_{tr} = \theta_{test} = 0°$	θ_{tr} and θ_{test} are iid $U(-10°, 10°)$	θ_{tr} and θ_{test} are iid $U(-20°, 20°)$
$K_{tr} = K_{test} = 1$	97.64	97.64	97.13
K_{tr} and K_{test} are iid $U(0.6, 1.4)$	97.29	96.98	96.26
K_{tr} and K_{test} are iid $U(0.2, 1.8)$	96.50	96.24	95.67

expansion-only vs shrinkage-only testing is only 0.04% up to the point that the testing range is equal to that of the training, i.e., $\zeta_{test} = \zeta_{tr}$. Similarly, for $\zeta_{tr} = 0.8$ the mean classification accuracy difference observed between the expansion-only vs shrinkage-only testing is ∼0.75%. When analyzing these results reported in Figure 5c, one should carefully consider the fact that for a fixed choice of ζ parameter there is an inherent asymmetry in expansion and shrinkage percentages; for example, for $\zeta_{test} = 0.8$, K can take values in the range (0.2, 1.8), where the extreme cases of 0.2 and 1.8 correspond to 5× shrinkage and 1.8× expansion of the input object, respectively. Therefore, the curves reported in Figure 5c for expansion-only vs shrinkage-only testing naturally contain different percentages of scaling with respect to the original size of the input objects.

Next, we expanded the presented framework to handle input object *rotations*. Figure 6 illustrates an equivalent analysis as in Figure 3, except that the input objects are now rotating, instead of shifting, around the optical axis, according to a uniformly distributed random rotation angle, $\Theta \sim U(-\theta, \theta)$, where $\Theta < 0$ and $\Theta > 0$ correspond to clockwise and counterclockwise rotation as depicted in Figure 1b, respectively. In this comparative analysis, six different diffractive network models trained with θ_{tr} values taken as 0° (standard design), 5°, 10°, 20°, 30°, and 60° were tested as a function of θ_{test} taking values between 0° and 60° with a step size of 1°, i.e., $\Theta \sim U(-\theta_{test}, \theta_{test})$. Similar to the case of scale-invariant designs reported in Figure 5, these diffractive network models trained with different θ_{tr} values can build up strong resilience against random object rotations, almost without a compromise in their inference. In fact, training with $\theta_{tr} \leq 20°$ (red, yellow, and purple curves in Figure 6b) improves the peak inference accuracy over the standard design ($\theta_{tr} = 0°$). When $\theta_{tr} = 30°$ (green curve in Figure 6b), the inference of the diffractive optical network is relatively flat as a function of θ_{test}, achieving a classification accuracy of 97.51% and 96.68% for $\theta_{test} = 0°$ and $\theta_{test} = 30°$, respectively, clearly demonstrating the advantages of the presented design framework.

Finally, we investigated the design of diffractive optical network models that were trained to simultaneously accommodate two of the three commonly encountered input objects transformations, i.e., random lateral shifting, scaling, and in-plane rotation. Table 1 reports the resulting classification accuracies of these newly trained D²NN models, where the inference performance of the corresponding diffractive optical network was tested with the same level of random object transformation as in the training, i.e., $\Delta_{tr} = \Delta_{test}$, $\zeta_{tr} = \zeta_{test}$, $\theta_{tr} = \theta_{test}$. The results in Table 1 reveal that these diffractive network designs can maintain their inference accuracies over 90%, building up resilience against unwanted, yet practically inevitable object transformations and variations. The thickness profile of the diffractive layers constituting the D²NN designs trained with the object transformation parameter pairs: ($\Delta_{tr} = 2.12\lambda$, $\theta_{tr} = 10°$), ($\Delta_{tr} = 2.12\lambda$, $\zeta_{tr} = 0.4$), and ($\theta_{tr} = 10°$, $\zeta_{tr} = 0.4$) reported in Table 1 are illustrated in Figure S1 of the Supporting Information. The confusion matrices provided by these three diffractive network models computed under $\Delta_{tr} = \Delta_{test}$, $\zeta_{tr} = \zeta_{test}$, and $\theta_{tr} = \theta_{test}$ are also reported in Figure S2.

CONCLUSIONS

In conclusion, we have quantified the sensitivity of diffractive optical networks against three fundamental object trans-

formations (lateral translation, scaling, and rotation), that are frequently encountered in various machine vision applications. Moreover, a new design scheme that formulates these input field transformations through uniformly distributed random variables as part of the optical forward model has been presented in deep learning-based training of D²NNs. Our analyses reveal that this training strategy significantly increases the robustness of diffractive networks against undesired object field transformations. Although we have taken input object classification as our target inference task, the presented ideas and the underlying methods can be extended to other optical machine learning tasks. As the presented training scheme enables the diffractive optical networks to achieve significantly higher inference accuracies in dynamic environments, we believe that this study will potentially expand the utilization of diffractive networks to a plethora of new applications that demand real-time monitoring and classification of fast and dynamic events.

METHODS

D²NN framework formulates the all-optical object classification problem from the point-of-view of training the physical features of matter inside a diffractive optical black-box. In this study, we modeled each D²NN using 5 successive modulation layers, each representing a two-dimensional, thin modulation component (Figure 1a). The optical modulation function of each diffractive layer was sampled with a period of 0.53λ over a regular 2D grid of coordinates, with each point representing the transmittance coefficient of a diffractive feature, i.e., an optical "neuron". Following earlier work,[36,38,39,42] we selected the material thickness, h, as the trainable physical parameter of each neuron,

$$h = Q_4\left(\frac{\sin(h_a) + 1}{2}(h_m - h_b)\right) + h_b \tag{2}$$

According to eq 2, the material thickness over each diffractive neuron is defined as a function of an auxiliary variable, h_a. The function, $Q_n(.)$, represents the n-bit quantization operator and h_m, h_b denote the predetermined hyperparameters of our forward model determining the allowable range of thickness values, $[h_b, h_m]$. The thickness in eq 2 is related to the transmittance coefficient of the corresponding diffractive neuron through the complex-valued refractive index (τ) of the optical material used to fabricate the resulting D²NN, i.e., $\tau(\lambda) = n(\lambda) + j\kappa(\lambda)$, with λ denoting the wavelength of the illumination light. On this basis, we can express the transmission coefficient, $t(x_q, y_p, z_k)$, of a diffractive neuron located at (x_q, y_p, z_k) as

$$t(x_q, y_p, z_k) = \exp\left(-\frac{2\pi\kappa h_{q,p}^k}{\lambda}\right)\exp\left(j(n - n_s)\frac{2\pi h_{q,p}^k}{\lambda}\right) \tag{3}$$

where $h_{q,p}^k$ refers to the material thickness over the corresponding neuron computed using eq 2, and n_s is the refractive index of the medium, surrounding the diffractive layers; without loss of generality, we assumed $n_s = 1$ (air). On the basis of the earlier demonstrations of diffractive optical networks,[31,36,38,39,42] we assumed the optical modulation surfaces in our diffractive optical networks are made of a material with $\tau = 1.7227 + j0.031$. Accordingly, the h_m and h_b were selected as 2λ and 0.66λ, respectively, as illustrated in Figure 2 and Figure S1.

The 2D complex modulation function, $T(x, y, z_k)$, of a diffractive surface, S_k, located at $z = z_k$, can be written as

$$T(x, y, z_k) = \sum_q \sum_p t(x_q, y_p, z_k) P(x - qw_x, y - pw_y, z_k)$$

(4)

where the w_x and w_y denote the width of a diffractive neuron in x and y directions, respectively (both taken as 0.53λ). $P(x, y, z_k)$ represents the 2D interpolation kernel which we assumed to be an ideal rectangular function in the following form,

$$P(x, y, z_k) = \begin{cases} 1, & |x| < \left(\dfrac{w_x}{2}\right) \text{ and } |y| < \left(\dfrac{w_y}{2}\right) \\ 0, & \text{otherwise} \end{cases}$$

(5)

The light propagation in the presented diffractive optical networks were modeled based on the digital implementation of the Rayleigh−Sommerfeld diffraction equation, using an impulse response defined as

$$w(x, y, z) = \frac{z}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right)$$

(6)

where $r = \sqrt{x^2 + y^2 + z^2}$. On this basis, the wave field synthesized over a surface at $z = z_{k+1}$, $U(x, y, z_{k+1})$, by a trainable diffractive layer, S_k, located at $z = z_k$ can expressed as

$$U(x, y, z_{k+1}) = U'(x, y, z_k) * w(x, y, z_{k+1} - z_k)$$

(7)

where $U'(x, y, z_k) = U(x, y, z_k)T(x, y, z_k)$ is the complex wave field immediately after the diffractive layer, k, and * denotes the 2D convolution operation. In this optical forward model, the layer-to-layer distances were taken as 40λ for the diffractive network architectures that have 40K neurons on each diffractive layer to induce optical connections between all the neurons of two successive diffractive layers based on eq 6. To provide a fair comparison, for the diffractive network architectures with $m = 4$ and $m = 9$ fold larger diffractive layers as depicted in Figure 4b, the layer-to-layer distances were also accordingly increased to be $(m)^{0.5} \times 40\lambda$, preserving the all-optical connections set by the diffraction cone angle between successive layers of these network models. Therefore, the improvement in inference accuracy for randomly shifting objects demonstrated in Figure 4b comes at the expense of using larger diffractive layers separated with larger distances, increasing both the lateral and the axial size of the diffractive network.

On the basis of the above outlined optical forward model, if we let the complex-valued object transmittance, $T(x, y, z_0)$, over the input FOV be located at a surface defined with $k = 0$, then the complex field and the associated optical intensity distribution at the *output/detector plane* of a 5-layer diffractive optical network architecture shown in Figure 1a, can be expressed as $U(x, y, z_6)$ and $I = |U(x, y, z_6)|^2$, respectively. In our forward training model, we assumed that each class detector collects an optical signal, Γ_c, that is computed through the integration of the output intensity, I, over the corresponding detector active area ($6.4\lambda \times 6.4\lambda$ per detector). For a given data set with C classes, the standard D^2NN architecture in Figure 1a employs C detectors at the output plane, each representing a data class; C = 10 for MNIST data set. Accordingly, at each training iteration, after the propagation of the input object to the output plane (based

on eqs 6 and 7), a vector of optical signals, $\boldsymbol{\Gamma}$, is formed and then normalized to get $\boldsymbol{\Gamma}'$ using the following relationship:

$$\boldsymbol{\Gamma}' = \frac{\boldsymbol{\Gamma}}{\max\{\boldsymbol{\Gamma}\}} \times T_s$$

(8)

where T_s is a constant temperature parameter.[43,44] Next, the class score of the c-th data class, O_c, is computed as

$$O_c = \frac{\exp(\Gamma'_c)}{\sum_{c \in C} \exp(\Gamma'_c)}$$

(9)

In eq 9, Γ'_c denotes the normalized optical signal collected by the detector, c, computed as in eq 8. At the final step, the classification loss function, $\mathcal{L}$, in the form of the cross-entropy loss defined in eq 10 is computed for the subsequent error-backpropagation and update of the diffractive layers:

$$\mathcal{L} = -\sum_{c \in C} g_c \log(O_c)$$

(10)

where $\mathbf{g}$ denotes the one-hot ground truth label vector.

For the digital implementation of the diffractive optical network training outlined above, we developed a custom-written code in Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.). The backpropagation updates were calculated using the Adam[45] optimizer with its parameters set to be the default values as defined by TensorFlow and kept identical in each model. The learning rate was set to be 0.001 for all the diffractive network models presented here. The training batch sizes were taken as 50 and 20 for the diffractive network designs with 40K neurons per layer and wider diffractive networks reported in Figure 4b, respectively. The training of a 5-layer diffractive optical network with 40K diffractive neurons per layer takes ∼6 h using a computer with a GeForce GTX 1080 Ti Graphical Processing Unit (GPU, Nvidia Inc.) and Intel Core i7−8700 Central Processing Unit (CPU, Intel Inc.) with 64 GB of RAM, running Windows 10 operating system (Microsoft). The training of a wider diffractive network presented in Figure 4b, on the other hand, takes ∼30 h based on the same system configuration due to the larger light propagation windows used in the forward optical model. Since the investigated object transformations were implemented through a custom-developed bilinear interpolation code written based on TensorFlow functions, it only takes ∼50 s longer to complete an epoch with the presented scheme compared to the standard training of D^2NNs.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphotonics.0c01583.

Supplementary Figures S1 and S2 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Aydogan Ozcan − *Electrical and Computer Engineering Department, Bioengineering Department, California NanoSystems Institute, and Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, California 90095, United States;* ⓘ orcid.org/0000-0002-0717-683X; Email: ozcan@ucla.edu

Authors

Deniz Mengu − *Electrical and Computer Engineering Department, Bioengineering Department, and California NanoSystems Institute, University of California, Los Angeles, California 90095, United States*

Yair Rivenson − *Electrical and Computer Engineering Department, Bioengineering Department, and California NanoSystems Institute, University of California, Los Angeles, California 90095, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsphotonics.0c01583

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016.

(2) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436−444.

(3) Rivenson, Y.; Wang, H.; Wei, Z.; de Haan, K.; Zhang, Y.; Wu, Y.; Günaydın, H.; Zuckerman, J. E.; Chong, T.; Sisk, A. E.; Westbrook, L. M.; Wallace, W. D.; Ozcan, A. Virtual Histological Staining of Unlabelled Tissue-Autofluorescence Images via Deep Learning. *Nat. Biomed Eng.* **2019**, *3* (6), 466−477.

(4) Wu, Y.; Rivenson, Y.; Wang, H.; Luo, Y.; Ben-David, E.; Bentolila, L. A.; Pritz, C.; Ozcan, A. Three-Dimensional Virtual Refocusing of Fluorescence Microscopy Images Using Deep Learning. *Nat. Methods* **2019**, *16* (12), 1323−1331.

(5) Rivenson, Y.; Wu, Y.; Ozcan, A. Deep Learning in Holography and Coherent Imaging. *Light: Sci. Appl.* **2019**, *8* (1), 1−8.

(6) Wu, Y.; Luo, Y.; Chaudhari, G.; Rivenson, Y.; Calis, A.; de Haan, K.; Ozcan, A. Bright-Field Holography: Cross-Modality Deep Learning Enables Snapshot 3D Imaging with Bright-Field Contrast Using a Single Hologram. *Light: Sci. Appl.* **2019**, DOI: 10.1038/s41377-019-0139-9.

(7) Amodei, D.; Ananthanarayanan, S.; Anubhai, R.; Bai, J.; Battenberg, E.; Case, C.; Casper, J.; Catanzaro, B.; Cheng, Q.; Chen, G.; Chen, J.; Chen, J.; Chen, Z.; Chrzanowski, M.; Coates, A.; Diamos, G.; Ding, K.; Du, N.; Elsen, E.; Engel, J.; Fang, W.; Fan, L.; Fougner, C.; Gao, L.; Gong, C.; Hannun, A.; Han, T.; Johannes, L. V.; Jiang, B.; Ju, C.; Jun, B.; LeGresley, P.; Lin, L.; Liu, J.; Liu, Y.; Li, W.; Li, X.; Ma, D.; Narang, S.; Ng, A.; Ozair, S.; Peng, Y.; Prenger, R.; Qian, S.; Quan, Z.; Raiman, J.; Rao, V.; Satheesh, S.; Seetapun, D.; Sengupta, S.; Srinet, K.; Sriram, A.; Tang, H.; Tang, L.; Wang, C.; Wang, J.; Wang, K.; Wang, Y.; Wang, Z.; Wang, Z.; Wu, S.; Wei, L.; Xiao, B.; Xie, W.; Xie, Y.; Yogatama, D.; Yuan, B.; Zhan, J.; Zhu, Z. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin, *arXiv*, Dec 8, 2015, arXiv:1512.02595.

(8) Liu, D.; Tan, Y.; Khoram, E.; Yu, Z. Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. *ACS Photonics* **2018**, *5* (4), 1365−1369.

(9) Rahmani, B.; Loterie, D.; Konstantinou, G.; Psaltis, D.; Moser, C. Multimode Optical Fiber Transmission with a Deep Learning Network. *Light: Sci. Appl.* **2018**, *7* (1), 1−11.

(10) Ma, W.; Cheng, F.; Liu, Y. Deep-Learning-Enabled On-Demand Design of Chiral Metamaterials. *ACS Nano* **2018**, *12* (6), 6326−6334.

(11) Malkiel, I.; Mrejen, M.; Nagler, A.; Arieli, U.; Wolf, L.; Suchowski, H. Plasmonic Nanostructure Design and Characterization via Deep Learning. *Light: Sci. Appl.* **2018**, *7* (1), 60.

(12) Wu, Y.; Rivenson, Y.; Zhang, Y.; Wei, Z.; Günaydın, H.; Lin, X.; Ozcan, A. Extended Depth-of-Field in Holographic Imaging Using Deep-Learning-Based Autofocusing and Phase Recovery. *Optica* **2018**, *5* (6), 704.

(13) Li, Y.; Xue, Y.; Tian, L. Deep Speckle Correlation: A Deep Learning Approach toward Scalable Imaging through Scattering Media. *Optica, OPTICA* **2018**, *5* (10), 1181−1190.

(14) Rivenson, Y.; Liu, T.; Wei, Z.; Zhang, Y.; de Haan, K.; Ozcan, A. PhaseStain: The Digital Staining of Label-Free Quantitative Phase Microscopy Images Using Deep Learning. *Light: Sci. Appl.* **2019**, *8* (1), 23.

(15) Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A. L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2018**, *40* (4), 834−848.

(16) Zhang, Y.; de Haan, K.; Rivenson, Y.; Li, J.; Delis, A.; Ozcan, A. Digital Synthesis of Histological Stains Using Micro-Structured and Multiplexed Virtual Staining of Label-Free Tissue. *Light: Sci. Appl.* **2020**, DOI: 10.1038/s41377-020-0315-y.

(17) Feldmann, J.; Youngblood, N.; Wright, C. D.; Bhaskaran, H.; Pernice, W. H. P. All-Optical Spiking Neurosynaptic Networks with Self-Learning Capabilities. *Nature* **2019**, *569* (7755), 208−214.

(18) Hughes, T. W.; Williamson, I. A. D.; Minkov, M.; Fan, S. Wave Physics as an Analog Recurrent Neural Network. *Science Advances* **2019**, *5* (12), eaay6946.

(19) Shastri, B. J.; Tait, A. N.; Ferreira de Lima, T.; Nahmias, M. A.; Peng, H.-T.; Prucnal, P. R. Neuromorphic Photonics, Principles Of. In *Encyclopedia of Complexity and Systems Science*; Meyers, R. A., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2018; pp 1−37. DOI: 10.1007/978-3-642-27737-5_702-1.

(20) Van der Sande, G.; Brunner, D.; Soriano, M. C. Advances in Photonic Reservoir Computing. *Nanophotonics* **2017**, *6* (3), 561−576.

(21) Cichos, F.; Gustavsson, K.; Mehlig, B.; Volpe, G. Machine Learning for Active Matter. *Nature Machine Intelligence* **2020**, *2* (2), 94−103.

(22) Bueno, J.; Maktoobi, S.; Froehly, L.; Fischer, I.; Jacquot, M.; Larger, L.; Brunner, D. Reinforcement Learning in a Large-Scale Photonic Recurrent Neural Network. *Optica* **2018**, *5* (6), 756.

(23) George, J. K.; Mehrabian, A.; Amin, R.; Meng, J.; de Lima, T. F.; Tait, A. N.; Shastri, B. J.; El-Ghazawi, T.; Prucnal, P. R.; Sorger, V. J. Neuromorphic Photonics with Electro-Absorption Modulators. *Opt. Express* **2019**, *27* (4), 5181.

(24) Shen, Y.; Harris, N. C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.; Larochelle, H.; Englund, D.; Soljačić, M. Deep Learning with Coherent Nanophotonic Circuits. *Nat. Photonics* **2017**, *11* (7), 441−446.

(25) Tait, A. N.; de Lima, T. F.; Zhou, E.; Wu, A. X.; Nahmias, M. A.; Shastri, B. J.; Prucnal, P. R. Neuromorphic Photonic Networks Using Silicon Photonic Weight Banks. *Sci. Rep.* **2017**, DOI: 10.1038/s41598-017-07754-z.

(26) George, J.; Amin, R.; Mehrabian, A.; Khurgin, J.; El-Ghazawi, T.; Prucnal, P. R.; Sorger, V. J. Electrooptic Nonlinear Activation Functions for Vector Matrix Multiplications in Optical Neural Networks. In *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*; OSA: Zurich, 2018; p SpW4G.3. DOI: 10.1364/SPPCOM.2018.SpW4G.3.

(27) Mehrabian, A.; Al-Kabani, Y.; Sorger, V. J.; El-Ghazawi, T. PCNNA: A Photonic Convolutional Neural Network Accelerator. In *2018 31st IEEE International System-on-Chip Conference (SOCC)*; 2018; pp 169−173. DOI: 10.1109/SOCC.2018.8618542.

(28) Miscuglio, M.; Mehrabian, A.; Hu, Z.; Azzam, S. I.; George, J.; Kildishev, A. V.; Pelton, M.; Sorger, V. J. All-Optical Nonlinear Activation Function for Photonic Neural Networks [Invited]. *Opt. Mater. Express* **2018**, *8* (12), 3851.

(29) Estakhri, N. M.; Edwards, B.; Engheta, N. Inverse-Designed Metastructures That Solve Equations. *Science* **2019**, *363* (6433), 1333−1338.

(30) Hughes, T. W.; Minkov, M.; Shi, Y.; Fan, S. Training of Photonic Neural Networks through in Situ Backpropagation and Gradient Measurement. *Optica* **2018**, *5* (7), 864.

(31) Lin, X.; Rivenson, Y.; Yardimci, N. T.; Veli, M.; Luo, Y.; Jarrahi, M.; Ozcan, A. All-Optical Machine Learning Using Diffractive Deep Neural Networks. *Science* **2018**, *361* (6406), 1004−1008.

(32) Prabhu, M.; Roques-Carmes, C.; Shen, Y.; Harris, N.; Jing, L.; Carolan, J.; Hamerly, R.; Baehr-Jones, T.; Hochberg, M.; Čeperić, V.; Joannopoulos, J. D.; Englund, D. R.; Soljačić, M. A Recurrent Ising Machine in a Photonic Integrated Circuit. *arXiv*, Sep 30, 2019, arXiv:1909.13877.

(33) Li, J.; Mengu, D.; Luo, Y.; Rivenson, Y.; Ozcan, A. Class-Specific Differential Detection in Diffractive Optical Neural Networks Improves Inference Accuracy. *Adv. Photon.* **2019**, *1* (4), 046001.

(34) Rahman, S. S.; Li, J.; Mengu, D.; Rivenson, Y.; Ozcan, A. Ensemble Learning of Diffractive Optical Networks. *arXiv*, Sep 15, 2020, arXiv:2009.06869.

(35) Mengu, D.; Luo, Y.; Rivenson, Y.; Ozcan, A. Analysis of Diffractive Optical Neural Networks and Their Integration With Electronic Neural Networks. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26* (1), 1−14.

(36) Li, J.; Mengu, D.; Yardimci, N. T.; Luo, Y.; Li, X.; Veli, M.; Rivenson, Y.; Jarrahi, M.; Ozcan, A. Machine Vision Using Diffractive Spectral Encoding. *arXiv*, May 15, 2020, arXiv:2005.11387.

(37) Kulce, O.; Mengu, D.; Rivenson, Y.; Ozcan, A. All-Optical Information Processing Capacity of Diffractive Surfaces. *arXiv*, Jul 25, 2020, arXiv:2007.12813.

(38) Veli, M.; Mengu, D.; Yardimci, N. T.; Luo, Y.; Li, J.; Rivenson, Y.; Jarrahi, M.; Ozcan, A. Terahertz Pulse Shaping Using Diffractive Legos. *arXiv*, Jun 30, 2020, arXiv:2006.16599.

(39) Luo, Y.; Mengu, D.; Yardimci, N. T.; Rivenson, Y.; Veli, M.; Jarrahi, M.; Ozcan, A. Design of Task-Specific Optical Systems Using Broadband Diffractive Neural Networks. *Light: Sci. Appl.* **2019**, *8* (1), 112.

(40) Psaltis, D.; Brady, D.; Gu, X.-G.; Lin, S. Holography in Artificial Neural Networks. *Nature* **1990**, *343* (6256), 325−330.

(41) Wagner, K.; Psaltis, D. Multilayer Optical Learning Networks. *Appl. Opt.* **1987**, *26* (23), 5061−5076.

(42) Mengu, D.; Zhao, Y.; Yardimci, N. T.; Rivenson, Y.; Jarrahi, M.; Ozcan, A. Misalignment Resilient Diffractive Optical Networks. *Nanophotonics* **2020**, *9* (13), 4207−4219.

(43) Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K. Q. On Calibration of Modern Neural Networks. *arXiv*, Jun 14, 2017, arXiv:1706.04599.

(44) Laha, A.; Chemmengath, S. A.; Agrawal, P.; Khapra, M.; Sankaranarayanan, K.; Ramaswamy, H. G. On Controllable Sparse Alternatives to Softmax. In *Advances in Neural Information Processing Systems 31*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc., 2018; pp 6422−6432.

(45) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv*, Dec 22, 2014, arXiv:1412.6980.