

## Research Article

Matan Kleiner, Lior Michaeli and Tomer Michaeli\*

# Can the success of digital super-resolution networks be transferred to passive all-optical systems?

<https://doi.org/10.1515/nanoph-2025-0294>

Received June 23, 2025; accepted August 12, 2025;

published online September 8, 2025

**Abstract:** The deep learning revolution has increased the demand for computational resources, driving interest in efficient alternatives like all-optical diffractive neural networks (AODNNs). These systems operate at the speed of light without consuming external energy, making them an attractive platform for energy-efficient computation. One task that could greatly benefit from an all-optical implementation is spatial super-resolution. This would allow overcoming the fundamental resolution limitation of conventional optical systems, dictated by their numerical aperture. Here, we examine whether the success of digital super-resolution networks can be replicated with AODNNs considering networks with phase-only nonlinearities. We find that while promising, super-resolution AODNNs face two key physical challenges: (i) a tradeoff between reconstruction fidelity and energy preservation along the optical path and (ii) a limited dynamic range of input intensities that can be effectively processed. These findings offer a first step toward understanding and addressing the design constraints of passive, all-optical super-resolution systems.

**Keywords:** diffractive neural networks; all-optical super-resolution; nonlinear optical computing

## 1 Introduction

The spatial resolution of conventional optical systems, such as microscopes, telescopes, and cameras, is fundamentally constrained by their numerical aperture (NA) and the wavelength of light. These systems act as low-pass filters, attenuating high spatial frequencies and thus limiting the ability to resolve fine details [1]. Overcoming this diffraction limit has been a longstanding challenge across numerous scientific fields, from biology [2] to astronomy [3]. Most existing super-resolution methods modify the acquisition process to allow measuring high spatial frequencies and use linear optics to collect light. For example, some super-resolution techniques rely on near-field measurements of the imaged object, extracting high-frequency information encoded in evanescent waves using metamaterials [4], [5]. Methods like structured illumination [6], [7], localization microscopy [8]–[10], and optical stethoscopy [11] trade temporal resolution for spatial resolution. Other techniques exploit super-oscillatory phenomena [12], [13] or the statistical behavior of speckle and scattering media [14]–[17]. However, the requirement to modify the acquisition process poses severe limitations, rendering these techniques impractical for many real-world scenarios.

In contrast to optical techniques, digital super-resolution methods attempt to computationally recover the lost high frequencies. This is done by exploiting prior knowledge on the typical behavior of high-resolution images in the domain of interest (*e.g.*, cellular organelles [18], galaxies [19], natural scenery images [20], *etc.*). Over the last decade, this field has seen significant advancements thanks to the adoption of deep learning methods [18], [20]–[26]. Unlike classical computational methods, such as Richardson–Lucy deconvolution [27], [28] and its extensions, which only enforce simple priors on the restored image (*e.g.*, smoothness), deep learning approaches leverage the ability of neural networks to learn complex nonlinear mappings between low-resolution images and their high-resolution counterparts based

\*Corresponding author: Tomer Michaeli, Faculty of Electrical and Computer Engineering, Technion, Haifa, Israel,  
E-mail: tomer.m@ee.technion.ac.il  
<https://orcid.org/0000-0003-0525-8054>

Matan Kleiner, Faculty of Electrical and Computer Engineering, Technion, Haifa, Israel, E-mail: matankleiner@campus.technion.ac.il  
<https://orcid.org/0009-0000-2621-5286>

Lior Michaeli, School of Electrical & Computer Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel; and Light Matter Interaction Center, Tel Aviv University, Tel Aviv 6997801, Israel,  
E-mail: liormic1@tauex.tau.ac.il <https://orcid.org/0000-0002-9351-1111>

on training examples [20]. Specifically, these methods implicitly extract complex image priors from the training data (e.g., the valid shapes of handwritten letters) and thus achieve state-of-the-art reconstruction accuracy, as well as robustness to noise [29]. Unfortunately, however, state-of-the-art digital super-resolution networks are often challenging to deploy on end devices due to their intense computational requirements [30]–[32]. Recently, it has been recognized that many computational tasks that can be performed with digital networks can also be effectively implemented all-optically [31], [33]–[36]. In particular, all-optical diffractive neural networks [34] emerged as a promising tool for visual information processing. Here, we numerically investigate whether all-optical diffractive networks, using only passive elements, can replicate the success of digital super-resolution networks.

We show that while passive diffractive networks can indeed enhance resolution all-optically, they suffer from two fundamental tradeoffs that are absent in digital networks: (i) a tradeoff between the reconstruction quality and energy preservation and (ii) high sensitivity to the input global intensity. These two tradeoffs stem from how light propagates and interacts with passive nonlinear media. In particular, each nonlinear layer generates high spatial frequencies that correspond to waves propagating at large angles. Some of these may inevitably escape the next layer's collection cone, thus causing a loss of energy. Additionally, while images in the digital domain are normalized, natural scenes may exhibit a large range of intensities that correspond to markedly different regions of the nonlinearity. This limits the system's applicability under changing illumination conditions.

## 2 Passive all-optical super-resolution neural network

The principles we would like to explore are not inherent to a particular passive diffractive network architecture. However, for concreteness, we focus on an all-optical super-resolution neural network (AOSRNN) architecture inspired by digital convolutional neural networks (CNNs), as illustrated in Figure 1a. This network consists of a series of diffractive convolutional units, each comprising a  $4f$  system with a learned phase mask in the Fourier plane and an all-optical nonlinear layer at the image plane. We assume energy-preserving, phase-only nonlinear layers, whose refractive index  $n$  at spatial location  $(x, y)$  depends nonlinearly on the intensity  $I(x, y)$  at that location, as

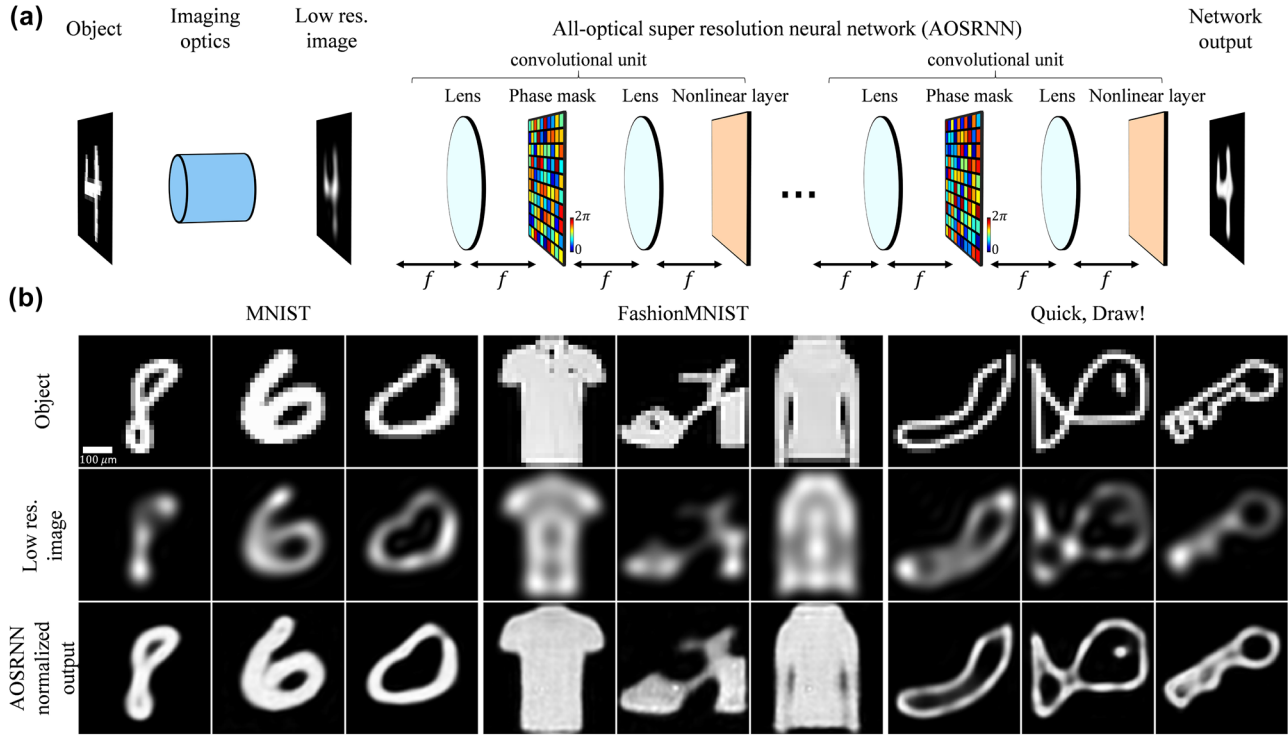
$$n(x, y) = n_0 + n_2 I(x, y). \quad (1)$$

Here,  $n_0$  is the linear refractive index,  $n_2$  is the nonlinear refractive index, and  $I(x, y) = |E(x, y)|^2$  is the intensity of the optical field, which we assume to be monochromatic [37]. While we focus here on phase-only nonlinearities, similar considerations apply to amplitude nonlinearities, which we analyze in Supplementary Note 1. It is important to note that achieving nonlinearity with sufficiently large  $n_2$  is a challenge in its own right and would come at a cost of slow response time. However, the goal of this paper is to illustrate that even if such nonlinear layers were available, other phenomena would still severely limit the network's utility (see Supplementary Note 1 for further discussions).

The input plane of AOSRNN contains a low-resolution image of an object, obtained from some imaging optics with small NA (left part of Figure 1a). This input image corresponds to a low-pass filtered version of the object, which we assume is illuminated by coherent light. The goal of AOSRNN is to reconstruct the object at its output plane, which coincides with the plane of the last nonlinear layer. Achieving this requires generating spatial frequencies higher than those present at the input image plane, a task facilitated by the nonlinear layers.

To explore the ability of AOSRNN to enhance spatial resolution, we simulated a network composed of 10 convolutional units (see Supplementary Note 2 for the effect of the network's depth). The input to the network is a low-resolution image, generated by imaging optics with NA of approximately 0.01.

Throughout the paper, we assume that we do not have access to the imaging optics, which produces images of a given resolution. Therefore, our goal is to enhance the input image resolution (relative improvement rather than absolute resolution) by placing the AOSRNN after the imaging system. Assuming a wavelength of  $\lambda = 550$  nm, this corresponds to a cutoff spatial frequency of  $f_c = \frac{\text{NA}}{\lambda} = \frac{0.01}{550 \text{ nm}} = 19,345.2 \text{ m}^{-1}$ , which sets the resolution at the input plane to  $\Delta x = 1/f_c = 51.7 \text{ }\mu\text{m}$ . We examined a case in which AOSRNN can potentially improve resolution by a factor of up to 4.3. To support this, the apertures of the Fourier plane convolutional units (i.e., the phase mask sizes) were chosen to correspond to a resolution of  $12 \text{ }\mu\text{m}$ , which translates to a cutoff frequency of  $f_c = 83,333.3 \text{ m}^{-1}$ . This represents the highest possible resolution achievable at the network's output. As shown below, AOSRNN is capable of improving the input image resolution, albeit not fully reaching this theoretical limit. AOSRNN was trained by minimizing the discrepancy between the object intensity,  $I(x, y)$ , and the intensity at the network's output,  $\hat{I}(x, y)$ . For simplicity, we represent these quantities as column vectors, omitting



**Figure 1:** All-optical super-resolution neural architecture (AOSRNN). (a) Schematic of the proposed network. An optical imaging system captures an object and produces a low-resolution image, limited by its NA, which serves as the input to the AOSRNN. This image is then processed by the all-optical neural network, which outputs a high-resolution reconstruction. The network is composed of consecutive convolutional units, each consisting of a  $4f$  system with a learnable phase mask layer in the Fourier plane, followed by a nonlinear optical layer. (b) Qualitative results on the MNIST, FashionMNIST, and Quick, Draw! datasets. The first, second, and third rows display the high-resolution object, low-resolution network input, and high-resolution network output, respectively. We note that the ripples observed in the low-resolution images result from the use of a low-NA imaging system with spatially and temporally coherent illumination.

their coordinates. We used the  $\ell_1$  norm between the object and the normalized network output as our loss function. Specifically, let  $I^{(i)}$  denote the  $i$ th example in the training set and  $\hat{I}^{(i)}$  denote the corresponding network prediction. We constructed a normalized variant of  $\hat{I}^{(i)}$  by dividing it by its maximal value and multiplying it by the maximal intensity value across all objects in the dataset, namely  $\hat{I}_{\text{norm}}^{(i)} = c \cdot \hat{I}^{(i)}$ , where  $c = \frac{\max_j \|I^{(j)}\|_\infty}{\|\hat{I}^{(i)}\|_\infty}$ . We then defined the training loss as

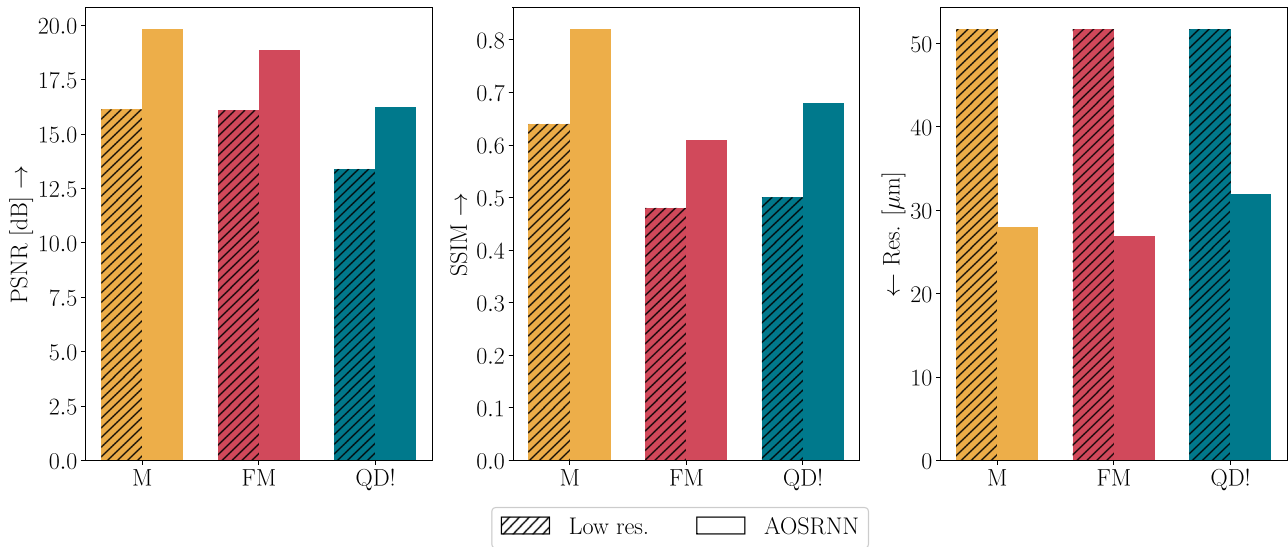
$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|\hat{I}_{\text{norm}}^{(i)} - I^{(i)}\|_1, \quad (2)$$

where  $N$  is the number of training samples. The normalization was chosen to address the decrease in image intensity that accompanies resolution enhancement, a key challenge that will be thoroughly discussed in the remainder of the paper.

Figure 1b shows several normalized results obtained from three different AOSRNNs, one trained on the MNIST dataset of handwritten digits [38], one on the FashionMNIST dataset of fashion items [39], and one on a subset of the Quick, Draw! dataset, which contains drawings of everyday

objects [40]. All three networks effectively recover high-resolution images from their low-resolution counterparts.

Figure 2 reports a quantitative evaluation of AOSRNN's reconstruction quality. The left and middle panes show two measures of discrepancy between the network's outputs and the original objects: the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [41]. For both metrics, higher scores indicate greater similarity. As can be seen, the similarity between the network's outputs and the ground-truth images is greater than the similarity between the input low-resolution images and the ground-truth images, suggesting that AOSRNN successfully improves resolution. The right pane of Figure 2 provides a more direct measure of resolution improvement. To quantify resolution, we look for a linear, NA-limited imaging system that achieves the same average PSNR as AOSRNN. We then define the resolution of AOSRNN as the inverse of the cutoff frequency of this linear imaging system (see Supplementary Note 3 for more details). In this case, lower values correspond to better resolution. The bar plot confirms that AOSRNN improves the resolution of the input, with an average enhancement factor of 1.8. However, we note that this



**Figure 2:** Quantification of resolution enhancement by the all-optical network. Bar plots compare the similarity between the original object and both the low-resolution input (diagonal-striped bars) and the AOSRNN output (solid bars). Similarity is quantified using three metrics: PSNR (left), SSIM (middle), and effective resolution (right). Results are shown for three datasets: MNIST (M), FashionMNIST (FM), and Quick, Draw! (QD).

improvement is still below the theoretical maximum factor of 4.3, which corresponds to the ultimate resolution bound of  $12\text{ }\mu\text{m}$ . We additionally evaluated AOSRNN's performance using input images generated by imaging optics with NAs of approximately  $\{0.006, 0.004, 0.003\}$ , corresponding to input resolutions of  $\Delta x = \{84, 134.4, 168\}\text{ }\mu\text{m}$ . For these cases, AOSRNN enhances resolution by an average factor of  $\{2, 2.6, 3\}$ , respectively. See Supplementary Note 4 for more details.

### 3 Fundamental tradeoffs

As demonstrated above, assuming the availability of suitable nonlinear layers, all-optical diffractive networks, like AOSRNN, can enhance resolution all-optically. However, as we now illustrate, nonlinear optical super-resolution systems suffer from fundamental tradeoffs that are absent in digital networks. In the following subsections, we discuss two such tradeoffs, both inherently tied to the properties of light and thus expected to persist in any type of passive all-optical neural network, regardless of the specific implementation of the nonlinearity. However, we discuss how the working point can be chosen along the tradeoffs.

#### 3.1 Resolution versus energy preservation

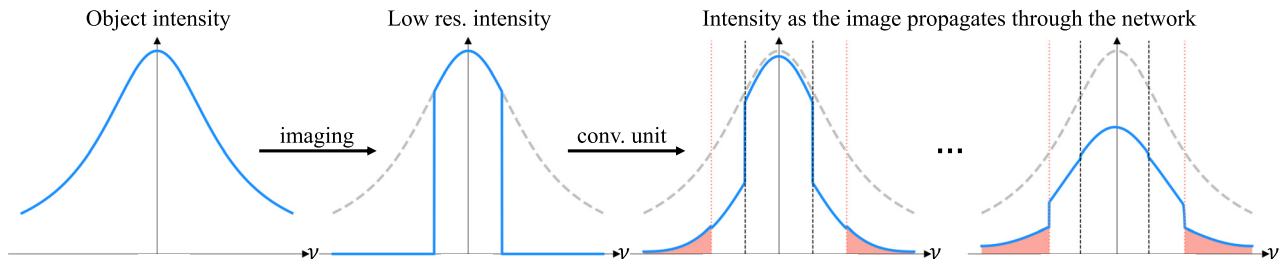
Our first observation is that, while the nonlinear AOSRNN achieves impressive reconstruction performance, it does so at the cost of significant optical power loss. Only a small fraction of the input field energy is preserved at the output plane of the network, just a few percent (6%) in the

example shown in Figure 1b. This energy loss is not specific to the architecture of AOSRNN but is instead inherent to the physics of light interacting with nonlinear optical layers. Specifically, as light propagates through such systems, the nonlinear layers generate new high spatial frequencies, corresponding to larger diffraction angles. Due to the limited numerical aperture (NA) of the system, some of these high-angle components lie outside the collection cone of the subsequent layers and are thus inevitably lost.

This effect is illustrated in Figure 3, using a simplified one-dimensional example. The spatial frequency content of a typical object naturally decays with increasing spatial frequency (leftmost plot). A low-resolution version of the object is obtained by the imaging optics that precede the network, effectively acting as a low-pass filter that suppresses frequencies beyond the system's NA (second plot).

As the field propagates through the network, nonlinear interactions give rise to new spatial frequency components that were not present in the initial low-resolution image. These new high-frequency components are crucial for recovering fine spatial details of the original object. The convolutional units act to shape this growing frequency content, steering the evolution of the field toward a high-fidelity reconstruction of the object up to a global normalization constant, as formalized by Eq. (2). However, the newly generated high-frequency components correspond to light propagating at large angles. Due to the finite NA of the network's convolutional units, some of these components fall outside the collection cone of subsequent layers (red dotted vertical lines) and cannot be captured. This excess bandwidth is inevitably lost, as illustrated by the red shaded





**Figure 3:** Energy loss in nonlinear networks. Illustration of wave propagation through a nonlinear network in the spatial frequency domain, plotted along one spatial dimension as a function of spatial frequency  $\nu$ . An imaging system with a low NA generates the low-resolution input image, effectively applying a low-pass filter  $H$ . The black dashed vertical lines mark the NA limit of this original system. As the wave propagates through the network, nonlinear interactions generate higher spatial frequencies, enabling partial reconstruction of previously lost high-frequency information, up to the NA defined by the collection cones of the convolutional units (red vertical dotted lines). Inevitably, this nonlinear spectral broadening also produces components beyond this NA, resulting in energy loss due to spectral content leaking outside the network's collection cone, highlighted by the red shaded regions. Consequently, the spectral intensity at low frequencies diminishes as the wave progresses through the network. At the final layer, the spectrum within the network's NA provides a scaled approximation of the original object, represented by the dashed gray line in all plots. The diagrams show the spatial frequency content at the frequency plane of each convolutional unit, just before its corresponding phase mask.

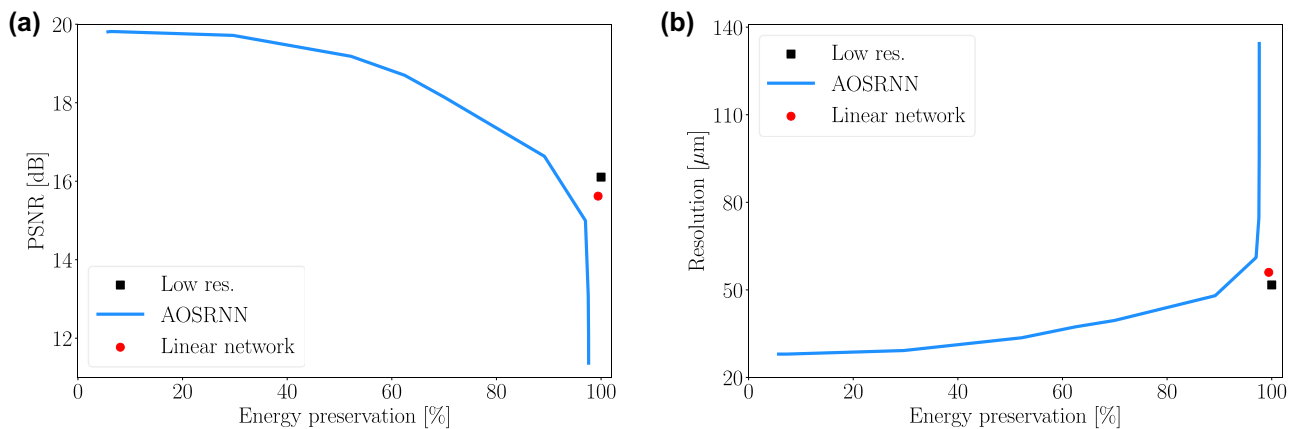
regions in Figure 3. As the field continues to propagate through the nonlinear network, each layer builds upon the previous one, reconstructing a progressively higher-resolution approximation of the object but at the cost of cumulative energy loss at every stage. A detailed frequency-domain derivation of this energy leakage is provided in Supplementary Note 5.

This inherent energy loss inevitably leads to output images with low total intensity. Such weak outputs are undesirable, as they are more susceptible to being overwhelmed by noise during the digital acquisition process. A simple approach to improve energy preservation is to amend the loss function of Eq. (2) with a regularization term that penalizes for low output intensities,

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left( \|\hat{I}_{\text{norm}}^{(i)} - I^{(i)}\|_1 - \gamma \|\hat{I}^{(i)}\|_1 \right), \quad (3)$$

where the regularization parameter  $\gamma$  controls the strength of the energy-preserving term. However, this solution turns out to introduce a strong tradeoff between energy preservation and performance, measured in terms of PSNR and resolution, as illustrated in Figure 4 for the MNSIT dataset. When the regularization parameter  $\gamma$  is very small, the network prioritizes image fidelity and achieves its best performance but preserves very little energy. As  $\gamma$  increases, the network retains more energy, but this comes at the cost of reduced reconstruction quality. For sufficiently large values of  $\gamma$ , the network's performance degrades to the point where it underperforms even the low-resolution input images, as indicated by the black squares in the figure.

A network that simply transmits low-resolution images to its output plane should, in principle, achieve performance metrics identical to those of the low-resolution inputs. Such a network would effectively reproduce the input field at the



**Figure 4:** Reconstruction fidelity vs. energy preservation. (a) PSNR in dB and (b) resolution in  $\mu\text{m}$  as a function of the fraction of energy preserved at the output (0–100 %). The blue line shows the results of AOSRNN, trained with 12 different values of the regularization parameter  $\gamma$  ranging from  $10^{-1}$  to  $10^{-9}$ . The black squares indicate the performance of the low-resolution input images. The red dots represent a linear network with the same architecture as AOSRNN but without the nonlinear layers.

output, without modification. However, this is not feasible with a nonlinear architecture, as the nonlinear layers inherently perturb the input during propagation. By removing these nonlinear elements from AOSRNN, the resulting linear network could, in theory, learn to perform this identity mapping. In practice, however, the optimization process rarely converges to this solution. Consequently, even the linear network introduces slight distortions to the input, as indicated by the red dots in Figure 4a and b.

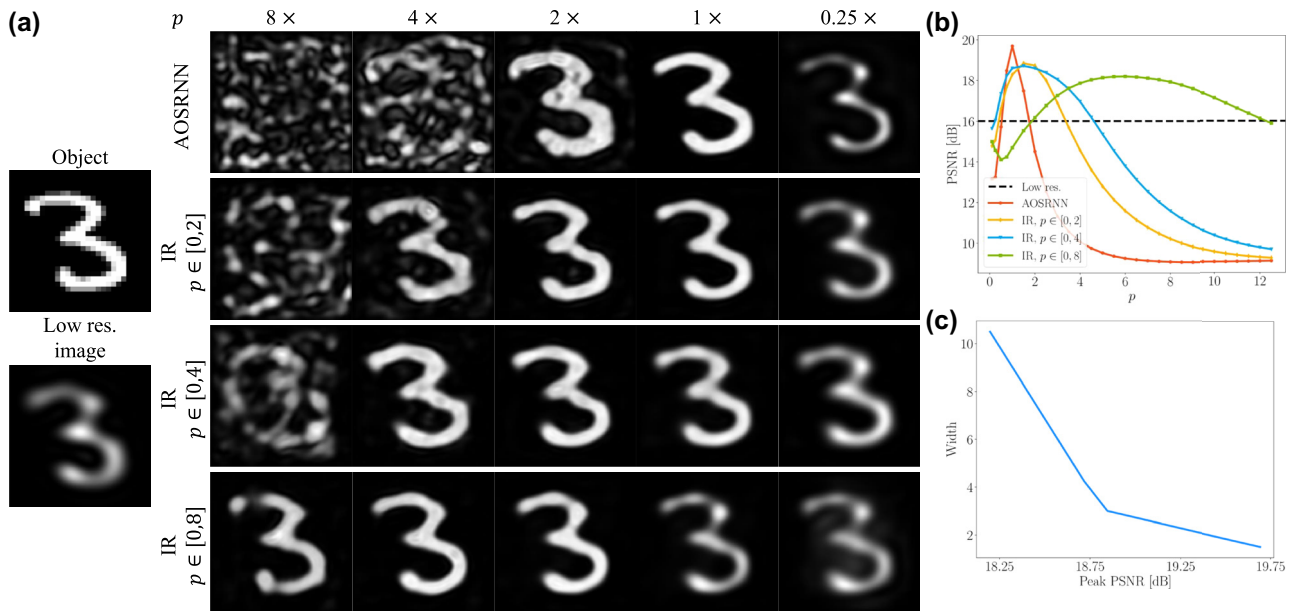
The above approach offers a simple way to improve the preservation of input energy at the cost of reducing reconstruction accuracy for a given network configuration. The tradeoff between energy preservation and reconstruction accuracy can be slightly improved by increasing the network's NA. For example, doubling the width and height of each layer in a network trained on the MNIST dataset without the energy preservation loss term improves energy preservation from  $\sim 6\%$  to  $\sim 10\%$  and reconstruction quality from 19.8 dB to 20.3 dB. However, increasing the physical size of the layers results in a larger system, which may be impractical in some applications. It may also introduce fabrication challenges that limit performance.

An alternative way to increase the NA is by reducing the focal length between the optical elements. This could help retain more input energy while also reducing the

system's physical footprint. In this context, optical metasurfaces offer a promising future direction. Their ability to precisely control local phase, amplitude, and polarization at subwavelength scales enables compact optical components with high NA [42]. As such, metasurfaces are an appealing platform for compact, energy-efficient optical neural networks and analog computing systems [43]–[45]. Notably, our modeling framework is general and directly applicable to metasurface-based implementations, since all optical components in the network are defined by their local phase profiles, regardless of physical realization.

### 3.2 Sensitivity to global intensity

Our second key observation relates to the sensitivity of the network to the global input intensity. In the digital domain, images are typically normalized to the range  $[0, 1]$ . However, in natural scenes, illumination intensity can vary across different scenarios. While this variation is generally not problematic for linear systems, where output intensities scale proportionally with the input, nonlinear systems behave quite differently. Even slight changes in input intensity can result in dramatically different outputs, as they shift the input field into different regimes of the system's nonlinearity.



**Figure 5:** Sensitivity to global intensity. (a) Representative output images from different networks evaluated at different  $p$  values. The standard AOSRNN fails to generalize to unseen intensity levels, whereas the IR networks produce high-quality reconstructions across varying  $p$ . (b) Reconstruction fidelity, quantified by PSNR in dB, as a function of the maximal image intensity  $p = \max_j |I^{(j)}|_\infty$ . The orange curve shows the performance of a standard AOSRNN, while the yellow, blue, and green curves show the results of an intensity-robust (IR) version of AOSRNN, each one trained with  $p$  values in different range. The black dashed line indicates the PSNR of the low-resolution input images. While the IR networks maintain consistent performance across a wide range of  $p$  values (denote as width), the regular network shows a sharp decline in performance when  $p \neq 1$ . (c) The blue curve denotes the inherent tradeoff between reconstruction fidelity (horizontal axis, measured by peak PSNR) and stability to different peak intensity values (vertical axis, measured by width).

Figure 5 illustrates the results of a network trained on the MNIST dataset with maximal intensity value,  $p = \max_j \|r^{(j)}\|_\infty$ , of 1, when evaluated on inputs with varying maximal intensities, ranging from 0.1 to 12.5. The results show significant degradation, both quantitatively and qualitatively, when the input intensity deviates from the training value. For example, increasing the intensity by a factor of 2 leads to a drop of approximately 5 dB in PSNR, as shown by the orange curve in Figure 5b. Qualitative deterioration is also evident in the first row of Figure 5a, where the network fails to generalize to inputs with  $p \neq 1$ .

A straightforward approach to accommodating a larger range of intensities is to expose the network to varying  $p$  values during training [46]. We trained three different networks, where the varying  $p$  values were in the range  $[0, 2]$ ,  $[0, 4]$ ,  $[0, 8]$ , indicated by the yellow, blue, and green curves in Figure 5b, respectively. Each of these intensity-robust (IR) networks demonstrate improved stability for different  $p$  values. Qualitative results for each of these networks are shown in Figure 5a. The IR networks improved stability can be seen by the increased range of input intensities over which their reconstruction quality exceeds the input quality, *i.e.*, the range of  $p$  values over which the different curves in Figure 5b exceed the black dashed line. We term this range width. However, a larger width comes at the expense of reconstruction fidelity, causing a drop in the peak PSNR achieved by each network. It can be seen in Figure 5b that the peak of the orange curve, representing a network trained with a single  $p$  value, is the highest. Figure 5c illustrates the inherent tradeoff between reconstruction fidelity (horizontal axis, measured by peak PSNR) and stability to different peak intensity values (vertical axis, measured by width), for passive all-optical nonlinear networks.

## 4 Discussion

We have demonstrated that passive all-optical neural networks are theoretically capable of achieving spatial super-resolution but suffer from two fundamental limitations: an inherent sensitivity to the input intensity and a tradeoff between energy preservation and performance. Although these sensitivities can be reduced by exposing the network to diverse conditions during training, this comes at the expense of deterioration in performance.

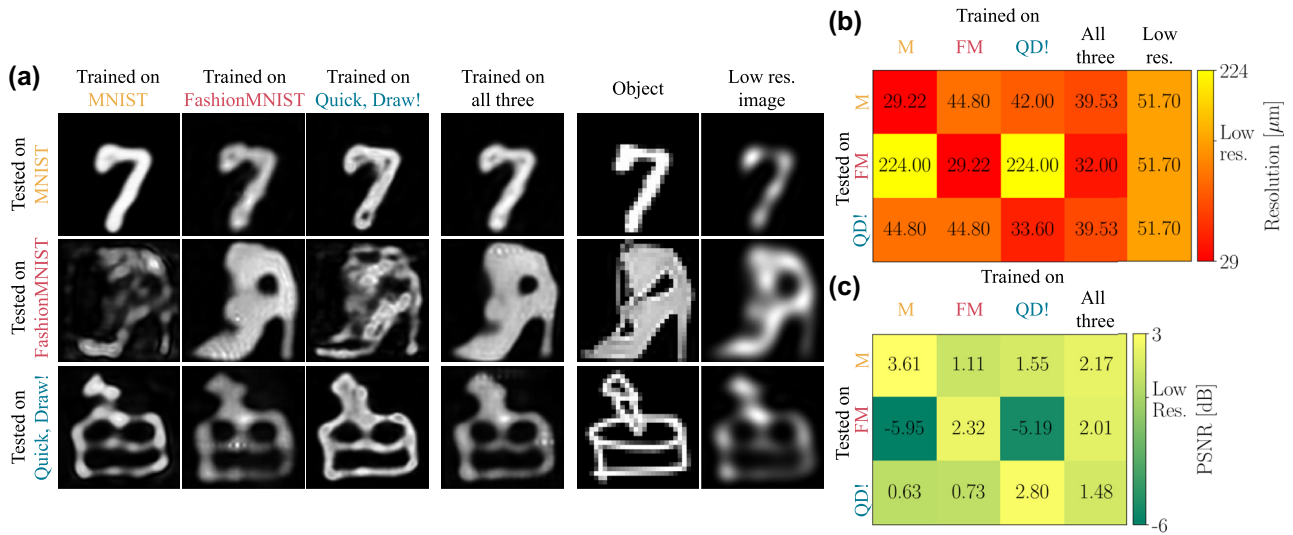
As mentioned in Section 2, our analysis, although independent from a specific choice of nonlinear layer, still assumes the availability of such a layer. Achieving an all-optical nonlinear response with sufficiently large  $n_2$  is a challenge in its own right [31]. In Supplementary Note 1,

we list various physical mechanisms that could be used to implement such nonlinear layers and discuss their shortcomings. Additionally, we demonstrate that AOSRNN can also be trained with a different nonlinear layer, a hybrid optoelectronic nonlinear device recently introduced by Zhang et al. [47].

It is instructive to note that, unlike conventional linear optical systems, diffractive networks are also fundamentally restricted by their training data, as they adapt to its underlying statistics. As a result, a network trained on one dataset may generalize poorly to images from another dataset. Figure 6 illustrates this effect using the MNIST, FashionMNIST, and Quick, Draw! datasets (see Supplementary Note 6 for more datasets). As shown, performance on out-of-distribution data highly depends on the degree of distribution shift. For example, networks trained on MNIST or Quick, Draw! datasets fail to generalize to the FashionMNIST dataset, which is characterized by larger smooth regions. In these cases, the results are even worse than the input low-resolution images. Figure 6 further shows the performance of a network trained on all three datasets. When tested on a specific dataset, this network achieves improved results with respect to the networks that were not trained on that dataset but slightly worse results than the network trained on that particular dataset. This highlights yet another fundamental tradeoff in all-optical super-resolution – the more diverse the data on which the network is trained, the smaller the resolution improvement. Supplementary Note 6 provides additional details. Supplementary Note 7 illustrates the cost of achieving immunity to all the discussed effects simultaneously – sensitivity to global intensity, to energy preservation, and to out-of-distribution data.

While this work primarily addresses all-optical spatial super-resolution, the tradeoffs we identify are inherently linked to the properties of light and may thus arise in different tasks as well. For example, when training our AOSRNN on image classification task, it achieves 94.8 % test classification accuracy on the MNIST dataset. However, this performance is achieved by scattering a significant portion of the incoming light. Additionally, this network is sensitive to the global input intensity, experiencing a 33 % drop in classification accuracy when increasing the input intensity by a factor of 2 (see Supplementary Note 8 for further details).

The fundamental limitations of energy-preservation versus resolution and sensitivity to global intensity are also not unique to a specific network architecture or to the choice of nonlinear layers but are instead intrinsic to passive nonlinear optical networks. This is illustrated



**Figure 6:** AOSRNN results on out-of-distribution data. (a) Qualitative results of networks trained using the MNIST, FashionMNIST, and Quick, Draw! datasets, as well as a network trained using all these three datasets. These networks were then evaluated on the MNIST, FashionMNIST, and Quick, Draw! datasets. All the results are given next to the objects and the low-resolution images. (b) Resolution results of the different networks when evaluated on the different datasets. (c) PSNR results of all networks w.r.t the low resolution images, when evaluated on the different datasets.

in Supplementary Note 9, where a nonlinear passive network with a different architecture [34] is trained to perform super-resolution. This network, which comprises of a sequence of diffractive layers interleaved by nonlinear layers (without  $4f$  systems), successfully enhances the resolution of the input images, achieving a PSNR of 17.4 dB for input images with a PSNR of 16.1 dB. However, it does so by preserving only  $\sim 1\%$  of the energy in its output plane. Furthermore, when multiplying the input intensity by 2 or by 4, the achieved PSNR deteriorates to 16.6 dB and 14.2 dB, respectively (see Supplementary Note 9 for further details).

Finally, the fundamental limitations of energy-preservation versus resolution and sensitivity to global intensity are not unique to monochromatic light. This is demonstrated in Supplementary Note 10, where we trained an AOSRNN for objects illuminated by polychromatic light comprising three different wavelengths –  $\{400, 550, 700\}$  nm. This network, again, successfully enhances the input image resolution, yet, still suffers from the identified tradeoff (see Supplementary Note 10 for further details.)

These analyses highlight the need for future designs to carefully consider these tradeoffs and identify operating points that are well suited to the task and deployment context. By exposing and quantifying these limitations, our work marks an important step toward enabling more robust development of passive all-optical neural systems and lays out the key considerations for implementing all-optical super-resolution neural networks.

## 5 Methods

For training AOSRNN, we used a dataset  $\mathcal{D}$  of objects (images),  $\mathcal{D} = \{I^{(i)}\}_{i=1}^N$ . The input to the network was the low-resolution images of  $I^{(i)}$ , resulting from the imaging optics. We implemented the imaging optics as ideal filtering in the frequency domain [1].

We used the MNIST [38], FashionMNIST [39], and Quick, Draw! [40] datasets, which contain  $28 \times 28$  grayscale images of handwritten digits, fashion items, and drawing of different everyday objects, respectively. In all of the numerical experiments, we first upsampled these images using nearest-neighbor interpolation to  $112 \times 112$  grayscale images. The input images were padded accordingly to prevent aliasing. The MNIST and FashionMNIST datasets include 10 different classes, while the Quick, Draw! dataset includes draws of 345 different classes. For the numerical experiments, we used 10 out of these 345 classes. From the Quick, Draw! dataset, we used the classes of aircraft carrier, banana, cake, diamond, fish, guitar, hexagon, key, microphone, and saxophone.

We trained the different networks with 10,000 training images ( $\sim 1,000$  per class) for 1,000 epochs, using the Adam optimizer [48] with a learning rate of  $\eta = 1 \times 10^{-1}$ . We used a scheduler that reduces the learning rate by a factor of 2 every 100 epochs. The simulations were implemented using the Pytorch deep learning framework [49] and were executed on a Linux machine with NVIDIA GeForce GTX 2080 Ti GPU. Standard training took  $\sim 7$  h on this machine.



The intensity robust (IR) networks were trained using the same hyperparameters and training process but with maximal intensity values chosen uniformly from the predefined set [0.1, 0.25, 0.5, 0.75, 1, 2, 3, ...]. The largest value of this set was either 2, 4, or 8, for the different networks.

For details regarding the forward propagation through AOSRNN, see Supplementary Note 11.

**Research funding:** None declared.

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and consented to its submission to the journal, reviewed all the results, and approved the final version of the manuscript.

**Conflict of interest:** Authors state no conflict of interest.

**Data availability:** The code used in the current study is available at ref. [50]. The datasets analyzed during the current study were derived from these public domain dataset: MNIST dataset – <https://docs.pytorch.org/vision/main/generated/torchvision.datasets.MNIST.html>, FashionMNIST dataset – <https://github.com/zalandoresearch/fashion-mnist>, Quick, Draw! dataset – <https://github.com/googlecreativelab/quickdraw-dataset>, KMNIST dataset – <https://github.com/rois-codh/kmnist>, EMNIST dataset – <https://www.nist.gov/itl/products-and-services/emnist-dataset>.

## References

- [1] J. W. Goodman, *Introduction to Fourier Optics*, Englewood, Colorado, Roberts and Company Publishers, 2005.
- [2] S. W. Hell, "Far-field optical nanoscopy," *Science*, vol. 316, no. 5828, pp. 1153–1158, 2007.
- [3] A. N. Kellerer, "Beating the diffraction limit in astronomy via quantum cloning (corrigendum)," *Astron. Astrophys.*, vol. 582, 2015, <https://doi.org/10.1051/0004-6361/201322665e>.
- [4] N. Fang, H. Lee, C. Sun, and X. Zhang, "Sub-diffraction-limited optical imaging with a silver superlens," *Science*, vol. 308, no. 5721, pp. 534–537, 2005.
- [5] D. O. Melville and R. J. Blaikie, "Super-resolution imaging through a planar silver layer," *Opt. Express*, vol. 13, no. 6, pp. 2127–2134, 2005.
- [6] M. A. Neil, R. Juškaitis, and T. Wilson, "Method of obtaining optical sectioning by using structured light in a conventional microscope," *Opt. Lett.*, vol. 22, no. 24, pp. 1905–1907, 1997.
- [7] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *J. Microsc.*, vol. 198, no. 2, pp. 82–87, 2000.
- [8] E. Betzig, et al., "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.
- [9] S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.*, vol. 91, no. 11, pp. 4258–4272, 2006.
- [10] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nat. Methods*, vol. 3, no. 10, pp. 793–796, 2006.
- [11] D. W. Pohl, W. Denk, and M. Lanz, "Optical stethoscopy: image recording with resolution  $\lambda/20$ ," *Appl. Phys. Lett.*, vol. 44, no. 7, pp. 651–653, 1984.
- [12] F. M. Huang, Y. Chen, F. J. G. De Abajo, and N. I. Zheludev, "Optical super-resolution through super-oscillations," *J. Opt. A: Pure Appl. Opt.*, vol. 9, no. 9, p. S285, 2007.
- [13] E. T. Rogers and N. I. Zheludev, "Optical super-oscillations: sub-wavelength light focusing and super-resolution imaging," *J. Opt.*, vol. 15, no. 9, p. 094008, 2013.
- [14] Y. Choi, et al., "Overcoming the diffraction limit using multiple light scattering in a highly disordered medium," *Phys. Rev. Lett.*, vol. 107, no. 2, p. 023902, 2011.
- [15] H. Yilmaz, E. G. van Putten, J. Bertolotti, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Speckle correlation resolution enhancement of wide-field fluorescence imaging," *Optica*, vol. 2, no. 5, pp. 424–429, 2015.
- [16] E. Edrei and G. Scarcelli, "Memory-effect based deconvolution microscopy for super-resolution imaging through scattering media," *Sci. Rep.*, vol. 6, no. 1, p. 33558, 2016.
- [17] D. Wang, S. K. Sahoo, X. Zhu, G. Adamo, and C. Dang, "Non-invasive super-resolution imaging through dynamic scattering media," *Nat. Commun.*, vol. 12, no. 1, p. 3150, 2021.
- [18] 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. 458–464, 2018.
- [19] Q.-Q. Shan, et al., "Galaxy image super-resolution reconstruction using diffusion network," *Eng. Appl. Artif. Intell.*, vol. 142, p. 109836, 2025.
- [20] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2015.
- [21] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.
- [22] C. Ledig, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.
- [23] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 pirm challenge on perceptual image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018.
- [24] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286–301.
- [25] X. Wang, et al., "Esrgan: enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018.
- [26] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: image restoration using swin transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1833–1844.
- [27] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am.*, vol. 62, no. 1, pp. 55–59, 1972.

- [28] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, no. 6, p. 745, 1974.
- [29] V. V. R. Bukka, M. Xu, M. Andrew, and A. Andreyev, "Assessment of deep-learning-based resolution recovery algorithm relative to imaging system resolution and feature size," *Methods Microsc.*, vol. 2, no. 2, pp. 183–201, 2025.
- [30] X. Xu, *et al.*, "Scaling for edge inference of deep neural networks," *Nat. Electron.*, vol. 1, no. 4, pp. 216–222, 2018.
- [31] G. Wetzstein, *et al.*, "Inference in artificial intelligence with deep optics and photonics," *Nature*, vol. 588, no. 7836, pp. 39–47, 2020.
- [32] P. N. Michelini, Y. Lu, and X. Jiang, "edge-sr: super-resolution for the masses," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1078–1087.
- [33] Y. Shen, *et al.*, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics*, vol. 11, no. 7, pp. 441–446, 2017.
- [34] X. Lin, *et al.*, "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018.
- [35] B. J. Shastri, *et al.*, "Photonics for artificial intelligence and neuromorphic computing," *Nat. Photonics*, vol. 15, no. 2, pp. 102–114, 2021.
- [36] P. L. McMahon, "The physics of optical computing," *Nat. Rev. Phys.*, vol. 5, no. 12, pp. 717–734, 2023.
- [37] D. N. Christodoulides, I. C. Khoo, G. J. Salamo, G. I. Stegeman, and E. W. Van Stryland, "Nonlinear refraction and absorption: mechanisms and magnitudes," *Adv. Opt. Photonics*, vol. 2, no. 1, pp. 60–200, 2010.
- [38] Y. LeCun, "The mnist database of handwritten digits," 1998. Available at: <https://yann.lecun.com/exdb/mnist/>.
- [39] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.
- [40] J. Jongejan, H. Rowley, T. Kawashima, J. Kim, and N. Fox-Gieg, "The quick, draw!-ai experiment," 2016. Available at: <https://quickdraw.withgoogle.com/data>.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [42] A. I. Kuznetsov, *et al.*, "Roadmap for optical metasurfaces," *ACS Photonics*, vol. 11, no. 3, pp. 816–865, 2024.
- [43] C. Qian, *et al.*, "Performing optical logic operations by a diffractive neural network," *Light Sci. Appl.*, vol. 9, no. 1, 2020, <https://doi.org/10.1038/s41377-020-0303-2>.
- [44] C. Liu, *et al.*, "A programmable diffractive deep neural network based on a digital-coding metasurface array," *Nat. Electron.*, vol. 5, no. 2, pp. 113–122, 2022.
- [45] T. Badloe, S. Lee, and J. Rho, "Computation at the speed of light: metamaterials for all-optical calculations and neural networks," *Adv. Photonics*, vol. 4, no. 6, p. 064002, 2022.
- [46] M. Kleiner, L. Michaeli, and T. Michaeli, "Coherence awareness in diffractive neural networks," *Laser Photonics Rev.*, vol. 19, no. 10, p. 2401299, 2025.
- [47] D. Zhang, *et al.*, "Broadband nonlinear modulation of incoherent light using a transparent optoelectronic neuron array," *Nat. Commun.*, vol. 15, no. 1, p. 2433, 2024.
- [48] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *International Conference on Learning Representations*, vol. 3, 2015.
- [49] A. Paszke, *et al.*, "Pytorch: an imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.
- [50] M. Kleiner, L. Michaeli, and T. Michaeli, "Github reposetery — All Optical Super Resolution," 2025. Available at: <https://github.com/matankleiner/All-Optical-SR>.

---

**Supplementary Material:** This article contains supplementary material (<https://doi.org/10.1515/nanoph-2025-0294>).