photonics

MDPI

Article

Generative Adversarial Optical Networks Using Diffractive Layers for Digit and Action Generation

Pei Hu [1,*], Tengyu Cui [2], Yuanyuan Zhang [3] and Shuai Feng [3,*]

[1] College of Humanities and Law, Beijing University of Chemical Technology, Beijing 100029, China
[2] Department of Physical Education, Beijing University of Posts and Telecommunications, Beijing 100876, China; cuitengyu1@bupt.edu.cn
[3] School of Science, Minzu University of China, Beijing 100081, China; 23302511@muc.edu.cn
* Correspondence: hupei01@buct.edu.cn (P.H.); shuaifeng@muc.edu.cn (S.F.)

Abstract

Within the traditional electronic neural network framework, Generative Adversarial Networks (GANs) have achieved extensive applications across multiple domains, including image synthesis, style transfer and data augmentation. Recently, several studies have explored the use of optical neural networks represented by the diffractive deep neural network (D^2NN) for GANs. However, most of these focus on applications of the generative network, and there is currently no well-established D^2NN architecture that simultaneously implements generative adversarial functionality. Here, we propose a novel implementation scheme for generative adversarial networks based on all-optical diffraction layers, demonstrating a complete all-optical adversarial architecture that simultaneously realizes both the generative network and the adversarial network (D^2NN-GAN). We validated this method on the MNIST handwritten digit dataset, achieving Nash equilibrium convergence with the discriminator accuracy stabilizing around 50%. Concurrently, the average SSIM parameter of generated images reached 0.9573, indicating that the generated samples possess high quality and closely resemble real samples. Furthermore, we extended the framework to the KTH human action dataset, successfully reconstructing the "running" action with a discriminator accuracy of approximately 75%. The D^2NN-GAN architecture introduces a fully optical generative adversarial model, providing a practical path for future optical modeling methods, such as image generation and video synthesis.

Keywords: diffractive deep neural network; optical neural network; generative adversarial networks; optical classification

1. Introduction

The exponential growth of deep neural networks has revolutionized artificial intelligence, enabling unprecedented capabilities in image recognition, natural language processing and generative modeling [1–3]. However, this remarkable progress has come at a substantial computational cost, with state-of-the-art models requiring massive computational resources that strain the limits of conventional electronic processors [4,5]. The energy consumption of training large-scale neural networks has become a critical concern, with some models consuming as much electricity as small cities during their training phase [6,7]. This computational bottleneck has motivated the exploration of alternative computing paradigms that can overcome the fundamental limitations of electronic architecture.

Optical neural networks (ONNs) have emerged as a promising solution to address these challenges, leveraging the inherent parallelism and energy efficiency of photonic

Check for updates

Received: 12 December 2025
Revised: 13 January 2026
Accepted: 16 January 2026
Published: 21 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

information processing [8–14]. Light propagation through optical media naturally performs matrix multiplications at the speed of light with minimal energy dissipation, making it ideally suited for neural network computations [15,16]. Recent breakthroughs in integrated photonics have demonstrated optical implementations of various neural network architectures, including convolutional neural networks [17], recurrent neural networks [18] and reservoir computing systems [19].

Among various optical neural network implementation schemes, Diffractive Neural Networks (D^2NNs) have garnered significant attention due to their advantages of large-scale computational matrices, low loss and high parallelism [20–25]. D^2NNs consist of successive diffractive layers that modulate the phase and/or amplitude of incident light, with each layer containing trainable neurons arranged in a two-dimensional array. The network parameters are encoded in the transmission coefficients of these layers, which can be physically realized using various technologies, including 3D-printed diffractive elements [26], spatial light modulators [27] and Reconfigurable/Plug-and-Play Meta surfaces [28]. D^2NNs offer several compelling advantages: they perform inference at the speed of light, consume no power except for the input illumination and can process entire images in parallel without scanning or serialization [29,30].

Generative adversarial networks (GANs), introduced by Goodfellow et al. [31], represent one of the most significant innovations in machine learning over the past decade. GANs employ two neural networks in an adversarial game: a generator that creates synthetic data and a discriminator that distinguishes real from fake samples [32]. This adversarial training process has proven remarkably effective for generating high-quality synthetic data, with applications ranging from photorealistic image synthesis [33–35] to drug discovery [36] and material design [37]. The success of GANs stems from their ability to capture complex data distributions without explicit likelihood estimation, learning through a minimax game that pushes both networks toward improved performance [38,39].

Recently, a groundbreaking study demonstrated optical generation of Van Gogh-style paintings using diffractive deep neural networks (D^2NN) [40]. This work reported" optical generative models" by distilling a digital diffusion model into an optical generation pipeline, experimentally demonstrating snapshot generation including Van Gogh-style artworks under monochrome and multiwavelength illumination. In parallel, Zhan et al. demonstrated photonic diffractive generators by sampling optical noises from scattering media [41]. Moreover, Qiu et al. proposed an optoelectronic GAN (OE-GAN) that distributes the generator and discriminator across optical and electronic components, using a diffractive-network-based optoelectronic generator with an electronic MLP discriminator [42]. These studies delved deeply into the potential applications of D^2NN and generative networks. However, to our knowledge, no research has yet successfully implemented a complete GAN architecture—including both optical generator and optical discriminator—using diffractive optical networks.

Here, we present a strategy for implementing generative adversarial networks employing two five-layer D^2NN networks—one as the generator and the other as the discriminator. By using the MNIST handwritten digit dataset, the generator network transforms spatially encoded Gaussian beams with varying positions and waist radii into synthetic digit patterns. The generated samples are mixed with real images from the original dataset and fed into the discriminator network. The output layer of the discriminator network is divided into two regions, classifying samples as real or fake based on the detected light intensity. In a generative adversarial network, the training cycle alternates between the generator network and the adversarial network. Our simulation results demonstrate that D^2NN-GAN successfully achieves Nash equilibrium: the discriminator's accuracy converges to approximately 50% (The generator's loss value cannot be reduced further, as

the images it produces are already sufficiently realistic. At this point, the discriminator can only perform blind tests, randomly determining whether an image is real or fake). Furthermore, the generator network achieves an average SSIM index of 0.9573 and an FID index of 106.31—indicating that the generated images approach the quality level of real samples.

Building upon this foundation, we further extended the framework to the more challenging KTH Human Action Dataset. Through generative adversarial training on the "running" action, D^{2}NN-GAN ultimately achieved a classification accuracy of 75%. The generated images attained an average SSIM index of 0.2658 and an FID index of 253.71. The failure to achieve Nash equilibrium with an accuracy rate of 50% is primarily attributed to the increased data complexity in the KTH dataset compared to MNIST (expanding from 28 × 28 pixel images to 120 × 160 pixel images). Therefore, we validated a diffractive network with 400 × 400 neurons after expanding the network scale. The average SSIM index of generated images improved to 0.4665, while the FID index decreased to 204.81, achieving a more effective D^{2}NN-GAN.

Diffractive neural network-based fully optical generative adversarial networks provide a practical fully optical generative adversarial approach. Our results demonstrate that this approach exhibits strong generative adversarial capabilities across both handwritten digit and human motion datasets. Leveraging multiple advantages of diffraction networks—such as ultra-fast generation and inference at minimal energy consumption, along with robust parallel processing capabilities that may support multi-sample parallel processing in the future [43]—D^{2}NN-GAN holds promise for effective application across diverse image generation scenarios.

2. Methods

2.1. Forward Propagation Model of D^{2}NN

Our D^{2}NN -GAN framework consists of two independent D^{2}NNs operating in cascade. Each network comprises five diffractive layers separated by free-space propagation distances of 40λ, with each layer containing 200 × 200 trainable neurons arranged in a square lattice with a pixel pitch of 0.5λ. The networks operate at a wavelength of λ = 1550 nm, chosen for its availability in coherent light sources and compatibility with standard optical components.

The forward propagation through each diffractive layer is modeled using the Rayleigh-Sommerfeld (RS) diffraction integral [44]:

$$U_{l+1}(x_{l+1}, y_{l+1}) = \iint U_l(x_l, y_l) \cdot t_l(x_l, y_l) \cdot h(x_{l+1} - x_l, y_{l+1} - y_l) dx_l dy_l, \quad (1)$$

where U_l represents the complex field at layer l, t_l denotes the complex transmission function of layer l, and h is the free-space transfer function given by:

$$h(x, y) = \frac{1}{i\lambda} \frac{exp(ikr)}{r} cos\theta, \quad (2)$$

with $r = \sqrt{x^2 + y^2 + z^2}$ and $cos\theta = z/r$. In the formula, $i^2 = -1$. The transmission function of each neuron is parameterized as $t = exp(i\phi)$, where $\phi \in [0, 2\pi]$ represents the trainable phase modulation. The RS is a rigorous scalar propagation model (not a paraxial Fresnel approximation) under the scalar-wave assumption. Moreover, the RS formulation is mathematically equivalent to the angular spectrum method, i.e., a real-space convolution with the free-space Green's function corresponds to a Fourier-domain transfer function $exp(ikr)$. Under our propagation distance (40λ) and discrete layer sampling, evanescent spectral components decay exponentially with z and therefore do not contribute

appreciably at the subsequent layer. These considerations justify the physical consistency of the RS-based forward model used throughout this study. In our work, backpropagation is performed numerically on a computer, and the trained phase masks are then used for optical inference/generation.

2.2. Gaussian Beam Parameters

The generator network transforms structured input light into synthetic digit patterns. We encode the input using Laguerre-Gaussian (LG) beams, which provide a rich basis for spatial mode representation [45]:

$$LG_{p,l}(r,\varphi,z) = \sqrt{\frac{2p!}{(\pi(p+|l|)!)}}\frac{1}{w(z)}\left(\frac{r\sqrt{2}}{w(z)}\right)^{|l|}L_p^{|l|}\left(\frac{2r^2}{w^2(z)}\right)exp\left(-\frac{r^2}{w^2(z)}\right)exp(il\varphi) \\ exp(ikz)exp(i(2p+|l|+1)\zeta(z)), \tag{3}$$

where p and l are the radial and azimuthal mode indices, w(z) is the beam waist, $L_p^{|l|}$ represents the associated Laguerre polynomial, and $\zeta(z) = arctan(z/zR)$ is the Gouy phase with Rayleigh range $z_R = \pi w_0^2/\lambda$.

For our implementation, we use fundamental Gaussian modes (p = l = 0) with varying beam waists ranging from 25λ to 35λ and lateral positions spanning a 3 × 3 grid across the input plane. This parameterization provides sufficient diversity in the input space while maintaining Simulation feasibility.

2.3. Discriminator Network Parameters

The discriminator network receives either real digits from the MNIST dataset or synthetic digits from the generator. The input images are encoded as amplitude distributions. The discriminator's output plane is divided into two square regions: a "True" region centered at (x_T, y_T) =(−23.25 μm, 0) and a "False" region at (x_F, y_F) = (23.25 μm, 0). The area diameter is 15.5 μm. The classification decision is based on the integrated intensity in each region:

$$I_{True} = \iint_{RT} |U_{out}(x,y)|^2 dxdy, \tag{4}$$

$$I_{False} = \iint_{RF} |U_{out}(x,y)|^2 dxdy, \tag{5}$$

where I represents the intensity, RT and RF denote the detected square regions. The category represented by the higher intensity region in the two square areas is the optical inference classification.

2.4. Network Training Parameters

We implement alternating gradient descent with the Adam optimizer [46], using learning rates of $\alpha_G = 0.001$ for the generator and $\alpha_D = 0.0005$ for the discriminator. The asymmetric learning rates help stabilize training by preventing the discriminator from overwhelming the generator. As stated in the main text, the discriminator network is solely responsible for recognizing and inferring the intensity distribution within ten designated rectangular regions, while the generator network must handle full-image generation. The differing task complexities dictate these hyperparameter distinctions. If both networks share the same learning rate, gradient training will reference both rates simultaneously. Since the discriminator network has already reached an optimal state, its loss value becomes difficult to further reduce. This prevents the generator network's loss from decreasing further, causing it to converge to a local optimum. All models in this work were constructed and trained based on PyTorch 2.1.2, implemented within a Python 3.9 computational framework. The simulations were executed on a 64-bit workstation

equipped with an Intel Core i9-10940k CPU, NVIDIA GeForce RTX 4080 Ti GPU, and 256 GB DDR4 system memory.

Each training batch consists of 8 real samples and 8 generated samples, with the discriminator updated twice for each generator update to maintain competitive balance. It is important to note that selecting a batch size within an appropriate range will not affect the overall generative and discriminative quality of the network. This is because the number of parameters in diffractive neural networks (approximately 200 × 200 × 5 = 0.2 million) is significantly smaller than that of convolutional neural networks like VGG-16 (approximately 138 million parameters). Thus, batchsize exerts minimal influence on the overall gradient training process.

2.5. FID Calculation Method

To quantitatively evaluate the quality and diversity of generated images, we employed the Fréchet Inception Distance (FID) as a comprehensive metric. FID measures the similarity between the distributions of real and generated images by comparing their feature representations in a high-dimensional embedding space. Unlike pixel-level metrics, FID captures both the perceptual quality and diversity of generated samples.

2.5.1. Feature Extraction with Inception-v3

The FID calculation relies on a pre-trained Inception-v3 network to extract high-level features from images. The Inception-v3 model, trained on ImageNet, serves as a feature extractor that maps images to a 2048-dimensional feature space. For our implementation, we modified the standard Inception-v3 architecture by removing the final classification layer and extracting features from the last average pooling layer.

Since our diffractive networks generate single-channel images (grayscale) with dimensions of 200 × 200 pixels for MNIST and 160 × 120 pixels for KTH dataset, we applied the following preprocessing steps:

1. Channel replication: Single-channel images were replicated across three channels to match the RGB input format expected by Inception-v3: IRGB(x, y) = [Igray(x, y), Igray(x, y), Igray(x, y)]

2. Spatial resizing: Images were resized to 299 × 299 pixels (the standard Inception-v3 input size) through zero-padding: Ipadded (x,y) = pad(IRGB (x,y), (hpad, wpad)), where hpad = $\lfloor (299 - \text{Horiginal})/2 \rfloor$ and wpad = $\lfloor (299 - \text{Woriginal})/2 \rfloor$

3. Intensity normalization: Pixel intensities were normalized to the range [0, 1]: Inorm(x,y) = (Ipadded(x,y) − min(Ipadded))/(max(Ipadded) − min(Ipadded) + ε), where ε = 10 − 8 ensures numerical stability.

The forward propagation through the modified Inception-v3 network can be expressed as: f = Inception-v3pool(Inorm), where f ∈ R2048 represents the feature vector for each image.

2.5.2. Statistical Computation

For a set of N images, we computed the activation statistics by passing all images through the feature extractor in batches to obtain the feature matrix F ∈ RN × 2048. The mean vector and covariance matrix of the feature distribution were calculated as:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} f_i, \tag{6}$$

$$\sum = \left(\frac{1}{N-1}\right) \sum_{i=1}^{N} (f_i - \mu)(f_i - \mu)^T, \tag{7}$$

where f_i denotes the feature vector of the i-th image, μ ∈ R2048 is the mean vector, and Σ ∈ R2048 × 2048 is the covariance matrix.

2.5.3. Fréchet Distance Calculation

The FID score quantifies the distance between two multivariate Gaussian distributions by computing the Fréchet distance (also known as the Wasserstein-2 distance) between them. Given the mean vectors (μ_{real}, μ_{fake}) and covariance matrices ($\sum_{real}$, $\sum_{fake}$) for real and generated image features, the FID is calculated as:

$$FID = \| \mu_{real} - \mu_{fake} \|^2 + Tr(\sum_{real} + \sum_{fake} - 2(\sum_{real} \cdot \sum_{fake})^{\frac{1}{2}}), \quad (8)$$

where $\|\cdot\|^2$ denotes the L2 norm, $Tr(\cdot)$ represents the matrix trace, and $\left(\sum_{real} \cdot \sum_{fake}\right)^{\frac{1}{2}}$ is the matrix square root of the product $\sum_{real} \cdot \sum_{fake}$.

The first term $\| \mu_{real} - \mu_{fake} \|^2$ measures the squared Euclidean distance between feature means, capturing the difference in average characteristics. The second term $Tr(\sum_{real} + \sum_{fake} - 2(\sum_{real} \cdot \sum_{fake})^{\frac{1}{2}})$ quantifies the difference in feature distributions, accounting for both variance and correlations within each distribution.

3. Results

3.1. GAN Principle Based on D²NN

Figure 1 illustrates the complete architecture of our diffractive GAN system. The generator network receives Gaussian beams with varying spatial parameters and transforms them through five diffractive layers to produce synthetic digit patterns. These generated samples are then mixed with real MNIST digits (or human movement digits) and fed to the discriminator network (bottom), which performs binary classification through spatially separated output regions. The adversarial feedback loop drives both networks toward improved performance, with the generator learning to produce increasingly realistic digits while the discriminator becomes more sophisticated at detecting fakes.

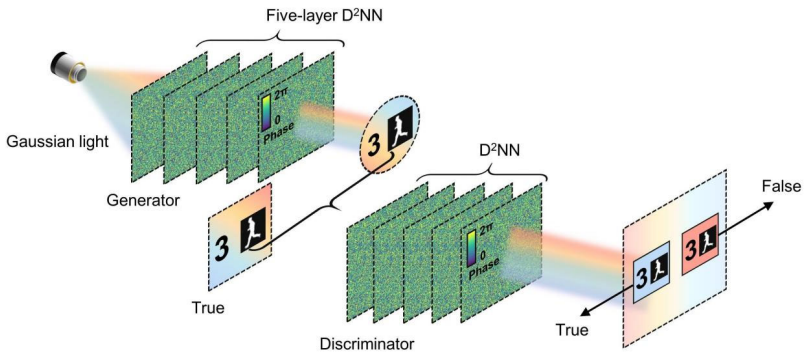


Figure 1. The operational principle follows the classical GAN framework adapted for optical implementation. In the generation phase, Laguerre–Gaussian beams with controlled spatial positions and waist radii illuminate the first diffractive layer of the generator network. As the optical field propagates through the five diffractive layers, each layer applies learned phase modulations that collectively transform the input beam into a synthetic digit pattern at the output plane. The generated optical patterns represent "fake" handwritten digits that aim to mimic the intensity distributions of real MNIST (or KTH) samples. The subsequent mixed numbers will be input together into the discriminator network for true and false classification.

3.2. D²NN-GAN for the MNIST Dataset

Figure 2 presents the learned phase distributions and optical field evolution through both networks after training convergence. Figure 2a,b shows the phase patterns of the five diffractive layers for the generator and discriminator networks, respectively. The Gaussian beam collectively transforms into complex digit morphologies. In contrast,

the discriminator's phase patterns (Figure 2b) display more uniform distributions with subtle variations, indicating a focus on feature extraction rather than pattern synthesis. This distinction aligns with the networks' complementary roles: the generator performs complex transformations while the discriminator implements analytical decomposition.

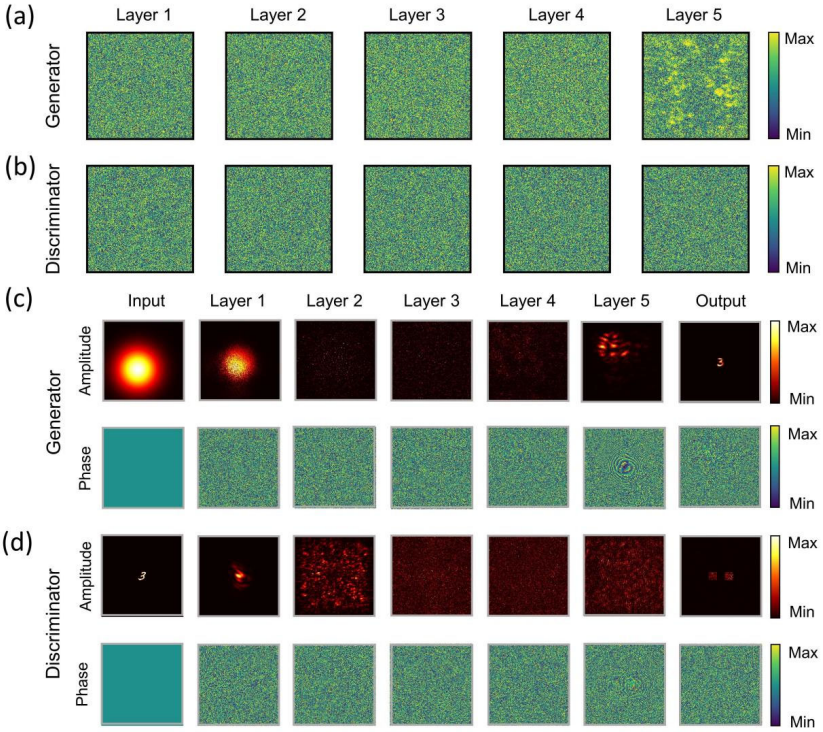


Figure 2. Phase distribution and optical field propagation in the D^2NN-GAN. (**a**) Phase layer parameters of the generator D^2NN across five diffractive layers. (**b**) Phase layer parameters of the discriminator D^2NN across five diffractive layers. (**c**) Amplitude and phase distributions of optical fields during propagation through the generator D^2NN, showing the evolution from input layer through each diffractive layer to the output layer. (**d**) Amplitude and phase distributions of optical fields during propagation through the discriminator D^2NN, displaying the field of digit "3" evolution from input layer through each diffractive layer to the output layer. The color scales represent normalized intensity values from minimum to maximum for both amplitude and phase distributions.

The optical field propagation through the generator network is visualized in Figure 2c, showing both amplitude and phase distributions at each layer for a representative input generating the digit "3". The input Gaussian beam undergoes progressive modulation, with early layers (1–2) performing coarse shaping and later layers (3–5) refining fine details. Layer 1 immediately disperses the Gaussian profile into multiple focal points, establishing the basic digit structure. Subsequent layers redistribute energy to form the characteristic curves of the numeral, with the final layer performing precise correction to achieve uniform brightness.

Figure 2d illustrates the discriminator's processing of a generated "fake" digit. The amplitude evolution shows progressive feature extraction, with early layers performing edge detection and spatial frequency analysis. The concentrated intensity patterns in layers 3–4 suggest the formation of characteristic features used for discrimination. The phase distributions maintain relative coherence throughout propagation, indicating that the discriminator preserves phase relationships for accurate classification. The final output layer shows clear intensity concentration in the "False" region, demonstrating successful discrimination.

The loss evolution curves in Figure 3a reveal characteristic GAN training behavior, with both networks exhibiting rapid initial improvement followed by stable oscillation around equilibrium values. The generator loss stabilizes at nearly 1.7 after approximately 100 epochs, while the discriminator loss converges to 0.69. These near-identical loss values indicate balanced competition between the networks, neither able to consistently outperform the other.

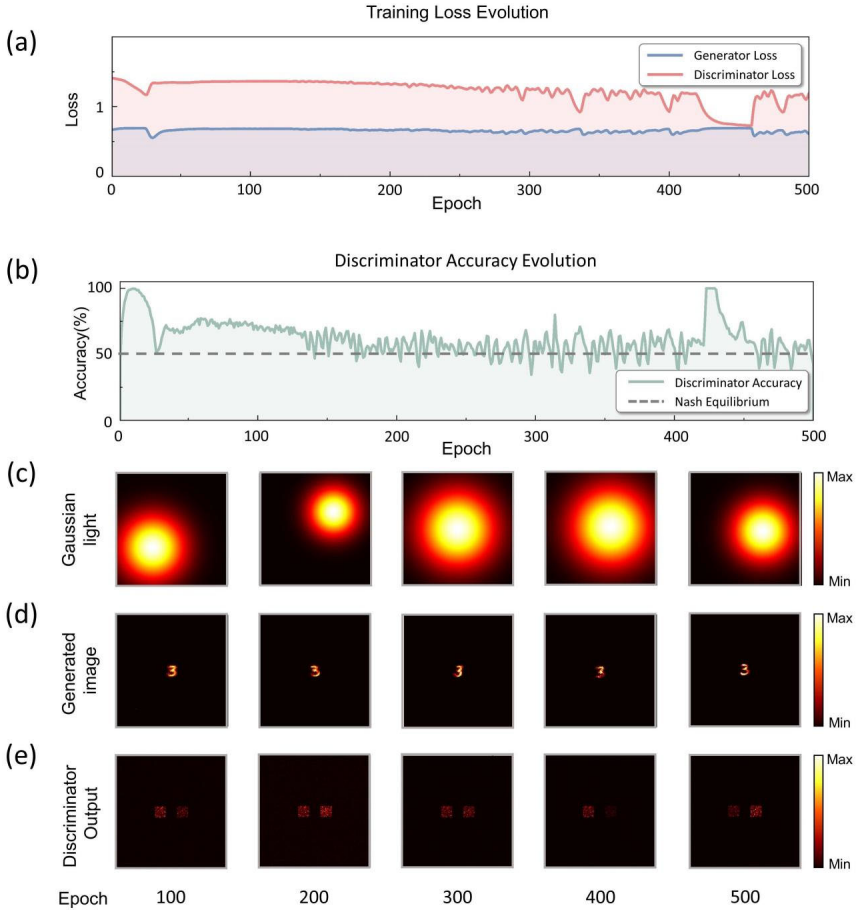


Figure 3. Training dynamics and performance evolution of the D^{2}NN-GAN. (**a**) Training loss evolution for both generator D^{2}NN (blue line) and discriminator D^{2}NN (red line) over 500 epochs, showing the adversarial learning process. (**b**) Discriminator accuracy evolution throughout the training process (green line), with the dashed horizontal line indicating the Nash equilibrium at 50% accuracy. (**c**) Gaussian light input patterns used for generating fake digit images. (**d**) Generated fake digit images produced by the generator D^{2}NN at different training epochs (100, 200, 300, 400 and 500 epochs) in response to the corresponding Gaussian light inputs. (**e**) Discriminator output responses to the generated images, showing the discrimination patterns at each training epoch.

The discriminator accuracy evolution in Figure 3b provides crucial insight into the quality of generated samples. Starting from near-perfect accuracy (>95%), the discriminator's performance gradually degrades as the generator improves, ultimately stabilizing around 50% accuracy—the theoretical Nash equilibrium for perfectly matched adversaries. At around 425 epochs, the discriminator suddenly suppressed the generator. Subsequently, as the network was trained, it returned to the equilibrium point and continued to oscillate. This demonstrated the network's robust adversarial equilibrium, effectively resisting mode collapse or discriminator overload. This convergence to random-chance discrimination definitively indicates that the generated digits have become statistically indistinguishable from real MNIST samples.

Figure 3c–e showcases the evolution of generated digit quality across training epochs. Before epoch 100, the generated "3" digits exhibit basic structure but lack fine details. The discriminator easily identifies these as fake, evidenced by strong activation in the "False" output region. By epoch 150, significant improvement is observed: the digits display clearer definition, consistent curvature and proper proportions. The discriminator's output becomes more ambiguous. At convergence (after epoch 150), the discriminator outputs show nearly equal intensity in both regions, confirming its inability to reliably distinguish generated samples from real samples. Additionally, we analyzed the adversarial generation of other digits (see Supplementary Material Section S3).

To evaluate the quality of generated images, we employed the SSIM and FID metrics. The average SSIM value across 1000 generated images reached 0.9573, while the FID index stood at 106.31. The reasons for the high FID index are multifaceted. Primarily, the number of images used for evaluation significantly impacts FID assessment. It should be noted that under the current settings (i.e., only 1000 evaluation images and single-class generation), the reported FID should be understood as a relative metric for comparison within the same framework, rather than an absolute benchmark directly comparable to reports from large-scale multi-class GANs. This is because FID estimates from limited samples may be biased, particularly when estimating covariance statistics from such small datasets. Typically, incorporating 10,000 or more images (more than 50,000) per calculation is required to achieve a lower index result. And we only had 1000 pictures to participate in the evaluation. Additionally, FID is influenced by the diversity of generated categories. Since we only performed generative adversarial training on the digit "3," this inevitably contributes to the elevated FID value. Therefore, considering the SSIM index as a reference, we attribute the high FID to the numerical relationship limitations of the calculation method rather than poor image quality. More detailed information and computational methods can be found in Supplementary Material Section S1.

The diversity of generated samples, shown across different input beam configurations, demonstrates that our optical GAN successfully captures the multi-modal nature of the MNIST distribution rather than converging to a single mode—a common failure mode in GAN training. Each input beam configuration (varying position and waist size) produces distinct stylistic variations while maintaining digit recognizability, analogous to different handwriting styles in the original dataset.

3.3. D^2NN-GAN for the KTH Dataset

To evaluate the scalability and generalization capability of the D^2NN-GAN architecture, we extended it to the substantially more challenging KTH human action dataset [47]. This dataset presents several complications: higher resolution (160 $\times$ 120 pixels versus 28 $\times$ 28 for MNIST), complex articulated motion patterns and significant intra-class variation. We focused on the "running" action category, centering the extracted human silhouettes to ensure consistent spatial alignment.

Figure 4a,b presents the network configurations for the KTH dataset processing. The optical field evolution for KTH generation (Figure 4c–e) reveals the increased complexity of this task. The generator must synthesize coherent body poses with proper limb articulation and realistic proportions. The evolution of the light field generated by KTH (see Figure 4c) reveals the increased complexity of this task. Among them, the generator must synthesize coherent body postures with appropriate limb joints and true proportions. Figure 4d,e, respectively, show the processing of false images and real images. Among them, the most obvious difference is that during input, the intensity distribution of false images is significantly different from that of real images—false images are more concentrated in the central area of the image, while the intensity at the edges is weaker. There is no significant

difference in intensity in real images. This phenomenon is mainly due to the fact that the false images are just generated by the generation network, and the limitation of modulation capability makes it impossible to balance the overall shape and intensity uniformity of the output. Therefore, it is also easier for the discriminator to identify the difference between the two in terms of intensity.

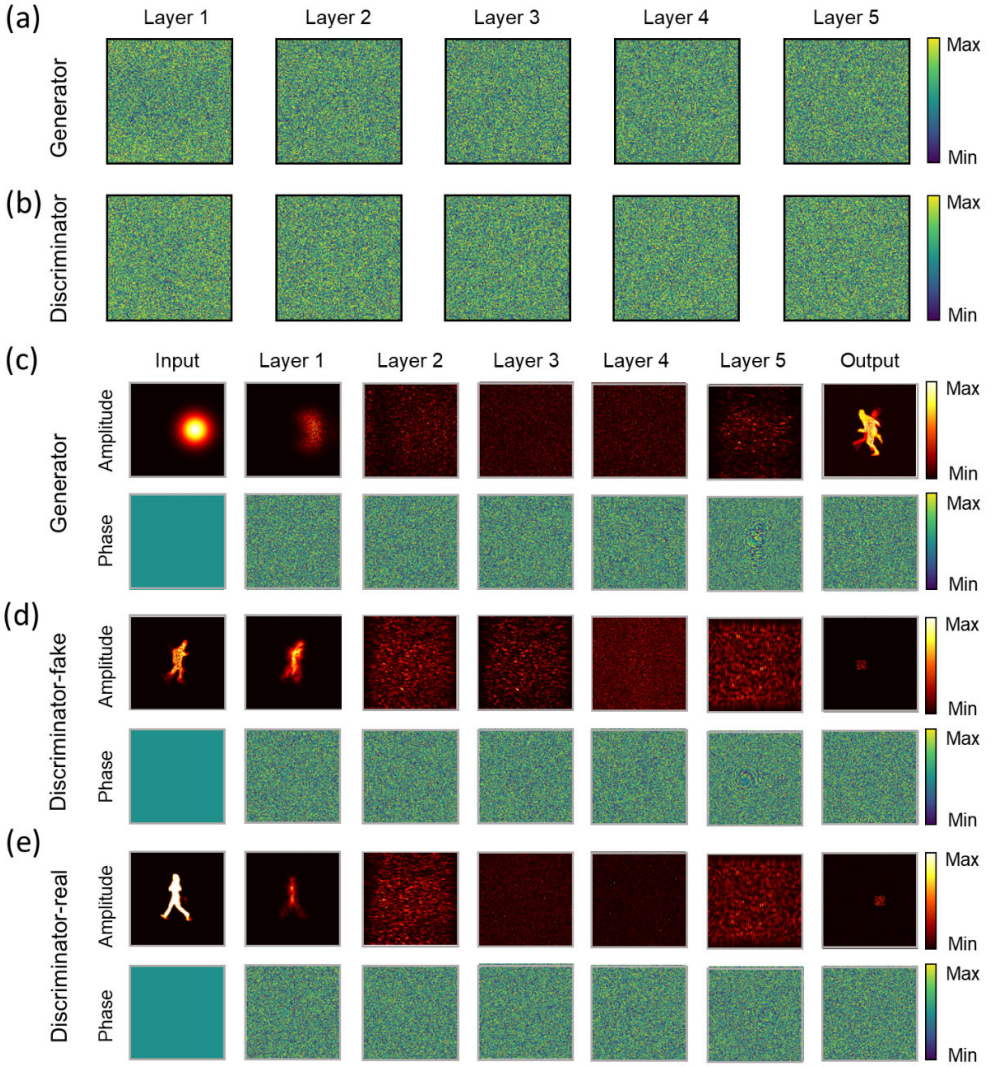


Figure 4. Visualization of D^2NN-GAN architecture and optical field propagation. (**a**) Phase layer parameters of the generator D^2NN across five diffractive layers. (**b**) Phase layer parameters of the discriminator D^2NN across five diffractive layers. (**c**) Amplitude and phase distributions of optical fields propagating through the generator D^2NN, showing the evolution from input to output across all layers. (**d**) Amplitude and phase distributions during the propagation of fake images through the discriminator D2NN, displaying the optical field characteristics at each layer from input to output. (**e**) Amplitude and phase distributions during the propagation of real images through the discriminator D^2NN, showing the corresponding optical field evolution at each diffractive layer.

Figure 5a,b, respectively, show the loss decay curve and discriminator accuracy curve of the KTH dataset during training. Compared to MNIST training, the discriminator accuracy rapidly drops to around 20% during the steep loss decay phase. At approximately 50 epochs, the discriminator accuracy recovers to 50%. At final convergence, the discriminator accuracy stabilizes at approximately 75%. This demonstrates that the discriminator network outperforms the generator network in the KTH task.

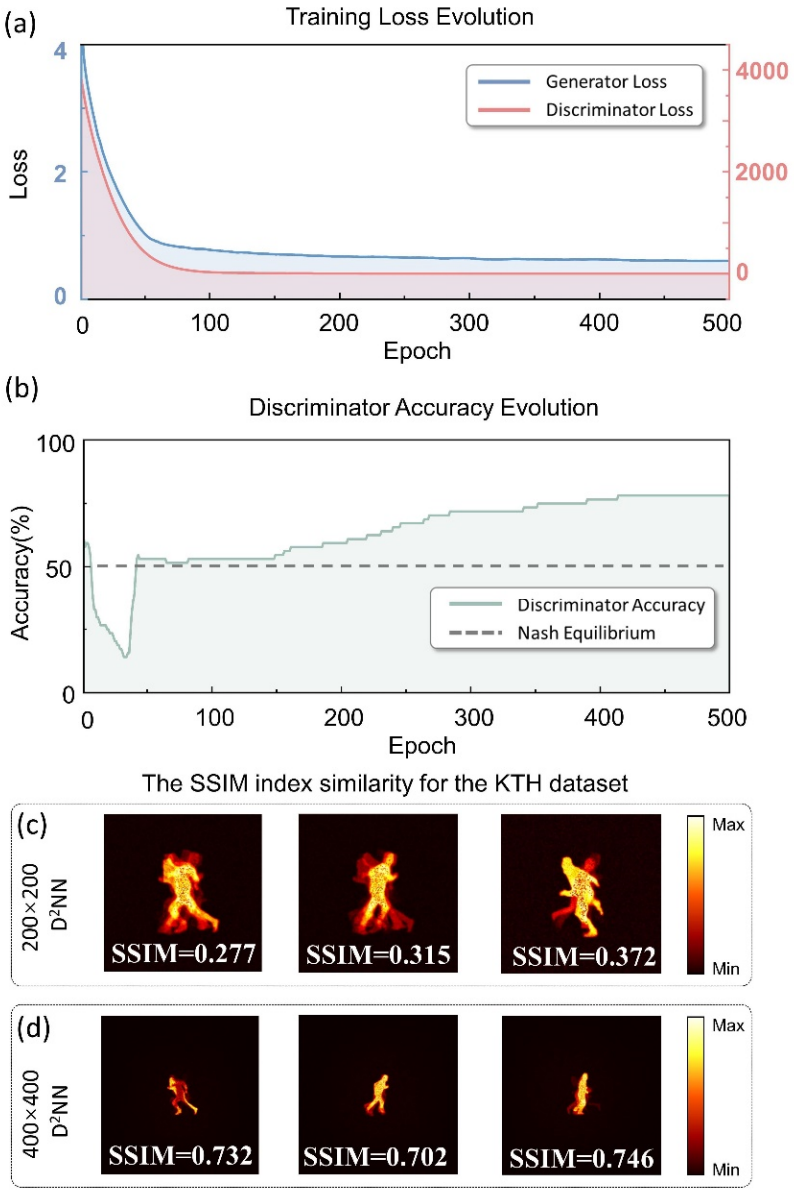


Figure 5. Training performance of D^{2}NN-GAN and the quality of the generated samples. (**a**) The variation in loss values during the 500-epoch training process, where the blue curve represents the generator D^{2}NN loss and the red curve represents the discriminator D^{2}NN loss. (**b**) The evolution of the accuracy rate of Discriminator D^{2}NN during the training process, with the green solid line representing the actual accuracy rate of Discriminator. (**c**) Human action samples reconstructed by 200 × 200 D^{2}NN and their corresponding SSIM indices. (**d**) Human action samples reconstructed by 400 × 400 D^{2}NN and their corresponding SSIM indices.

To investigate factors limiting generation quality in the KTH dataset, we explored various input beam configurations and network architectures. Initially, we attempted to enhance generation quality by diversifying input beam modes, including higher-order Laguerre–Gaussian (LG) modes and vortex beams carrying orbital angular momentum (OAM). These alternative input configurations were hypothesized to provide richer encoding spaces. However, simulation results indicate that multiple Gaussian input modes did not significantly improve generation performance on the KTH dataset. Evaluation metrics (SSIM and FID) reveal negligible differences between outputs generated by the fundamental Gaussian mode and higher-order optical modes, suggesting that beam input diversity alone is insufficient to overcome generation challenges posed by complex motion patterns.

Following these initial investigations, we systematically explored the impact of network scale on generation quality by comparing architectures with different neuron densities per diffractive layer. Specifically, we evaluated two network configurations: 200 × 200 neurons per layer (40,000 trainable parameters per layer) and 400 × 400 neurons per layer (160,000 trainable parameters per layer). This four-fold increase in network capacity provides substantially greater representational power for learning complex optical transformations.

The generated human motion images reveal significant performance differences between network scales (see Figure 5c,d). Among these, the 200 × 200 network generated image samples exhibited significant ghosting artifacts, with SSIM indices hovering around 0.3. In contrast, the 400 × 400 network demonstrated substantial improvements, producing samples with more coherent body structures, better-defined limbs and more realistic proportions, while elevating the SSIM index to approximately 0.7. Simultaneously, the FID score for the 100 generated images decreased from 253.71 to 204.81. While still imperfect, these larger network structures captured recognizable running poses and established appropriate spatial relationships between body parts. The discriminator accuracy saturating at ~75% indicates a persistent advantage of the discriminator over the generator in this high-dimensional setting. This is because both the discriminator and generator utilize the same parameterized network architecture; generating a 160 × 120 image is significantly more challenging than inferring its authenticity. In practice, GAN stability can be improved by explicitly rebalancing generator–discriminator strength, for example, by adopting asymmetric architectures (e.g., using a larger-capacity generator while keeping a lighter discriminator) and/or by weakening discriminator updates. Therefore, for more complex tasks, matching smaller discriminator networks with larger generator networks should be considered to achieve better D^2NN-GAN performance. Such asymmetric G–D configurations are particularly relevant to KTH datasets because synthesizing structured human silhouette generation is inherently more demanding than binary authenticity inference.

4. Experimental Error Discussion

Although the present work verifies D^2NN-GAN in simulation, several engineering factors are critical for physical deployment, including interlayer alignment, environmental disturbances and fabrication tolerances of large-scale diffractive elements.

4.1. Quantitative Tolerance to Inter-Layer Translational Misalignment

In practical multi-layer diffractive implementations, precise layer-to-layer alignment is a key engineering challenge, because translational misalignment introduces spatial mismatch between adjacent phase masks and effectively breaks the trained optical transformation. To quantify the sensitivity of our D^2NN-GAN to such alignment errors, we performed a systematic misalignment sweep based on the trained MNIST configuration.

Specifically, we considered cases where N = 1–5 diffractive layers are misaligned ("misaligned_layers"), with a translational offset of 5% (shift_percent = 0.05) or 10% (shift_percent = 0.10) of the 200 × 200-layer, corresponding to 10 px and 20 px, respectively. For each setting, the offset was applied along either the horizontal or vertical direction ("axis"). After convergence (epoch = 500), we evaluated the generation quality using FID and SSIM.

Overall, Table 1 shows that SSIM remains relatively robust under moderate misalignment in many cases (typically remaining ≥ 0.93), whereas FID is noticeably more sensitive and exhibits larger fluctuations across different misalignment configurations. In this tolerance test, FID should be interpreted as a relative indicator across misalignment conditions rather than an absolute benchmark.

Table 1. Misalignment tolerance analysis of the D^2NN-GAN generator (MNIST).

Misaligned Layers	Shift	Axis	Final FID	SSIM (Mean ± Std)
1	5% (10 px)	Horizontal	250.79	0.9530 ± 0.0192
	5% (10 px)	Vertical	255.58	0.9530 ± 0.0163
	10% (20 px)	Horizontal	245.38	0.9480 ± 0.0279
	10% (20 px)	Vertical	181.91	0.9505 ± 0.0208
2	5% (10 px)	Horizontal	143.66	0.9523 ± 0.0239
	5% (10 px)	Vertical	157.96	0.9540 ± 0.0150
	10% (20 px)	Horizontal	242.57	0.9504 ± 0.0212
	10% (20 px)	Vertical	233.92	0.9530 ± 0.0201
3	5% (10 px)	Horizontal	167.73	0.9551 ± 0.0179
	5% (10 px)	Vertical	157.89	0.9542 ± 0.0174
	10% (20 px)	Horizontal	215.57	0.9438 ± 0.0262
	10% (20 px)	Vertical	253.72	0.9355 ± 0.0299
4	5% (10 px)	Horizontal	221.43	0.9449 ± 0.0217
	5% (10 px)	Vertical	241.58	0.9511 ± 0.0223
	10% (20 px)	Horizontal	72.96	0.9739 ± 0.0077
	10% (20 px)	Vertical	73.97	0.9733 ± 0.0119
5	5% (10 px)	Horizontal	202.62	0.9325 ± 0.0235
	5% (10 px)	Vertical	233.88	0.9354 ± 0.0304
	10% (20 px)	Horizontal	108.33	0.9586 ± 0.0116
	10% (20 px)	Vertical	154.31	0.8963 ± 0.0514

4.2. Temperature/Mechanical Variations

Temperature variations have negligible effects on diffractive networks (passive systems). Temperature variations typically introduce only weak perturbations in passive diffractive networks because the optical computation is carried out by free-space propagation and static phase profiles (i.e., no active biasing is required during inference). Recent integrated diffractive neural network chips have experimentally demonstrated high-temperature robustness and stable inference in harsh environments, supporting the practical stability of passive diffractive processors [48]. Mechanical vibration mainly manifests as relative interlayer misalignment (lateral shift/axial spacing/tilt) rather than intrinsic phase noise of the passive layers.

4.3. Manufacturing Precision

Regarding manufacturing precision for large-scale diffractive elements, experimental implementation of D^2NN typically uses a spatial light modulator to modulate the light source and 3D printing to fabricate metasurfaces designed by an electronic computer. Limited by the precision size of 3D printing, this fabrication method is typically only available for terahertz bands [20]. Here, the D^2NN-GAN operates at the wavelength of 1550 nm, which corresponds to pixel sizes of approximately 800 nm. The diffractive layer of D^2NN-GAN can be fabricated by micro/nano processing technology compatible with CMOS technology, as the current state-of-the-art e-beam lithography (EBL) technology has a fabrication resolution of only a few nanometers. Nevertheless, multilayer integration still requires careful control of overlay/alignment and interlayer spacing, which remains an important engineering topic for scalable diffractive/metasurface computing hardware.

4.4. Real-Time Optimization

Beyond static fabrication, practical "real-time" optimization can be achieved through SLM-based prototyping and calibration to compensate system imperfections [49], and/or measurement-in-the-loop in situ optimization/training methods [50] recently demonstrated

for diffractive optical processors and trainable on-chip diffractive architectures with massive tunable elements.

The lack of reconfigurability is an inherent characteristic of passive diffractive hardware rather than a limitation specific to the D^2NN-GAN framework. Similarly to conventional large-scale generative models, once the diffractive parameters are optimized offline and deployed, the inference/generation process can run with ultra-low incremental energy cost and massive parallelism; task switching can be realized by re-fabricating/swapping phase masks or by introducing programmable diffractive layers when on-demand reconfiguration is required.

5. Conclusions

In this work, we have numerically demonstrated the feasibility of implementing generative adversarial networks using all-optical diffractive layers. The simulation results demonstrate that a dual-network architecture composed of five layers of diffractors and discriminators can achieve dynamic generative adversarial learning in the optical domain. For the MNIST digit dataset, the simulated system reaches Nash equilibrium with the discriminator accuracy stabilizing around 50%, while the generator achieves an average SSIM of 0.9573. The extension to the KTH human action dataset, though not achieving Nash equilibrium (discriminator accuracy ~75%), demonstrates the potential scalability of this approach to more complex visual data.

However, the current network still exhibits certain limitations: constrained by the depth of modulation (five layers) and the number of modulating neurons, coupled with potentially limited input encoding diversity, the network's adversarial generation capability is somewhat restricted. This limitation is particularly evident in the results from the KTH dataset.

Despite these limitations, optical D^2NN-GAN has potential natural advantages over electronic neural networks in terms of processing speed (achieved through multidimensional multiplexing of diffractive networks) and energy efficiency (the entire network consists of passive components, resulting in significantly lower overall power requirements for image generation compared to traditional neural networks). Compared to other optical GANs, D^2NN-GAN integrates both optical generation and inference models within a single architecture. Especially at the input end, we can achieve the demonstration effect by using Gaussian beams of different positions and beam waist radius; that is, no active devices are needed for encoding at the input end.

Beyond the image generation demonstrated in this work, the D^2NN-GAN framework can be extended in several practical directions. First, conditional or controllable generation can be realized by multiplexing additional optical input channels (e.g., spatial/wavelength/polarization encoding) to incorporate conditions, consistent with the task categories explored in optoelectronic GAN demonstrations (image generation and conditional generation) [41]. Second, image restoration (e.g., denoising or inpainting) can be formulated by training the generator with paired corrupted/target intensity distributions. Third, for spatiotemporal data, the architecture can be combined with time-multiplexed illumination or multi-frame outputs to support motion-related synthesis, providing a physically interpretable route toward optical video-oriented generative modeling. Meanwhile, it is possible to consider matching and combining large-sized generative networks with small-sized discriminative networks to enhance their ability to handle generative adversarial situations in high-dimensional complex data. In conclusion, we believe the D^2NN-GAN architecture introduces a fully optical generative adversarial model, providing a practical path for future optical modeling methods such as image generation and video synthesis.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/photonics13010094/s1, Figure S1: Category generalization results of the proposed D^2NN-GAN on MNIST; Table S1: Quantitative comparison of generation quality across different network scales and datasets; Table S2: Quantitative comparison between this work and representative optical generative models.

Author Contributions: Conceptualization, P.H., T.C., Y.Z. and S.F.; methodology, P.H. and T.C.; software, P.H. and Y.Z.; validation, P.H. and T.C.; formal analysis, P.H. and T.C.; investigation, P.H. and T.C.; resources, P.H. and S.F.; data curation, P.H. and Y.Z.; writing—original draft preparation, P.H.; writing—review and editing, P.H., T.C., Y.Z. and S.F.; visualization, P.H. and Y.Z.; supervision, P.H. and S.F.; project administration, P.H.; funding acquisition, P.H. and S.F. All authors have read and agreed to the published version of the manuscript.

Funding: The National Natural Science Foundation of China (Grant No. 62571562, 12274478); the Fundamental Research Funds for the Central Universities (ZY2525).

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the author upon reasonable request.

Acknowledgments: The authors wish to thank the anonymous reviewers and the associate editor for their valuable suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
2. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Lake Tahoe, NV, USA, 3–6 December 2012; Volume 25, pp. 1097–1105.
3. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [CrossRef]
4. Strubell, E.; Ganesh, A.; McCallum, A. Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 3645–3650.
5. Thompson, N.C.; Greenewald, K.; Lee, K.; Manso, G.F. The computational limits of deep learning. *arXiv* **2020**, arXiv:2007.05558.
6. Patterson, D.; Gonzalez, J.; Le, Q.V.; Liang, C.; Munguia, L.-M.; Rothchild, D.; So, D.; Texier, M.; Dean, J. Carbon emissions and large neural network training. *arXiv* **2021**, arXiv:2104.10350. [CrossRef]
7. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. Green AI. *Commun. ACM* **2020**, *63*, 54–63.
8. Shastri, B.J.; Tait, A.N.; Ferreira de Lima, T.; Pernice, W.H.P.; Bhaskaran, H.; Wright, C.D. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **2021**, *15*, 102–114.
9. Wetzstein, G.; Ozcan, A.; Gigan, S.; Fan, S.; Englund, D.; Soljačić, M.; Denz, C.; Miller, D.A.B.; Psaltis, D. Inference in artificial intelligence with deep optics and photonics. *Nature* **2020**, *588*, 39–47. [CrossRef]
10. Kitayama, K.; Notomi, M.; Naruse, M.; Inoue, K.; Kawakami, S.; Uchida, A. Novel frontier of photonics for data processing—Photonic accelerator. *APL Photonics* **2019**, *4*, 090901. [CrossRef]
11. Wang, Y.; Liao, K.; Zhang, K.; Du, Z.; Wang, Z.; Ni, B.; Xu, T.; Feng, S.; Yang, Y.; Yang, Q.-F.; et al. Reconfigurable versatile integrated photonic computing chip. *eLight* **2025**, *5*, 20. [CrossRef]
12. Li, D.; Zhang, K.; Hu, X.; Feng, S. Integrated convolutional kernel based on two-dimensional photonic crystals. *Opt. Lett.* **2024**, *49*, 6297–6300. [CrossRef]
13. Tian, X.; Li, R.; Peng, T.; Xue, Y.; Min, J.; Li, X.; Bai, C.; Yao, B. Multi-prior physics-enhanced neural network enables pixel super-resolution and twin-image-free phase retrieval from single-shot hologram. *Opto-Electron. Adv.* **2024**, *7*, 240060. [CrossRef]
14. Gigan, S. Data-driven polarimetric approaches fuel computational imaging expansion. *Opto-Electron. Adv.* **2024**, *7*, 240158. [CrossRef]
15. Miller, D.A.B. Attojoule optoelectronics for low-energy information processing and communications. *J. Light. Technol.* **2017**, *35*, 346–396. [CrossRef]
16. Prucnal, P.R.; Shastri, B.J. *Neuromorphic Photonics*; CRC Press: Boca Raton, FL, USA, 2017.
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*, 208–214. [CrossRef] [PubMed]

18. Larger, L.; Baylón-Fuentes, A.; Martinenghi, R.; Udaltsov, V.S.; Chembo, Y.K.; Jacquot, M. High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification. *Phys. Rev. X* **2017**, *7*, 011015. [CrossRef]
19. Vandoorne, K.; Mechet, P.; Van Vaerenbergh, T.; Fiers, M.; Morthier, G.; Verstraeten, D.; Schrauwen, B.; Dambre, J.; Bienstman, P. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **2014**, *5*, 3541. [CrossRef]
20. 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*, 1004–1008. [CrossRef]
21. Zhang, K.; Liao, K.; Cheng, H.; Feng, S.; Hu, X. Advanced all-optical classification using orbital-angular-momentum-encoded diffractive networks. *Adv. Photonics Nexus* **2023**, *2*, 066006. [CrossRef]
22. Lin, X.; Fu, Y.; Zhang, K.; Huang, H.; Xu, F.; Fan, J.; Lin, X.; Dai, Q.; Hu, X. Polarization and wavelength routers based on diffractive neural network. *Front. Optoelectron.* **2024**, *17*, 22. [CrossRef]
23. Lin, X.; Zhang, K.; Liao, K.; Huang, H.; Fu, Y.; Zhang, X.; Feng, S.; Hu, X. Polarization-based all-optical logic gates using diffractive neural networks. *J. Opt.* **2024**, *26*, 035701. [CrossRef]
24. Zhang, Y.; Zhang, K.; Hu, P.; Li, D.; Feng, S. Multi-wavelength diffractive optical neural network integrated with 2D photonic crystals for joint optical classification. *Nanophotonics* **2025**, *14*, 2891–2899. [CrossRef] [PubMed]
25. 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*, 112. [CrossRef]
26. 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*, 1–14. [CrossRef]
27. Rahman, M.S.S.; Ozcan, A. Computer-free, all-optical reconstruction of holograms using diffractive networks. *ACS Photonics* **2021**, *8*, 3375–3384. [CrossRef]
28. He, C.; Zhao, D.; Fan, F.; Zhou, H.; Li, X.; Li, Y.; Li, J.; Dong, F.; Miao, Y.-X.; Wang, Y.; et al. Pluggable multitask diffractive neural networks based on cascaded metasurfaces. *Opto-Electron. Adv.* **2024**, *7*, 230005. [CrossRef]
29. Yan, T.; Wu, J.; Zhou, T.; Xie, H.; Xu, F.; Fan, J.; Fang, L.; Lin, X.; Dai, Q. Fourier-space diffractive deep neural network. *Phys. Rev. Lett.* **2019**, *123*, 023901. [CrossRef] [PubMed]
30. Chen, H.; Feng, J.; Jiang, M.; Wang, Y.; Lin, J.; Tan, J.; Jin, P. Diffractive deep neural networks at visible wavelengths. *Engineering* **2021**, *7*, 1483–1491. [CrossRef]
31. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Montreal, QC, Canada, 8–11 December 2014; Volume 27, pp. 2672–2680.
32. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
33. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
34. Karras, T.; Aittala, M.; Laine, S.; Härkönen, E.; Hellsten, J.; Lehtinen, J.; Aila, T. Alias-free generative adversarial networks. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Online, 6–14 December 2021; Volume 34, pp. 852–863.
35. Brock, A.; Donahue, J.; Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
36. Méndez-Lucio, O.; Baillif, B.; Clevert, D.-A.; Rouquié, D.; Wichard, J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat. Commun.* **2020**, *11*, 10. [CrossRef]
37. Dan, Y.; Zhao, Y.; Li, X.; Li, S.; Hu, M.; Hu, J. Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials. *npj Comput. Mater.* **2020**, *6*, 84. [CrossRef]
38. Arjovsky, M.; Bottou, L. Towards principled methods for training generative adversarial networks. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
39. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp. 214–223.
40. Chen, S.; Li, Y.; Wang, Y.; Chen, H.; Ozcan, A. Optical generative models. *Nature* **2025**, *644*, 903–911. [CrossRef] [PubMed]
41. Zhan, Z.; Wang, H.; Liu, Q.; Fu, X. Photonic diffractive generators through sampling noises from scattering media. *Nat. Commun.* **2024**, *15*, 10643. [CrossRef] [PubMed]
42. Qiu, J.; Lu, G.; Liu, T.; Zhang, D.; Xiao, S.; Yu, T. Optoelectronic generative adversarial networks. *Commun. Phys.* **2025**, *8*, 162. [CrossRef]
43. Dong, Y.; Bai, Y.; Zhang, Q.; Luan, H.; Gu, M. High-throughput optical neuromorphic graphic processing at millions of images per second. *eLight* **2025**, *5*, 29. [CrossRef]
44. Goodman, J.W. *Introduction to Fourier Optics*; Roberts and Company Publishers: Greenwood Village, CO, USA, 2005.

45. Allen, L.; Beijersbergen, M.W.; Spreeuw, R.J.C.; Woerdman, J.P. Orbital angular momentum of light and the transformation of Laguerre–Gaussian laser modes. *Phys. Rev. A* **1992**, *45*, 8185–8189. [CrossRef]
46. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
47. Schuldt, C.; Laptev, I.; Caputo, B. Recognizing human actions: A local SVM approach. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 23–26 August 2004; Volume 3, pp. 32–36.
48. Dong, Y.; Lin, D.; Chen, L.; Li, B.; Chen, X.; Zhang, Q.; Luan, H.; Fang, X.; Gu, M. Compact eternal diffractive neural network chip for extreme environments. *Commun. Eng.* **2024**, *3*, 64. [CrossRef]
49. Cheng, J.; Huang, C.; Zhang, J.; Wu, B.; Zhang, W.; Liu, X.; Zhang, J.; Tang, Y.; Zhou, H.; Zhang, Q.; et al. Multimodal deep learning using on-chip diffractive optics with in situ training capability. *Nat. Commun.* **2024**, *15*, 6189. [CrossRef] [PubMed]
50. Li, Y.; Chen, S.; Gong, T.; Ozcan, A. Model-free optical processors using in situ reinforcement learning with proximal policy optimization. *Light Sci. Appl.* **2026**, *15*, 32. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.