**Research article**

Deniz Mengu, Yifan Zhao, Nezih T. Yardimci, Yair Rivenson, Mona Jarrahi and Aydogan Ozcan\*

# Misalignment resilient diffractive optical networks

**Abstract:** As an optical machine learning framework, Diffractive Deep Neural Networks ($D^2NN$) take advantage of data-driven training methods used in deep learning to devise light–matter interaction in 3D for performing a desired statistical inference task. Multi-layer optical object recognition platforms designed with this diffractive framework have been shown to generalize to unseen image data achieving, e.g., >98% blind inference accuracy for hand-written digit classification. The multi-layer structure of diffractive networks offers significant advantages in terms of their diffraction efficiency, inference capability and optical signal contrast. However, the use of multiple diffractive layers also brings practical challenges for the fabrication and alignment of these diffractive systems for accurate optical inference. Here, we introduce and experimentally demonstrate a new training scheme that significantly increases the robustness of diffractive networks against 3D misalignments and fabrication tolerances in the physical implementation of a trained diffractive network. By modeling the undesired layer-to-layer misalignments in 3D as continuous random variables in the optical forward model, diffractive networks are trained to maintain their inference accuracy over a large range of misalignments; we term this diffractive network design as *vaccinated* $D^2NN$ (v-$D^2NN$). We further extend this vaccination strategy to the training of diffractive networks that use differential detectors at the output plane as well as to jointly-trained hybrid (optical-electronic) networks to reveal that all of these diffractive designs improve their resilience to misalignments by taking into account possible 3D fabrication variations and displacements during their training phase.

# 1 Introduction

Deep learning has been redefining the state-of-the-art for processing various signals collected and digitized by different sensors, monitoring physical processes for, e.g., biomedical image analysis [1–4], speech recognition [5, 6] and holography [7–10], among many others [11–17]. Furthermore, deep learning and related optimization tools have been harnessed to find data-driven solutions for various inverse problems arising in, e.g., microscopy [18–22], nanophotonic designs and plasmonics [23–25]. These demonstrations and others have been motivating some of the recent advances in optical neural networks and related optical computing techniques that aim to exploit the computational speed, power-efficiency, scalability and parallelization capabilities of optics for machine intelligence applications [26–45].

Toward this broad goal, Diffractive Deep Neural Networks ($D^2NN$) [36–39] have been introduced as a machine learning framework that unifies deep learning-based training of matter with the physical models governing light propagation to enable all-optical inference through a set of diffractive layers. The training stage of a diffractive network is performed using a computer and relies on deep learning and error backpropagation methods to tailor the light–matter interaction across a set of diffractive layers that collectively perform a given machine learning task,

**\*Corresponding author: Aydogan Ozcan**, Electrical and Computer Engineering Department, University of California, Los Angeles, CA, 90095, USA; Bioengineering Department, University of California, Los Angeles, CA, 90095, USA; California NanoSystems Institute, University of California, Los Angeles, CA, 90095, USA; and Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA, E-mail: ozcan@ucla.edu. https://orcid.org/0000-0002-0717-683X
**Deniz Mengu and Yair Rivenson:** Electrical and Computer Engineering Department, University of California, Los Angeles, CA, 90095, USA; Bioengineering Department, University of California, Los Angeles, CA, 90095, USA; California NanoSystems Institute, University of California, Los Angeles, CA, 90095, USA. https://orcid.org/0000-0003-1051-1501 (D. Mengu)
**Yifan Zhao, Nezih T. Yardimci and Mona Jarrahi:** Electrical and Computer Engineering Department, University of California, Los Angeles, CA, 90095, USA; California NanoSystems Institute, University of California, Los Angeles, CA, 90095, USA

e.g., object classification. Previous studies on D²NNs have demonstrated the generalization capability of these multi-layer diffractive network designs to new, unseen image data. For example, using a 5-layer diffractive network architecture, >98% and >90% all-optical blind testing accuracies have been reported [38] for the classification of the images of handwritten digits (MNIST) [46] and fashion products (Fashion-MNIST) [47] that are encoded in the amplitude and phase channels of the input plane, respectively. Successful experimental demonstrations of these all-optical classification systems have been reported using 3D-printed diffractive layers that conduct inference by modulating the incoming object wave at terahertz (THz) wavelengths.

Despite the lack of nonlinear optical elements in these previous implementations, diffractive optical networks have been shown to offer significant advantages in terms of (1) inference accuracy, (2) diffraction efficiency and (3) signal contrast, when the number of successive diffractive layers in the network design is increased [37]. A similar depth advantage was also demonstrated in [39], where instead of a statistical inference task such as image classification, the D²NN framework was utilized to solve an inverse design problem to achieve, e.g., spatially-controlled wavelength de-multiplexing of a broadband source. While these multi-layer diffractive architectures offer significantly better performance for generalization and application-specific design merits, they also pose practical challenges for the fabrication and opto-mechanical assembly of these trained diffractive models.

Here, we present a training scheme that substantially increases the robustness of diffractive optical networks against physical misalignments and fabrication tolerances. Our scheme models and introduces these undesired system variations and layer-to-layer misalignments as continuous random variables during the deep learning-based training of the diffractive model to significantly improve the error tolerance margins of diffractive optical networks; this process of introducing random misalignments during the training phase will be termed as *vaccination* of the diffractive network, and the resulting designs will be referred to as vaccinated D²NNs (v-D²NNs). To demonstrate the efficacy of our strategy, we trained diffractive network models composed of five diffractive layers for all-optical classification of handwritten digits, where we utilized in the training phase independent and uniformly distributed displacement/misalignment vectors for $x$, $y$, and $z$ directions of each diffractive layer. Our results indicate that v-D²NN framework enables the design of diffractive optical networks that can maintain their object recognition performance against severe layer-to-layer

misalignments, providing nearly flat blind inference accuracies within the displacement/misalignment range adopted in the training.

To experimentally demonstrate the success of v-D²NN framework, we also compared two 3D-printed diffractive networks, each with five diffractive layers that were designed for hand-written digit classification under monochromatic THz illumination ($\lambda$ = ~0.75 mm): the first network model was designed without the presence of any misalignments (non-vaccinated) and the second one was designed as a v-D²NN. After the fabrication of each diffractive network, the third diffractive layer was on purpose misaligned to different 3D positions around its ideal location. The experimental results confirmed our numerical analysis to reveal that the v-D²NN design can preserve its inference accuracy despite a wide range of physical misalignments, while the standard D²NN design frequently failed to recognize the correct data class due to these purposely introduced misalignments.

We also combined our v-D²NN framework with the differential diffractive optical networks [38] and the jointly-trained optical-electronic (hybrid) neural network systems. Differential diffractive classification systems assign a pair of detectors (generating one positive and one negative signal) for each data class to mitigate the strict non-negativity constraint of optical intensity and were demonstrated to offer superior inference accuracy compared to standard diffractive designs [38]. When trained against misalignments using the presented v-D²NN framework, differential diffractive networks are also shown to preserve their performance advantages for all-optical classification. However, both differential and standard diffractive networks fall short in matching the adaptation capabilities of a hybrid diffractive network system that uses a modest, single-layer fully connected architecture with only 110 learnable parameters in the electronic domain, following the diffractive optical front-end.

In addition to misalignment-related errors, the presented vaccination framework can also be adopted to mitigate other error sources in diffractive network models, e.g., detection noise and fabrication imperfections or artefacts, provided that the approximate analytical models and the probability distributions of these factors are utilized during the training stage. We anticipate that v-D²NNs will be the gateway of diffractive optical networks and the related hybrid neural network schemes toward practical machine vision and sensing applications, by mitigating various sources of error between the training forward models and the corresponding physical hardware implementations. Furthermore, the presented methodology of designing misalignment and noise resilient physical

machine learning models can be broadly applicable to other optical learning platforms, regardless of their physical dimensions and selected operation wavelengths.

## 2 Results

Figure 1 illustrates three different types of diffractive optical network-based object recognition systems investigated in this work. We focused on 5-layer diffractive optical

network architectures as shown in Figure 1 that are fully-connected, meaning that the half cone angle of the secondary wave created by the diffractive features (neurons) of size, e.g., $\delta = 0.53\ \lambda$, is large enough to enable communication between all the features on two successive diffractive layers that are placed, e.g., 40 $\lambda$ apart in axial direction. On the transverse plane, each diffractive layer extends from $-100 \times \delta$ to $100 \times \delta$ on $x$ and $y$ directions around the optical axis, and therefore the edge length of each diffractive surface in total is $200 \times \delta$ ($\sim$106.66 $\lambda$). With this
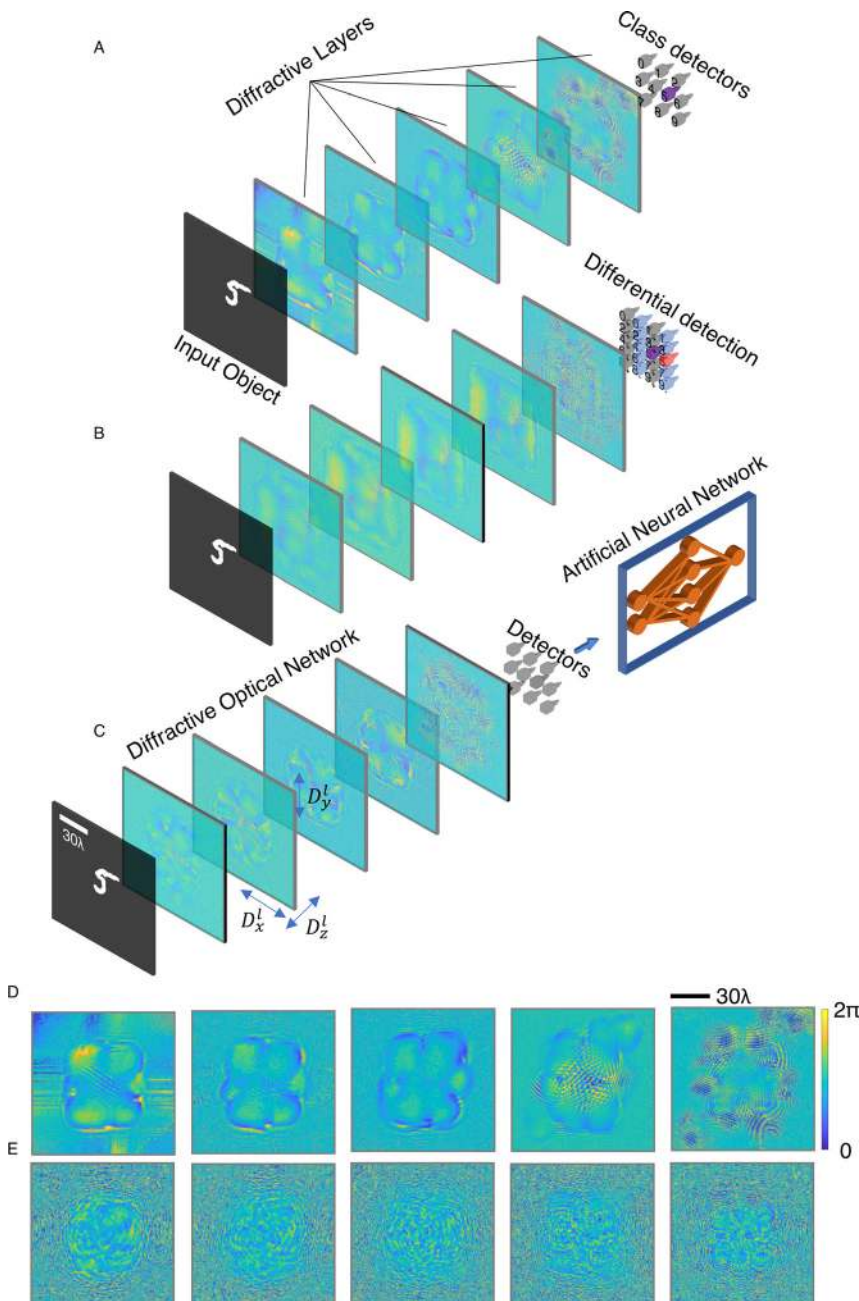


**Figure 1:** Different types of D$^2$NN-based image classification systems. (A) Standard D$^2$NN framework trained for all-optical classification of handwritten digits. Each detector at the output plane represents a data class. (B) Differential D$^2$NN trained for all-optical classification of handwritten digits. Each data class is represented by a pair of detectors at the output plane, where the normalized difference between these detector pairs represents the class scores. (C) Jointly-trained hybrid (optical-electronic) network system trained for classification of handwritten digits. The optical signals collected at the output detectors are used as inputs to the electronic neural network at the back-end, which is used to output the final class scores. (D) Phase profiles computed by the deep learning-based training for a 5-layer diffractive optical network that is vaccinated against both lateral and axial misalignments for the task of handwritten digit classification. The layers of this diffractive network were fabricated using 3D printing as shown in Figure 5D and experimentally tested using the setup shown in Figure 5E. (E) Same as D, except the diffractive network represents a non-vaccinated, error-free design.

outlined diffractive network architecture, the standard D²NN training routine updates the trainable parameters of the diffractive layers at every iteration based on the mean gradient computed over a batch of training samples with respect to a loss function, specifically tailored for the desired optical machine learning application, e.g., cross-entropy for supervised object recognition systems [37], until a convergence criterion is satisfied. Since this conventional training approach assumes perfect alignment throughout the training, the sources of statistical variations in the resulting model are limited to the initial condition of the diffractive network parameter space and the sequence of the training data introduced to the network.

## 2.1 Training and testing of v-D²NNs

The training of vaccinated diffractive optical networks mainly follows the same steps as the standard D²NN framework; except, it additionally incorporates system errors, e.g., misalignments, based on their probability distribution functions into the optical forward model. In this work, we modeled each orthogonal component of the undesired 3D displacement vector of each diffractive layer, $D = (D_x, D_y, D_z)$, as uniformly distributed, independent random variables as follows;

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

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

$$D_Z \sim U(-\Delta_z, \Delta_z) \tag{1c}$$

where $\Delta_\star$ denotes the shift along the corresponding axis, $(\star)$, reflecting the uncertainty in our physical assembly/fabrication of the diffractive model. During the training, the random displacement vector of *each* diffractive layer, **D**, takes different values sampled from the probability distribution of its components, $D_x$, $D_y$ and $D_z$, for each batch of training samples. Consequently, the location of layer $l$ at $i$th iteration/batch, $\mathbf{L}^{(l,i)}$, can be expressed as;

$$\boldsymbol{L}^{(l,i)} = \left(L_x^1, L_y^1, L_z^1\right) + \left(D_x^{(l,i)}, D_y^{(l,i)}, D_z^{(l,i)}\right) \tag{2}$$

where the first and the second vectors on the right-hand side denote the ideal location of the diffractive layer $l$, and a random realization of the displacement vector, $\mathbf{D}^{(l,i)}$, of layer $l$ at the training iteration $i$, respectively. The displacement vector of each layer is independently determined, i.e., each layer of a diffractive network model can move within the displacement ranges depicted in Eq. (1) without any dependence on the locations of the other diffractive layers.

Opto-mechanical assembly and fabrication systems, in general, use different mechanisms to control the lateral and axial positioning of optical components. Therefore, we split our numerical investigation of the vaccination process into two: the lateral and axial misalignment cases. For the vaccination of diffractive optical network models against layer-to-layer misalignments on the transverse plane, we assumed $D_x$ and $D_y$ are i.i.d random variables during the training, i.e., they are independent with a parameter of $\Delta_x = \Delta_y = \Delta_{tr}$, and D was set to be 0. The axial case, on the other hand, sets $\Delta_{tr}$ to be 0 throughout the training leaving $D_z \sim U(-\Delta_{z,tr}, \Delta_{z,tr})$ as the only source of inter-layer misalignments.

Following a similar path with the training, the blind testing of the presented diffractive network models updates the random displacement vector of each layer $l$, $\mathbf{D}^{(l,m)}$, for *each* test sample $m$. The reported accuracies throughout our analyses reflects the *blind testing accuracies* computed over the 10K image test set of MNIST digits where *each test sample propagates through a diffractive network model that experiences a different realization of the random variables* depicted in Eq. (1) for each diffractive layer, i.e., *there are 10K different configurations that a diffractive network model was misaligned throughout the testing stage*. Furthermore, similar to the training process, during the blind testing against lateral misalignments, it was assumed that $D_x$ and $D_y$ are i.i.d random variables with $\Delta_x = \Delta_y = \Delta_{test}$, and similarly, the axial displacements or misalignments were determined by $D_z \sim U(-\Delta_{z,\,test}, \Delta_{z,\,test})$.

## 2.2 Misalignment analysis of all-optical and hybrid diffractive systems

Figure 2A and D illustrate the blind testing accuracies provided by the standard diffractive optical network architecture (Figure 1A) trained against various levels of undesired axial and lateral misalignments, respectively. Focusing on the testing accuracy curve obtained by the error-free design (dark blue) in Figure 2A and D, it can be noticed that the diffractive optical networks are more susceptible to lateral misalignments compared to axial misalignments. For instance, when $\Delta_{test}$ is taken as 2.12 $\lambda$, inducing random lateral fluctuations on each diffractive layer's location around the optical axis, the blind testing accuracy achieved by the non-vaccinated standard diffractive optical network decreases to 38.40% from 97.77% (obtained in the absence of misalignments). As we further increase the level of lateral misalignments, the error-free diffractive optical network almost completely loses its inference capability by achieving, e.g., 19.24% blind
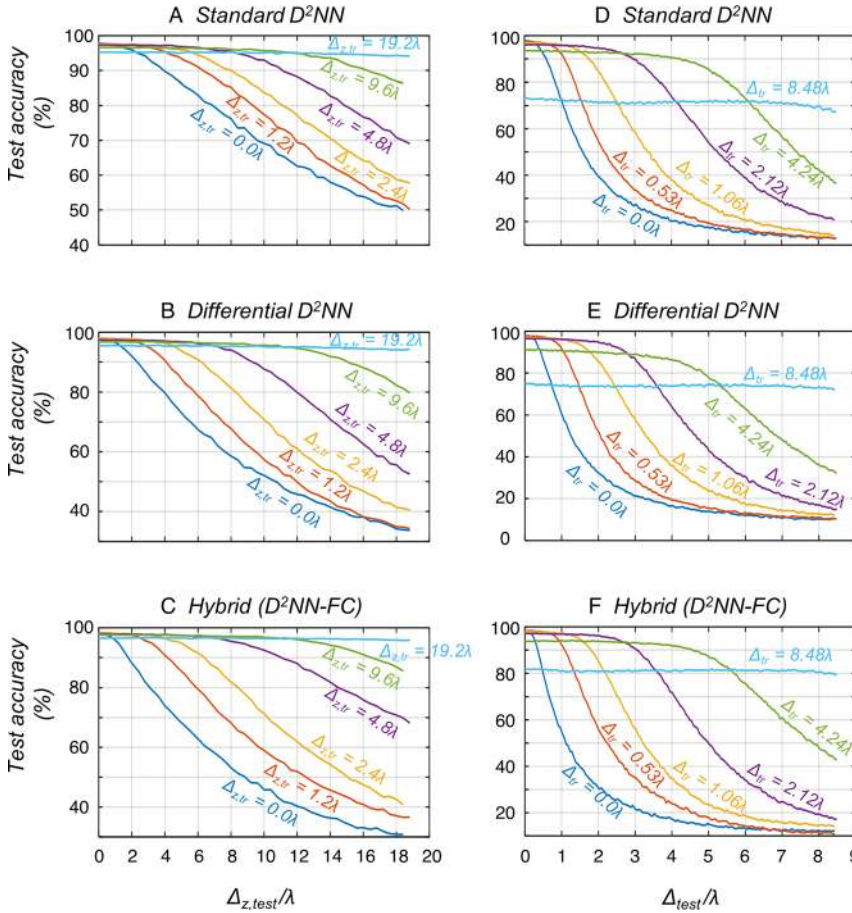
**Figure 2:** The sensitivity of the blind inference accuracies of different types of D²NN-based object classification systems against various levels of misalignments. (A) Standard D²NN systems trained for all-optical handwritten digit classification with and without vaccination were tested against various levels of *axial* misalignments, determined by $\Delta_{z,test}$. (B) Same as A, except for differential D²NN architectures. (C) Same as A and B, except for hybrid (D²NN-FC) systems comprised of a jointly-trained 5-layer D²NN optical front-end and a single-layer fully-connected neural network at the electronic back-end, combined through 10 discrete opto-electronic detectors (see Figure 1C). The comparison of these blind testing results reveals that as the axial misalignment increases during the training, $\Delta_{Z,tr}$, the inference accuracy of these machine vision systems decrease slightly but at the same time they are able to maintain their performance over a wider range of misalignments during the blind testing, $\Delta_{z,test}$. (D) Standard D²NN systems trained for all-optical handwritten digit recognition with and without vaccination were tested against various levels of *lateral* misalignment levels, determined by $\Delta_{test}$. (E) Same as D except for differential D²NNs architectures. (F) Same as E and F, except for hybrid object recognition systems comprised of a jointly-trained 5-layer D²NN optical front-end and a single-layer fully-connected neural network at the electronic back-end, combined through 10 discrete opto-electronic detectors. The proposed vaccination-based training strategy improves the resilience of these diffractive networks to uncontrolled *lateral* and *axial* displacements of the diffractive layers with a modest compromise of the inference performance depending on the misalignment range used in the training phase.

inference accuracy for $\Delta_{test} = 4.24\lambda$ (i.e., the misalignment range in each lateral direction of a diffractive layer is $-8\delta$ to $8\delta$). On the other hand, when the diffractive layers are randomly misaligned on the longitudinal direction alone, the inference performance does not drop as excessively as the lateral misalignment case; for example, even when $\Delta_{z,test}$ becomes as large as 19.2 λ, the error-free diffractive network manages to obtain an inference accuracy of 49.8%.

As demonstrated in Figure 2D, the rapid drop in the testing accuracy of diffractive optical classification systems under physical misalignments can be mitigated by using the v-D²NN framework. Since v-D²NN training introduces displacement errors in the training stage, the diffractive optical networks can adopt to those variations preserving their inference performance over large misalignment margins. As an example, the 38.40% blind testing accuracy achieved by the non-vaccinated diffractive design with a lateral misalignment range of $\Delta_{test} = 2.12$ λ, can be

increased to 94.44% when the same architecture is trained with a similar error range using the presented vaccination framework (see the purple line in Figure 2D). On top of that, the vaccinated design does not compromise the performance of the all-optical object recognition systems when the ideal conditions are satisfied. Compared to the 97.77% accuracy provided by the error-free design, this new vaccinated network (purple line in Figure 2D) obtains 96.1% in the absence of misalignments. In other words, the ~56% inference performance gain of the vaccinated diffractive network under physical misalignments comes at the expense of only 1.67% accuracy loss when the opto-mechanical assembly perfectly matches the numerical training model. In case the level of misalignment-related imperfections in the fabrication of the diffractive network is expected to be even smaller, one can design improved v-D²NN models that achieve, e.g., 97.38%, which corresponds to only 0.39% inference accuracy loss compared to the error-free models at their peak (perfect alignment case)

while at the same time providing >4% blind testing accuracy improvement under mild misalignment, i.e., $\Delta_{test}$ = 0.53 λ. Similarly, when we compare the blind inference curves of the error-free and vaccinated network designs in Figure 2A, one can notice that the v-D²NN framework can easily recover the performance of the diffractive digit classification networks in the case where the displacement errors are restricted to be on the longitudinal axis. For example, with $\Delta_{z,test}$ = 2.4 λ, the inference accuracy of the

error-free diffractive network (dark blue) is reduced to 94.88%, while a vaccinated diffractive network that was already trained against the same level of misalignment, $\Delta_{z,tr}$ = 2.4 λ (yellow), retains 97.39% blind inference accuracy under the same level of axial misalignment.

Next, we combined our v-D²NN framework with the *differential* diffractive network architecture: the blind testing results of various differential handwritten digit recognition systems under axial and lateral misalignments
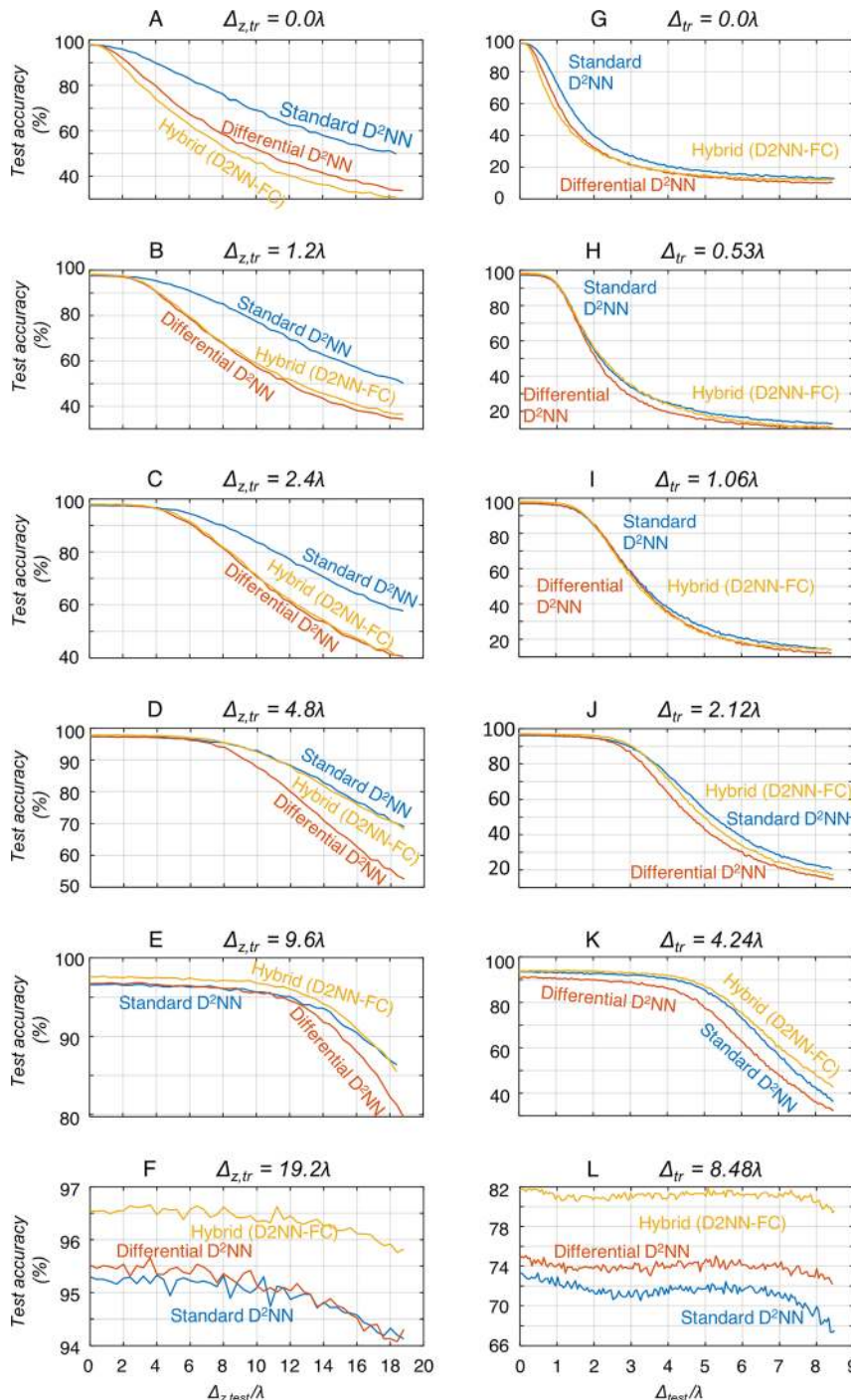


**Figure 3:** Comparison of different types of D²NN-based object classification systems trained with the same range of misalignments. (A) Comparison of error-free designs, $\Delta_{z,tr}$ = 0.0 λ, for standard (blue), differential (red) and hybrid (yellow) object classification systems against different levels of *axial* misalignments, $\Delta_{z,test}$. (B) Comparison of standard (blue), differential (red) and hybrid (yellow) object classification systems against different levels of *axial* misalignments when they are trained with $\Delta_{z,tr}$ = 1.2 λ. C, D, E and F are same as B, except during the training of the diffractive models the axial misalignment ranges are determined by $\Delta_{z,tr}$, taken as 2.4, 4.8, 9.6 and 19.2 λ, respectively. (G) Comparison of error-free designs, $\Delta_{tr}$ = 0.0 λ, for standard (blue), differential (red) and hybrid (yellow) object recognition systems against different levels of *lateral* misalignments, $\Delta_{test}$. (H) Comparison of standard (blue), differential (red) and hybrid (yellow) object classification systems against different levels of *lateral* misalignments when they are trained with $\Delta_{tr}$ = 0.53 λ. I, J, K and L are same as H, except the *lateral* misalignment ranges during the training are determined by $\Delta_{tr}$, taken as 1.06, 2.12, 4.24 and 8.48 λ, respectively.
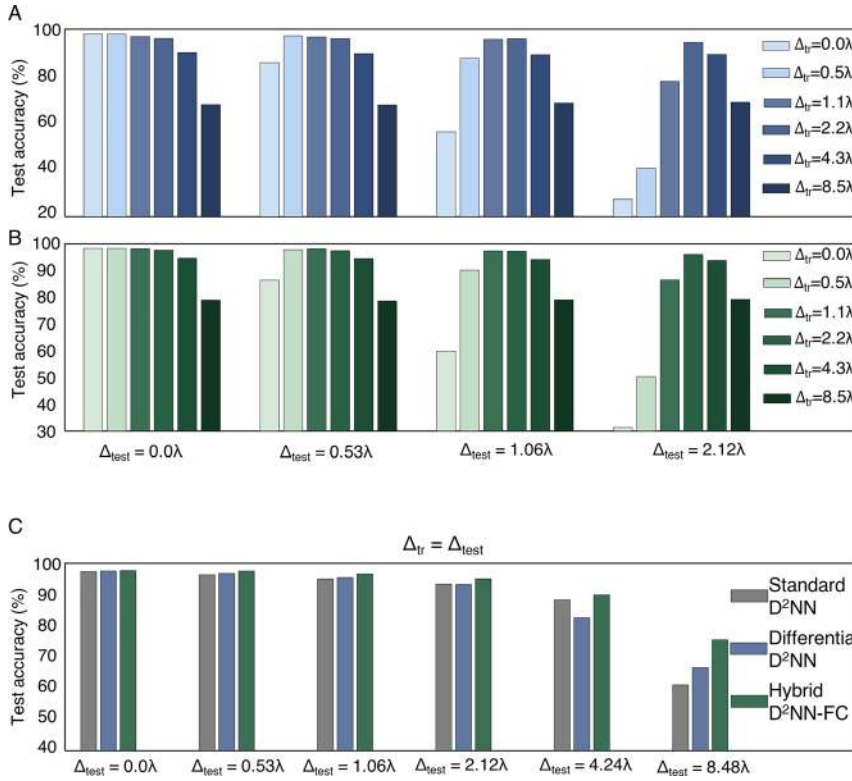
**Figure 4:** Summary of the numerical results for vaccinated D²NNs. (A) The inference accuracy of the non-vaccinated ($\Delta_{tr}$ = 0.0 λ) and the vaccinated ($\Delta_{tr}$ > 0.0 λ) differential D²NN systems trained for all-optical hand-written digit recognition quantified at different levels of testing misalignment ranges. The v-D²NN framework allows the all-optical classification systems to pre-serve their inference performance over a large range of misalignments. (B) Same as A, except for hybrid (D²NN-FC) systems comprised of a jointly-trained 5-layer D²NN optical front-end and a single-layer fully-connected neural network at the electronic back-end combined through 10 discrete opto-electronic detectors (see Figure 1C). (C) Vaccination comparison of three diffractive network-based machine learning architectures depicted in Figure 1; $\Delta_{tr} = \Delta_{test}$.

are reported in Figure 2B and E, respectively. Figure 3 also provides a direct comparison of the blind inference accuracies of these two all-optical diffractive machine learning architectures under different levels of misalignments. Figure 3A and G compare the error-free designs of differential and standard diffractive network architectures, which reveal that although the differential design achieves slightly better blind inference accuracy, 97.93%, in the absence of alignment errors, as soon as the misalignments reach beyond a certain level, the performance of a differential design decreases faster than the standard diffractive network. This means that they are more vulnerable against the system variations that they were not trained against. Since the number of detectors inside an output region-of-interest is twice as many in differential diffractive networks compared to the standard diffractive network architecture (see Figure 1A, B), the detector signals are more prone to have cross-talk when the diffractive layers are experiencing uncontrolled mechanical displacements. With the introduction of vaccination during the training phase, however, differential diffractive network models can adapt to these system variations as in the case of standard diffractive optical networks. Compared to standard diffractive optical networks, the differential counterparts that are vaccinated generate higher inference accuracies when the misalignment levels are small. In Figure 3H, for instance,

the vaccinated differential design (red curve) achieves 97.3% blind inference accuracy while the vaccinated standard diffractive network (blue curve) can provide 96.91% for the case $\Delta_{tr} = \Delta_{test} = 0.53\lambda$. In Figure 3I, where the vaccination range on *x* and *y* axis is twice as large compared to Figure 3H, the differential network reveals the correct digit classes with an accuracy of 96.18% when it is tested at an equal displacement/misalignment uncertainty to its vaccination level; on the other hand, the standard diffractive network can achieve 95.79% under the same training and testing conditions. Beyond this level of misalignment, the differential systems slowly lose their performance advantage and the standard diffractive networks starts to perform on par with their differential counterparts. One exception to this behavior is shown in Figure 3K, where the misalignment range of the diffractive layers during the training causes cross-talk among the differential detectors at a level that hurts the evolution of the differential diffractive network, leading to a consistently worse inference performance compared to the standard diffractive design. A similar effect also exists for the case illustrated in Figure 3L; however, this time, the standard diffractive optical network design also experiences a similar level of cross-talk among the class detectors at the output plane. Therefore, as demonstrated in Figure 3L, the differential diffractive optical network

recovers its performance gain thriving over the standard diffractive network design with a higher optical classification accuracy. This performance gain of the differential design depicted in Figure 3L, can be translated to the smaller misalignment cases, e.g., $\Delta_{tr} = \Delta_{test} = 4.24\lambda$, simply by increasing the distance between the detectors at the output plane for differential diffractive optical network designs, i.e., setting the region-of-interest covering the detectors to be larger compared to the standard diffractive network architecture.

Figure 3 also outlines a comparison of the differential and standard diffractive all-optical object recognition systems against hybrid diffractive neural networks under various levels of misalignments. For the hybrid neural network models presented here, we jointly trained a 5-layer diffractive optical front-end and a single-layer fully-connected electronic network, communicating through discrete detectors at the output plane. To provide a fair comparison with the all-optical diffractive systems, we used 10 discrete detectors at the output plane of these hybrid configurations, same as in the standard diffractive optical network designs (see Figures 1A and C). The blind inference accuracies obtained by these hybrid neural network systems under different levels of misalignments are shown in Figure 2C and F. When the opto-mechanical assembly of the diffractive network is perfect, the error-free, jointly-optimized hybrid neural network architecture can achieve 98.3% classification accuracy surpassing the all-optical counterparts as well as the all-electronic performance of a single-layer fully-connected network, which achieves 92.48% classification accuracy using >75-fold more trainable connections without the diffractive optical network front-end. As the level of misalignments increases, however, the error-free hybrid network fails to maintain its performance and its inference accuracy quickly falls. The v-D²NN framework helps the hybrid neural systems during the joint evolution of the diffractive and the electronic networks and makes them resilient to misalignments. For example, the handwritten digit classification accuracy values presented for the standard diffractive networks in Figure 3H (96.91%) and Figure 3I (95.79%) have improved to 97.92 and 97.15%, respectively, for the hybrid neural network system (yellow curve), indicating ~1% accuracy gain over the all-optical models under the same level of misalignment (i.e., 0.53 λ for Figure 3H and 1.06 λ for Figure 3I). As the level of misalignments in the diffractive optical front-end increases, the cross-talk between the detectors at the output plane also increases. However, for a hybrid network design there is no *direct* correspondence between the data classes and the output detectors, and therefore the joint-training under the vaccination scheme

introduced in this work directs the evolution of the electronic network model accordingly and opens up the performance gap further between the all-optical diffractive classification networks and the hybrid systems as illustrated in Figure 3K and L. A similar comparative analysis, along the lines of Figures 2 and 3, is also conducted for phase-encoded input objects (Fashion-MNIST dataset), which is reported in Supplementary Figures S4 and S5.

## 2.3 Experimental results

The error-free standard diffractive network design that achieves 97.77% blind inference accuracy for the MNIST dataset as presented in Figures 2A, D, 3A and G, offers a power efficiency of ~0.07% on average over the blind testing samples (see Supplementary Information for details). This relatively low power efficiency is mostly due to the absorption of our 3D printing material at THz band. Specifically, ~88.62% of the optical power right after the object is absorbed by the five diffractive layers, while 11.17% is scattered around during the light propagation. Due to the limited optical power in our THz source and the noise floor of our detector, we trained an error-free standard diffractive optical network model with a slightly compromised digit classification performance for the experimental verification of our v-D²NN framework. This new error-free diffractive network provides a blind inference accuracy of 97.19%, and it obtains ~3× higher power efficiency of ~0.2%. In addition to improved power efficiency, this new diffractive network model with 97.19% classification accuracy also achieves ~10× better signal contrast ($\psi$) [37] between the optical signal collected by the detector corresponding to the true object label and its closest competitor, i.e., the second maximum signal (see Supplementary Information for details). The layers of this error-free diffractive network are shown in Figure 1E. In addition, the comparison between the error-free, high-contrast standard diffractive optical network model and its lower contrast, lower efficiency counterpart in terms of their inference performance under misalignments is reported in Supplementary Figure S1A.

Following the same power-efficient design strategy, we trained another diffractive optical network that is *vaccinated* against *both the lateral and axial misalignments* with the training parameters ($\Delta_{tr}$, $\Delta_{z,tr}$) taken as (4.24 λ, 4.8 λ). As in the case of the error-free design, the inference accuracy of this new vaccinated diffractive network shown in Figure 5A is also compromised compared to the standard diffractive networks presented in Figures 2D and 3K since it was trained to improve power
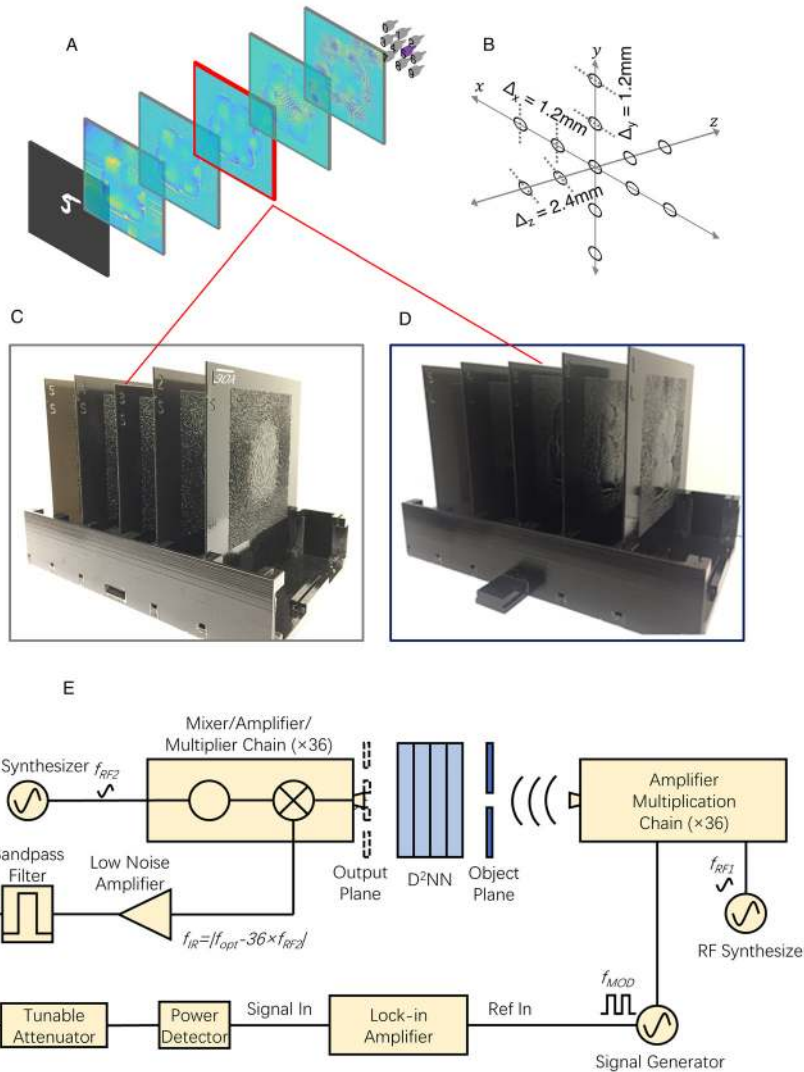
**Figure 5:** Experimental testing of v-D$^2$NN framework. (A) A diffractive optical network that is vaccinated against misalignments. This network is vaccinated against *both* lateral, $\Delta_{tr} = 4.24\,\lambda$, and axial, $\Delta_{z,tr} = 4.8\,\lambda$, misalignments. (B) The location of the third diffractive layer was on purpose altered throughout our measurements. Except the central location, the remaining 12 spots induce an inter-layer misalignment. (C) The 3D printed error-free design shown in Figure 1E. (D) The 3D printed vaccinated design shown in A and Figure 1D. (E) The schematic of the experimental setup.

efficiency and signal contrast. This design can achieve 89.5% blind classification accuracy for handwritten digits under ideal conditions, with the diffractive layers reported in Figure 1D. A comprehensive comparison of the blind inference accuracies of the vaccinated diffractive networks shown in Figures 2 and 3 and their high-contrast, high-efficiency counterparts are reported in Supplementary Figure S1B.

The experimental verification of our v-D2NN framework was based on the comparison of the vaccinated and the error-free standard diffractive optical network designs in terms of the accuracy of their optical classification decisions under inter-layer misalignments. To this end, we fabricated the diffractive layers of the non-vaccinated and the vaccinated networks shown in Figure 1D–E using 3D printing. The fabricated diffractive networks are depicted in Figure 5C–D. In addition, we fabricated six MNIST digits

selected from the blind testing dataset that are numerically correctly classified by both the vaccinated and the non-vaccinated diffractive network models without any misalignments. For a fair comparison, we grouped the correctly classified handwritten digits based on the signal contrast statistics provided by the non-vaccinated design. With $\mu_{SC}$, $\sigma_{SC}$, denoting the mean and the standard deviation of the signal contrast generated by the error-free diffractive network over the correctly classified blind testing MNIST digits, we selected two handwritten digits (Set 1) that satisfies the condition $\mu_{SC}+\sigma_{SC} < \{\psi, \psi'\} < \mu_{SC}+2\sigma_{SC}$, where $\psi$ and $\psi'$ denote the signal contrasts created by the error-free and the vaccinated designs for a given input object, respectively. The condition on $\psi$ and $\psi'$ for the second set of 3D printed handwritten digits (Set 2), on the other hand, is slightly less restrictive, $\mu_{SC} < \{\psi, \psi'\} < \mu_{SC}+\sigma_{SC}$. By using this outlined approach, we selected
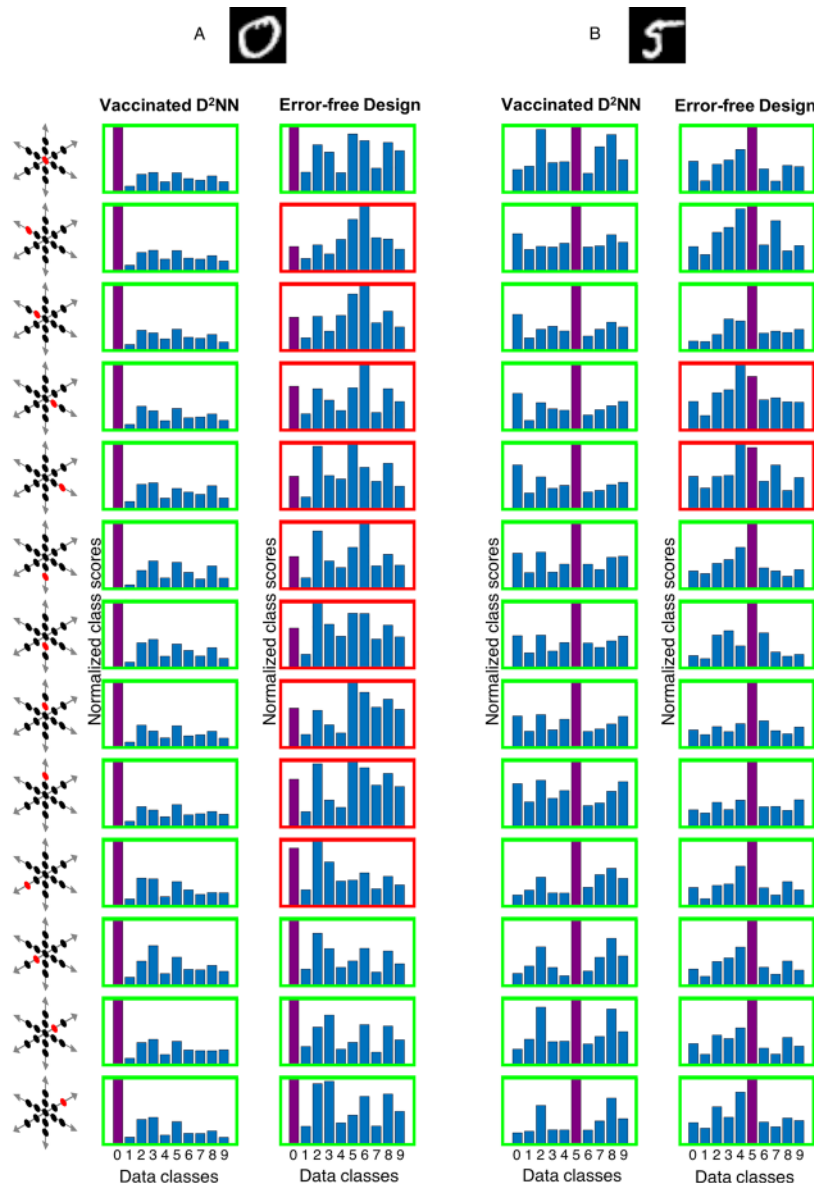
**Figure 6:** Experimental image classification results as a function of misalignments. (A) The experimentally measured class scores for handwritten digit '0' selected from Set 1. (B) Same as A, except the input object is now a handwritten digit '5' selected from Set 2. The red dot within the coordinate system shown on the left-hand side represents the physical misalignment for each case (see Figure 5B). Red (green) rectangles mean incorrect (correct) inference results. Refer to Supplementary Information (Figures S2 and S3) for more examples of our experimental comparisons between these vaccinated and error-free diffractive designs.

six experimental test objects in total that are equally favorable for both the vaccinated and non-vaccinated diffractive networks.

To test the performance of the error-free and vaccinated diffractive network designs under different levels of misalignments, we shifted the third layer of both diffractive systems to 12 different locations around its ideal location as depicted in Figure 5B. The perturbed locations of the third diffractive layer covers four different spots on each orthogonal direction. The distances between these locations are 1.2 mm (1.6 λ) along $x$ and $y$, and 2.4 mm (3.2 λ) along $z$ axes. These shifts cover a total length of 6.4 λ (12 times the smallest feature size) along ($x,y$) and 12.8 λ (0.32 × 40 λ) along $z$ axis, respectively.

Figure 5E shows a schematic of our THz setup that was used to test these diffractive networks and their misalignment performances (see Supplementary Information). Figure 6 reports the experimentally obtained optical signals for a handwritten digit '0' from Set 1 and a handwritten digit '5' from Set 2, received by the class detectors at the output plane based on the 13 different locations of the third diffractive layer of the vaccinated and the error-free networks. The first thing to note is that both the vaccinated and non-vaccinated networks can classify the two digits correctly when the third layer is placed at its ideal location within the set-up. As illustrated in Figure 6A, as we perturb the location of the third layer, the error-free diffractive network fails at nine locations while the vaccinated

network correctly infers the object label at all the 13 locations for the handwritten digit '0'. In addition, the vaccinated network maintains its perfect record of experimental inference for the digit '5' despite the inter-layer misalignments as depicted in Figure 6B. The error-free design, on the other hand, fails at two different locations of its third layer misalignment (see Figure 6B). The experimental results for the remaining four digits are presented in Supplementary Figures S2 and S3, confirming the same conclusions. In our experiments, all the objects were correctly classified when the third layer was placed at its ideal location. Out of the remaining 72 measurements (6 objects × 12 shifted/misaligned locations of the third layer), the error-free design failed to infer the correct object class in 23 cases, while the vaccinated network failed only two times, demonstrating its robustness against a wide range of misalignments as intended by the v-D$^2$NN framework.

# 3 Discussion

As an example of a severe case of lateral misalignments, we investigated a scenario where each diffractive layer can move within the range ($-8.48\,\lambda$, $8.48\,\lambda$) around the optical axis in $x$ and $y$ directions. As demonstrated in Figures 2D and 3G, when the error-free design (dark blue) is exposed to such large lateral misalignments, it can only achieve 12.8% test accuracy, i.e., it barely surpasses random guessing of the object classes. A diffractive optical network that is vaccinated against the same level of uncontrolled layer movement can partially recover the inference performance providing 67.53% blind inference accuracy. As the best performer, the hybrid neural network system composed of a 5-layer diffractive optical network and a single-layer fully-connected network can take this accuracy value up to 79.6% under the same level of misalignments, within the range ($-8.48\,\lambda$, $8.48\,\lambda$) for both $x$ and $y$ direction of each layer. When we compare the total allowed displacement range of each layer within the diffractive network (i.e., $16.96\,\lambda$ in each direction) and the size of our diffractive layers ($106.66\,\lambda$), we can see that they are quite comparable. If we imagine a lens-based optical imaging system and an associated machine vision architecture, in the presence of such serious opto-mechanical misalignments, this system would also fail due to acute aberrations substantially decreasing the image quality and the resolution. Our main motivation to include this severe misalignment case in our analyses was to test the limits of the adaptability of our vaccinated systems.

Figure 4A–B further summarize the inference accuracies of the differential diffractive networks and hybrid neural network systems at discrete points sampled from the corresponding curves depicted in Figure 3G–L. In Figure 4A, the best inference accuracy is achieved by the error-free (non-vaccinated) differential diffractive network model under perfect alignment of its layers. However, its performance drops in the presence of an imperfect opto-mechanical assembly. The vaccinated, diffractive all-optical classification networks provide major advantages to cope with the undesired system variations achieving higher inference accuracies despite misalignments. The joint-training of hybrid systems that are composed of a diffractive optical front-end and a single-layer electronic network (back-end) can adapt to uncontrolled mechanical perturbations achieving higher inference accuracies compared to all-optical image classification systems. These results further highlight that, operating with only a few discrete opto-electronic detectors at the output plane, the D$^2$NN-based hybrid architectures offer unique opportunities for the design of low-latency, power-efficient and memory-friendly machine vision systems for various applications.

On top of the translational layer-to-layer alignment errors, the presented framework can also be extended to accommodate 3D rotational misalignments of diffractive layers. While undesired in-plane rotations of diffractive layers can be readily addressed based on the 2D coordinate transformations performed through unitary rotation matrices incorporated into the optical forward model detailed in Supplementary Information, handling possible out-of-plane rotations of diffractive optical network layers requires further modifications to the formulation of wave propagation between tilted planes [48, 49]. Beyond misalignments or displacements of diffractive layers, the presented vaccination framework can also be harnessed to decrease the sensitivity of diffractive optical networks to various error sources, e.g., detection noise or fabrication defects. At its core, the presented framework can be interpreted as a training regularization method that avoids overfitting of a machine learning hardware to the specific 3D physical structure, distances and operational conditions, which are often assumed to be deterministic, precise and ideal during the training phase. In this respect, beyond its application to practically improve diffractive optical neural networks, the core principles introduced in our work can be extended to train other machine learning platforms [35, 50, 51] to mitigate various physical error sources that can cause deviations between the designed inference models and their corresponding physical implementations.

In conclusion, we presented a design framework that introduces the use of probabilistic layer-to-layer

misalignments during the training of diffractive neural networks to increase their robustness against physical misalignments. Although the experimental demonstrations of our vaccinated design framework were conducted using THz wavelengths and 3D printed diffractive layers, the presented principles and methods can readily be applicable to diffractive optical networks that operate at other parts of the electromagnetic spectrum, including, e.g., visible wavelengths. In fact, as the wavelength of operation gets smaller, the impact and importance of the presented framework will be better highlighted. We believe the presented training strategy will find use in the design of diffractive optical network-based machine vision and sensing systems, spanning different applications.

# References

[1] Y. Rivenson, H. Wang, Z. Wei, et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nat. Biomed. Eng.*, vol. 3, no. 6, pp. 466–477, 2019.

[2] Y. Zhang, K. de Haan, Y. Rivenson, J. Li, A. Delis, and A. Ozcan, "Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue," *Light Sci. Appl.*, vol. 9, no. 1, 2020. http://dx.doi.org/10.1038/s41377-020-0315-y.

[3] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.

[4] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light Sci. Appl.*, vol. 8, no. 1, p. 23, 2019.

[5] D. Amodei, S. Ananthanarayanan, R. Anubhai, and J. Bai, "Deep speech 2: end-to-end speech recognition in English and Mandarin," in *Proceedings of the 33rd International Conference on Machine Learning*, New York, NY, USA, JMLR: W&CP, vol. 48, p. 10, 2016.

[6] M. D. Zeiler, M. Ranzato, R. Monga, et al., "On rectified linear units for speech processing," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 3517–3521, https://doi.org/10.1109/ICASSP.2013.6638312.

[7] Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," *Light Sci. Appl.*, vol. 8, no. 1, pp. 1–8, 2019.

[8] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.*, vol. 7, no. 2, pp. 17141, 2018.

[9] Y. Wu, Y. Rivenson, Y. Zhang, et al., "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," *Optica*, vol. 5, no. 6, p. 704, 2018.

[10] Y. Wu, Y. Luo, G. Chaudhari, et al., "Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram," *Light Sci. Appl.*, vol. 8, no. 1, 2019. https://doi.org/10.1038/s41377-019-0139-9.

[11] Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," *Optica*, vol. 5, no. 10, pp. 1181–1190, 2018.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.

[14] R. Collobert and J. Weston, "A unified architecture for natural language processing: deep neural networks with multitask learning," in *Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5–9*, 2008, p. 8. https://doi.org/10.1145/1390156.1390177.

[15] B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," *Light Sci. Appl.*, vol. 7, no. 1, pp. 1–11, 2018.

[16] S. Bo, F. Schmidt, R. Eichhorn, and G. Volpe, "Measurement of anomalous diffusion using recurrent neural networks," *Phys. Rev. E*, vol. 100, no. 1, 2019. https://doi.org/10.1103/PhysRevE.100.010102.

[17] F. Cichos, K. Gustavsson, B. Mehlig, and G. Volpe, "Machine learning for active matter," *Nat. Mach. Intell.*, vol. 2, no. 2, pp. 94–103, 2020.

[18] K. de Haan, Y. Rivenson, Y. Wu, and A. Ozcan, "Deep-learning-based image reconstruction and enhancement in optical microscopy," *Proc. IEEE*, vol. 108, no. 1, pp. 30–50, 2020.

[19] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458, 2018.

[20] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, p. 1117, 2017.

[21] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica*, vol. 4, no. 11, p. 1437, 2017.

[22] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nat. Biotechnol.*, vol. 36, no. 5, pp. 460–468, 2018.

[23] D. Liu, Y. Tan, E. Khoram, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photonics*, vol. 5, no. 4, pp. 1365–1369, 2018.

[24] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, and H. Suchowski, "Plasmonic nanostructure design and characterization via deep learning," *Light Sci. Appl.*, vol. 7, no. 1, p. 60, 2018.

[25] W. Ma, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials," *ACS Nano*, vol. 12, no. 6, pp. 6326–6334, 2018.

[26] Shen Y., Harris, N., Skirlo, S., et al., "Deep learning with coherent nanophotonic circuits," *Nat. Photon*, vol. 11, no. 7, pp. 441–446, 2017.

[27] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, 2019.

[28] A.N. Tait, T.F. de Lima, E. Zhou, et al., "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Rep.*, vol. 7, no. 1, 2017, https://doi.org/10.1038/s41598-017-07754-z.

[29] J. George, R. Amin, A. Mehrabian, et al., "Electrooptic nonlinear activation functions for vector matrix multiplications in optical neural networks," in *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF), Zurich*, 2018, p. SpW4G.3, https://doi.org/10.1364/SPPCOM.2018. SpW4G.3.

[30] A. Mehrabian, Y. Al-Kabani, V. J. Sorger, and T. El-Ghazawi, "PCNNA: a photonic convolutional neural network accelerator," in *2018 31st IEEE International System-on-Chip Conference (SOCC)*, 2018, pp. 169–173, https://doi.org/10.1109/SOCC. 2018.8618542.

[31] M. Miscuglio, A. Mehrabian, Z. Hu, et al., "All-optical nonlinear activation function for photonic neural networks [Invited]," *Opt. Mater. Exp.*, vol. 8, no. 12, p. 3851, 2018.

[32] D. Psaltis, D. Brady, X.-G. Gu, and S. Lin, "Holography in artificial neural networks," *Nature*, vol. 343, no. 6256, pp. 325–330, 1990.

[33] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, H.-T. Peng, and P. R. Prucnal, "Neuromorphic photonics, principles of," in *Encyclopedia of Complexity and Systems Science*, R. A. Meyers, Ed. Berlin, Heidelberg, Springer, 2018, pp. 1–37. https://doi.org/10.1007/978-3-642-27737-5_702-1.

[34] N. M. Estakhri, B. Edwards, and N. Engheta, "Inverse-designed metastructures that solve equations," *Science*, vol. 363, no. 6433, pp. 1333–1338, 2019.

[35] J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Sci. Rep.*, vol. 8, no. 1, p. 12324, 2018.

[36] X. Lin, Y. Rivenson, N. T. Yardimci, et al., "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018.

[37] D. Mengu, Y. Luo, Y. Rivenson, and A. Ozcan, "Analysis of diffractive optical neural networks and their integration with electronic neural networks," *IEEE J. Select. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–14, 2020.

[38] J. Li, D. Mengu, Y. Luo, Y. Rivenson, and A. Ozcan, "Class-specific differential detection in diffractive optical neural networks improves inference accuracy," *Adv. Photonics*, vol. 1, no. 4, p. 046001, 2019.

[39] Y. Luo, D. Mengu, N. T. Yardimci, et al., "Design of task-specific optical systems using broadband diffractive neural networks," *Light Sci. Appl.*, vol. 8, no. 1, p. 112, 2019.

[40] G. V. der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, no. 3, pp. 561–576, 2017.

[41] A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto, "Network of time-multiplexed optical parametric oscillators as a coherent Ising machine," *Nat. Photonics*, vol. 8, no. 12, pp. 937–942, 2014.

[42] B. Penkovsky, X. Porte, M. Jacquot, L. Larger, and D. Brunner, "Coupled nonlinear delay systems as deep convolutional neural networks," *Phys. Rev. Lett.*, vol. 123, no. 5, 2019, https://doi.org/10.1103/PhysRevLett.123.054101.

[43] J. Bueno, S. Maktoobi, L. Froehly, et al., "Reinforcement learning in a large-scale photonic recurrent neural network," *Optica*, vol. 5, no. 6, p. 756, 2018.

[44] J. K. George, A. Mehrabian, R. Amin, et al., "Neuromorphic photonics with electro-absorption modulators," *Opt. Express*, vol. 27, no. 4, p. 5181, 2019.

[45] A. Ribeiro, A. Ruocco, L. Vanacker, and W. Bogaerts, "Demonstration of a $4 \times 4$-port universal linear circuit," *Optica*, vol. 3, no. 12, p. 1348, 2016.

[46] Y. LeCun, L. Bottou, Y. Bengio, and P. Ha, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[47] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," arXiv: 1708.07747 [cs, stat], Sep.2017. [Online]. Available at: http://arxiv.org/abs/1708.07747 [accessed: May 13, 2020].

[48] N. Delen and B. Hooker, "Free-space beam propagation between arbitrarily oriented planes based on full diffraction theory: a fast Fourier transform approach," *J. Opt. Soc. Am. A*, vol. 15, no. 4, p. 857, 1998.

[49] K. Matsushima, H. Schimmel, and F. Wyrowski, "Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves," *J. Opt. Soc. Am. A*, vol. 20, no. 9, p. 1755, 2003.

[50] G. Marcucci, D. Pierangeli, and C. Conti, "Theory of neuromorphic computing by waves: machine learning by rogue waves, dispersive shocks, and solitons," arXiv:1912.07044 [cond-mat, physics:nlin, physics:physics], Dec.2019. [Online]. Available at: http://arxiv.org/abs/1912.07044 [accessed: Jun. 12, 2020].

[51] T. W. Hughes, I. A. D. Williamson, M. Minkov, and S. Fan, "Wave physics as an analog recurrent neural network," *Sci. Adv.*, vol. 5, no. 12, p. eaay6946, 2019.