DOI: 10.14529/jsfi260208

An Analog Photonic Computing Device Based on a Diffractive Neural Network for Video Stream Processing

Roman V. Skidanov[1], ***Leonid L. Doskolovich***[1], ***Nikolay L. Kazanskiy***[1], ***Alexandr E. Morozov***[1], ***Alexey S. Pronin***[1], ***Daniil M. Sorokin***[1], ***Yuriy V. Khanenko***[1], ***Victor A. Soifer***[1]

This paper investigates the design principles of analog photonic computing devices for pattern recognition tasks. As a result of the conducted research, an analog photonic computing device (APCD) based on a diffractive neural network (DNN) is developed and implemented in two configurations. The DNN is realized using a phase spatial light modulator (SLM) placed in the Fourier plane of a two-lens optical system (4F system). Benefits of combining an optical DNN with a compact computer neural network for post-processing the optical recognition results are demonstrated. The presented results show the effectiveness of using the APCD for processing and recognition of high-dimensional imagery in real-time video streams. A comparative performance analysis of the APCD and a family of GPU-based YOLO neural networks is conducted for object image recognition in video streams. The results demonstrate that when processing high-dimensional frames, the power consumption of the APCD is an order of magnitude lower than that of modern GPU-class graphics platforms such as the RTX 4090.

Keywords: optical computing, diffractive neural network, image processing and classification, spatial light modulator, YOLO neural networks, video-streams.

Introduction

Artificial neural networks have achieved significant success in a wide range of machine learning tasks, including image and video classification and analysis, medical diagnostics, speech recognition, solving various inverse problems, and so on. At the same time, the training and inference processes of neural networks used, for example, in modern models of video imagery recognition and analysis impose enormous demands on digital processors in terms of computational volume and speed, which cannot always be met. These constraints have driven the rapid advancement of a research field focused on the optical implementation of neural networks [1–3]. It should be noted that this line of research lies within the general trend of augmenting or substituting digital computing devices with their optical analogs, which are expected to overcome the limitations of digital computing systems related to insufficient computational speed and high energy consumption [4]. As shoun in Fig. 1, interest in this research area has grown exponentially since 2017 The data presented in demonstrate an exponential growth in interest in this research direction since 2017.

Among optical neural networks, particular interest is attracted by so-called diffractive neural networks (DNNs), which are realized as a cascade of phase diffractive optical elements (DOEs) [5–16], (see Fig. 2). It is important to note that the research field concerned with the design of DOEs – including cascaded DOEs – for solving various problems of laser light transformation and focusing has a long history and has been actively evolving for over 50 years [17–24]. At the same time, the use of cascaded DOEs to optically solve machine learning problems was first demonstrated only in 2018 in the seminal work [5]. In that work, several analogies were drawn between a cascade of DOEs and an artificial neural network. Specifically, it was noted that, by the Huygens-Fresnel principle, each pixel of a DOE emits a spherical wave that illumi-

[1]S.P. Korolyov Samara National Research University, Samara, Russian Federation

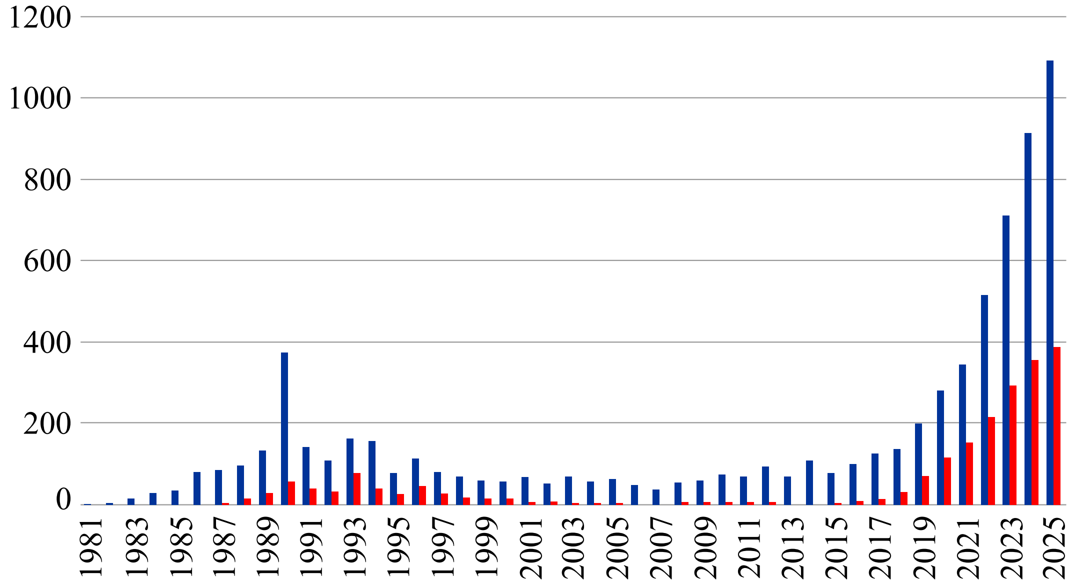


Figure 1. Annual number of publications in the field of optical computing. Blue: total publications in the Scopus database; Red: including publications on the subject of optical neural networks

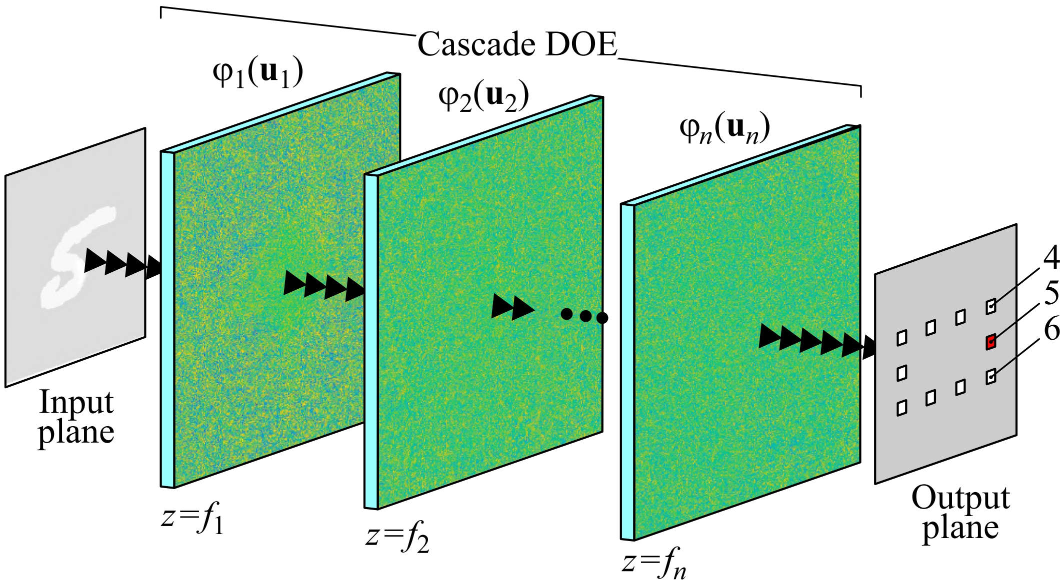


Figure 2. Geometry of a DNN implemented as a cascade of DOEs for solving a classification problem

nates the pixels of the subsequent DOE in the cascade, thereby mimicking the interconnections between neurons in a multilayer perceptron. On this basis, the term "diffractive neural network" was introduced, and the possibility of optically solving classification problems using cascaded DOEs was demonstrated both theoretically and experimentally.

The optical solution to the classification problem is schematically shown in Fig. 2. We assume that optical images belonging to N distinct object classes – for instance, images of handwritten digits – are generated in an input plane by means of an SLM. These images propagate through the optical system (a DOE cascade) towards the output plane. In the output plane, a set of energy detection regions are specified, with the number of regions corresponding to the number of input image classes. In a particular case, the registration regions can be implemented as an array of photodetectors that measure the total energy of the incident optical signal. The number of the region (photodetector) with the highest energy then determines the class of the input image. By way of example, Fig. 2 schematically illustrates that for an input image of the digit '5', the maximum energy is attained in the region associated with that digit. The above-described operation of the DNN presupposes its prior 'training', which involves solving the inverse problem of determining the phase function of the constituent DOEs such that the desired operational condition is satisfied (i.e., forming the maximum energy in the registration region that corresponds to the class of the input object). Typically, the afore-mentioned inverse problem is solved in the framework of scalar diffraction theory under the thin-element approximation. The phase functions of the DOEs are calculated using a gradient-based method within the artificial neural network methodology, utilizing a training dataset (a set of characteristic input images) [5–16].

It is important to note that in the DNN shown in Fig. 2, the relationship between the fields in the input and output planes is described by a linear operator. Considering that a multilayer neural network without nonlinearities is equivalent to a single-layer linear neural network, the use of a cascade of DOEs instead of a single DOE is not always justified. In particular, the authors of [25] showed that with properly chosen parameters, a single DOE can solve classification tasks with high accuracy, not only matching but even exceeding the classification accuracy achieved with complex cascaded DOEs. Moreover, it is well known that in the physical realization of a DNN designed to operate with visible light, a major challenge is the stringent requirement for precise positioning of the constituent DOEs [10, 12, 15, 16]. Moreover, cascaded DOEs prove to be substantially more sensitive to misalignment than single DOEs, so that with positioning errors of only 10–15 wavelengths, they become inferior to single DOEs in terms of operational performance [15, 16]. In this regard, the present authors consider DNNs implemented as a single DOE as the most suitable option for practical applications. Furthermore, studies conducted by the present authors (not reported in this paper) have shown that the variant most tolerant to positioning errors is the DNN realization in a 4F-system configuration (see Fig. 3), in which a single DOE is placed between two lenses that perform Fourier transforms. This particular DNN configuration is discussed further in this paper.

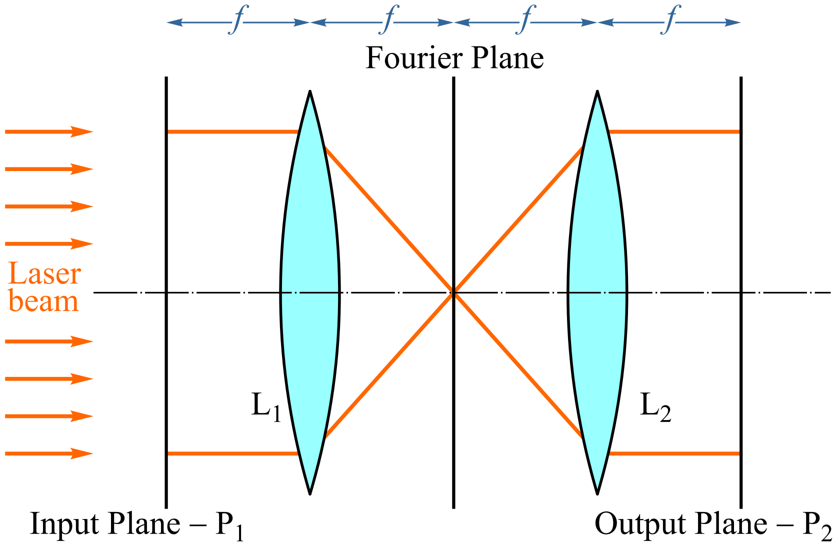


Figure 3. A 4F-system [26] (f are focal lengths of lenses L_1 and L_2)

In this work, we discuss two versions of an analog photonic computing device (APCD) employing a DNN in a 4F-system configuration. These devices were developed at S. P. Korolyov Samara University under a contract with RFNC-VNIIEF in 2024 and 2025. The APCD uses two UPOLabs HDSLM36R liquid-crystal SLMs with a resolution of 4096×2140 and a maximum frame rate of 60 Hz. One modulator serves for image input, whereas the other realizes a diffractive neural network as a single phase-only DOE positioned in the frequency plane. The results of the processing are captured using two DAHENG ME2P-1231-32U3M cameras with a resolution of 4096×3000 and a maximum frame rate of 60 Hz. A more detailed description of the APCDs is provided in Section 1. The methods for designing DOEs employed in the APCD are set forth in [15, 16].

1. APCD-2024 and APCD-2025

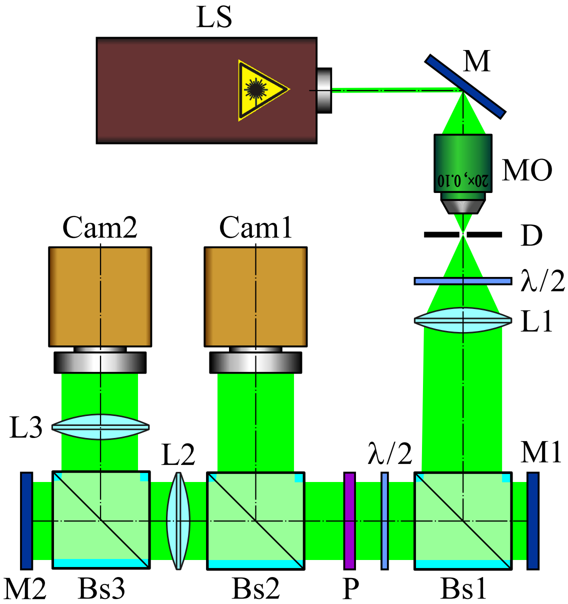
In 2024 and 2025, two experimental APCD prototypes based on a 4F-system were designed and studied.

1.1. APCD-2024

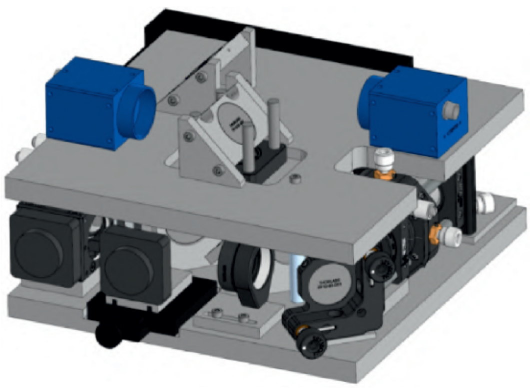
Figure 4a depicts a schematic optical arrangement of the developed APCD. The linearly polarized output beam from a solid-state laser LS (LSR532HX, $\lambda = 532$ nm, maximum output power 2.5 W) was converted into a near-plane-wave distribution by using a collimation system comprising a microscope objective MO (20×) and a pinhole D with a 10-μm diameter. The plane wave passed through a half-wave plate (denoted $\lambda/2$ in Fig. 4a), which served to rotate the linear polarization plane of the laser beam. This linearly polarized beam was then directed via a beam splitting cube Bs1 onto a reflective SLM M1 (SPSLM36R, 4096×2240 pixels, pixel size 3.6 μm). To perform amplitude modulation, a polarizer P was placed behind the SLM. A combination of lens L2 (focal length 150 mm) and reflective SLM M2 (SPSLM36R) was utilized to implement a 4F optical correlator system, wherein the phase mask of the DNN, displayed on modulator M2, served as the spatial filter. We note that the beam splitting cube Bs2 served as an extra correlator component to observe the reflected modulated light field with video camera Cam1 (Daheng ME2P-1231-32U3M, 4096×3000 pixel array, pixel size 3.45 μm). Light splitting cube BS3 combined with lens L3 (focal length 75 mm) enabled observation of the intensity distribution of the optical field generated in the display plane of the reflective modulator M2. Here, the lens distance to both the modulator display plane and the video-camera Cam 2 sensor (Daheng ME2P-1231-32U3M, sensor size 4096×3000, pixel size 3.45 μm) was equal to double the focal length of the lens. In this case, the actually recorded intensity corresponded to the optical field spectrum generated in the display plane of the first light modulator M1. The experimentally realized device (Fig. 4b and 4c) basically corresponds to the optical setup in Fig. 4a; however, to minimize the system's footprint, a large number of rotating mirrors were utilized, and the beam splitter cubes Bs2 and Bs3 were replaced with semi-transparent mirrors. To further reduce the device footprint, a two-tier optical setup was realized, with the second tier containing both cameras (Fig. 4b and 4c). A 3D model of the APCD assembly is shown in Fig. 4b and a APCD photograph is shown in Fig. 4c.

Camera Cam 2 is intended for capturing the intensity distribution in the Fourier plane. Data from this distribution allows implementation of an electronically controlled nonlinear operational mode of the APCD-2024, with the transmittance of each pixel varying nonlinearly as a function of the detected intensity level. Nonlinearity is introduced into the modulator M2 via correcting the DNN phase function.

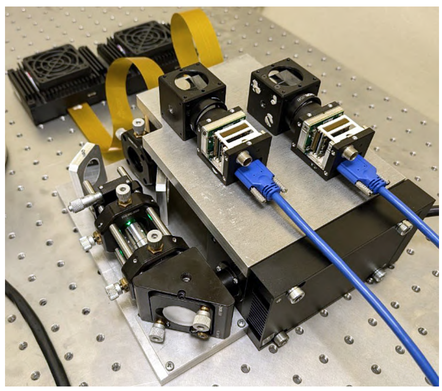
The APCD-2024 setup mounted in a compact housing demonstrated a fairly good performance: in an experiment on recognition of handwritten digits from the MNIST dataset [27], which comprises 60,000 training images and 10,000 test images, a recognition error rate of 3.14% was achieved (corresponding to an accuracy of 100% – 3.14% = 96.86%). According to the original website of MNIST creators [28], error rates of 12% are observed when using simple linear classifiers without preprocessing [29].



(a) Schematic optical layout: LS – a 532-nm laser, M – a rotating mirror, MO – a 20× microlens, D – a pinhole, $\lambda/2$ – half-wave plates, L1 – a collimating lens, Bs1, Bs2, Bs3 – beam splitters, M1 – an input SLM, P – a polarizer, L2 – a lens performing a Fourier transform, L3 – an imaging lens generating an intensity pattern in the Fourier plane, Cam1 – an output camera, and Cam2 – a camera capturing the intensity pattern in the Fourier plane



(b) A 3D model of the assembled APCD



(c) An APCD photograph

Figure 4. A 4F-system-enabled APCD

The two-tier arrangement suffers from several drawbacks: access to the components on the first tier for alignment is only possible after removal of the second tier. The high density of heat-dissipating elements within a confined volume induces distortions in the optical layout. Furthermore, the forced-air cooling system calls for frequent cleaning of the optical surface.

In addition, the optical scheme depicted in Fig. 4a proved to be suboptimal in terms of the position of beam splitter Bs3, thus leading to extra light losses.

In 2025, a new version of the APCD was developed, in which the optical components and electronic modules were integrated into a standard form-factor housing designed for installation in a server rake.

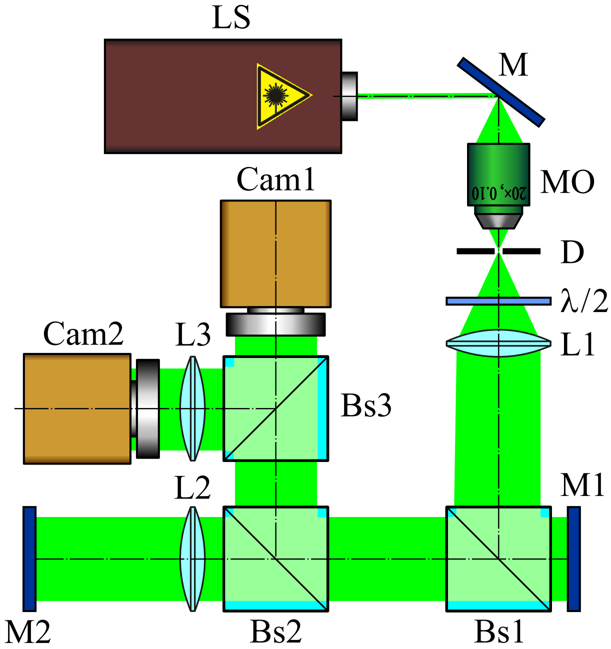
1.2. APCD-2025

Two main modifications were introduced into the APCD-2025 optical scheme. The first modification consists in image feeding via phase – rather than amplitude – modulation of the electromagnetic wave. All known approaches – starting from the classic work [26] to recent publications [30–32] – have relied on amplitude-based input of the image under recognition. This stems from the inertia of human visual perception, as the eye perceives the intensity distribution, which is determined by the light amplitude. The use of phase-modulated input nade it possible to exclude the half-wave plate located behind the beam splitter and the polarizer from the optical scheme. The second modification affected the beam splitter Bs3 deflecting light to the camera Cam2. With the light passing through the beam splitter twice, it was moved to the arm of back reflection from beam splitter Bs2. Figure 5a shows the new optical layout of the APCD-2025, while Fig. 5b presents a photograph of the APCD-2025 in a housing intended for mounting in a server rack.

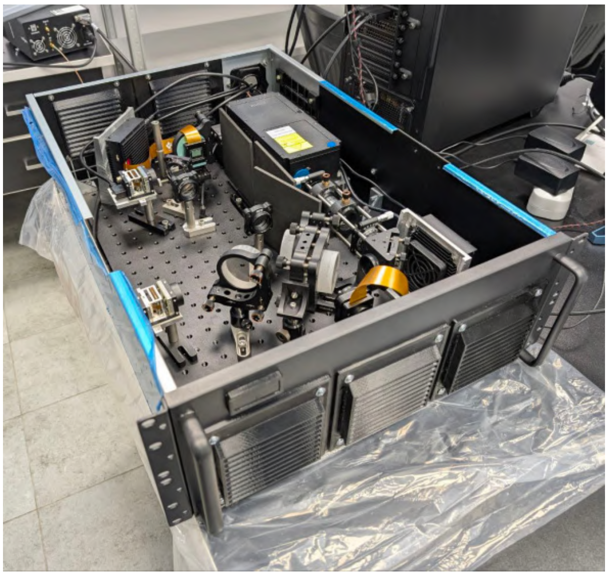
As seen from Fig. 5, the beam splitter Bs3 is relocated from modulator M2 closer to cameras Cam1 and Cam2, enabling a single passage of light through the beam splitter, so reducing energy losses in the element by more than twice (minus one beam splitting and two Fresnel reflections). With the half-wave plate and polarizer not found behind the beam splitter Bs1, the modulator M1 can be used in a regular mode of phase modulation. Another benefit is an essential reduction in light energy losses. A polarizer typically transmits only half of the incident power and about 4% is additionally reflected at each surface of the two elements. Thus, their combined transmittance was $0.5*0.96^4=0.42$, so that their removal from the optical scheme reduced energy losses by nearly a factor of 2.5. Overall, by eliminating both the half-wave plate and the polarizer, reduced the energy losses by approximately a factor of 5, allowing a five-fold reduction in the laser power in the APCD. The total energy consumption decreased from 72 W to 56 W.

2. Experiments on Hand-Wwritten Digit Recognition Using the APCD-2024

In image recognition (classification) tasks, a sequence of images from several distinct classes is fed to the input of the APCD-2024 (with image dimensions up to 2048×2048 pixels). Typically, handwritten digit images from the MNIST database [29] are used as a standard test [5–16]. In accordance with the description of the classification task in the Introduction, several energy registration zones are defined at the output plane of the optical system, with the zones used in the experiment shown in Fig. 6a. The diffractive neural network realized as a phase-only DOE (the DOE design methods employed are described in detail in [15, 16]) on phase modulator M2 is required to generate the maximum energy in the detection zone whose index matches the class of



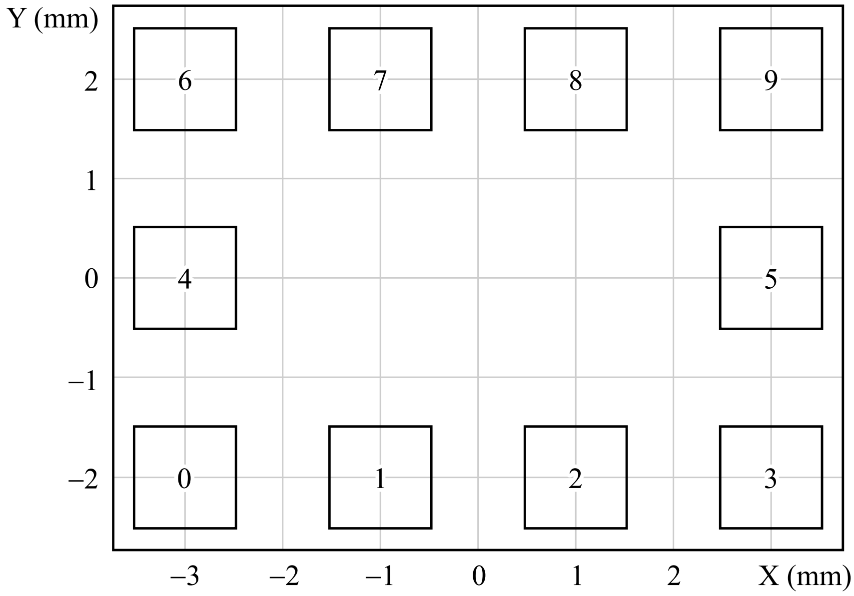
(a) Optical layout of the APCD-2025: LS – a 532-nm laser, M – a rotating mirror, MO – a 20× microlens, D – a pinhole, $\lambda/2$ – a half-wave plate, L1 – a collimating lens, Bs1, Bs2, Bs3 – beam splitters, M1 – an input modulator, L2 – a Fourier transforming lens, M2 – a modulator in the Fourier plane, L3 – a lens for imaging the Fourier-plane intensity distribution, Cam1 – an output camera, Cam2 – a camera for capturing the Fourier-plane intensity distribution



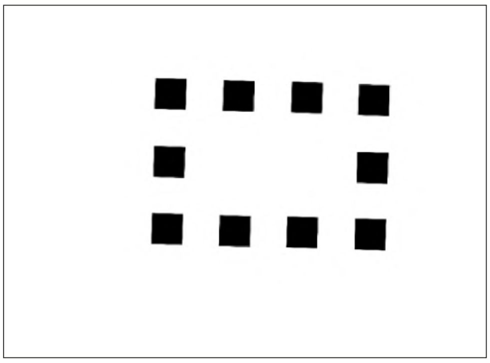
(b) A photograph of the APCD in the housing

Figure 5. APCD-2025

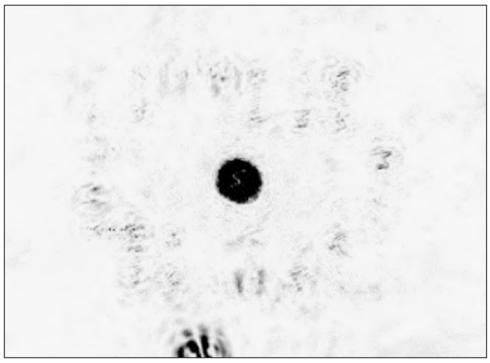
the input image. In the experimental setup, it is hardly possible to ensure the exact alignment of the modulator in the Fourier plane as well as precise camera positioning, so the positions of the registration zones are determined experimentally by means of a special phase mask uniformly illuminating each zone. Figure 6b shows actual positions of the detection zones within the full field of view of camera Cam1, while Fig. 6c represents the actual intensity distribution when an image of the digit 5 is fed to the input.



(a) Layout of the detection zones of the diffractive neural network (diagram)



(b) Actual positions of the zones in the field of view of camera Cam1



(c) Output intensity pattern when an image of 5 is fed to the input

Figure 6. Layout of the detection zones of the diffractive neural network

A series of experiments on handwritten digit recognition and classification using the MNIST database was conducted with the APCD-2024. In the experiments, a sequence of 5,000 digit images was loaded into the input SLM. The resulting intensity pattern across 10 detection zones at the optical system output plane was recorded and classification outcomes were subsequently determined. The classification accuracy achieved did not exceed 85%.

The relatively low classification accuracy can be attributed to several shortcomings of the optical system:

1. The wavefront from the collimator deviates from an ideal plane wave;
2. There are spurious reflections from numerous optical surfaces;
3. The effective active area of the modulators is only 80% of their total active surface area;
4. A significant proportion of the incident light is directed to the zero diffraction order (at the center of Fig. 6c);
5. Geometric inaccuracies result in errors when locating the detection zones.

A technique for reducing the recognition error that eliminates all systematic inaccuracies of the optical system was proposed. To achieve this, every other image in the sequence should be formed as a background image. Then, the intensity pattern from the object under recognition is analyzed and the intensity distribution corresponding to the background input image is subtracted from it. Using this technique, the classification accuracy reached 96% in the very first experiments. Unfortunately, this approach reduced the APCD's performance by half. Therefore, to compensate for the influence of systematic errors on the classification results, a tiny digital neural network – a linear perceptron digitally implemented on a computer – was introduced to post-process the recognition results. A similar approach was discussed in [33, 34].

The input layer of the digital neural network was formed by the integrated intensities measured in the registration zones. The hidden layer comprised 10 neurons and the output layer also comprised 10 neurons, associated with the classes of the input objects. Training of this neural network was performed using the output images generated by the APCD. From each image, between 10 and 90 numerical features were extracted – namely, the total intensities within each registration zone, or within subdivisions of the zones into 4 and 9 subregions. These features were computed for 50,000 handwritten digit images taken from the MNIST database. The neural network employed a linear neuron activation function. At the output layer, the computer neural network returned the object class, enhancing recognition accuracy. Table 1 presents the experimental recognition results obtained using the APCD-2024 augmented with the above-described neural network, evaluated on 5,000 characters from the MNIST test set.

The first column of Tab. 1 specifies the character presented at the APCD-2024 input, and the rows indicate the counts of correct and incorrect recognition outcomes. In the confusion matrix shown in Tab. 1, the rows represent the true classes, while the columns give the predicted classes. Recognition accuracy is computed as the sum of all values on the diagonal from the top-left to the bottom-right corner (highlighted in green in Tab. 1) divided by the total number of digits under recognition (5,000): $(476 + 552 + 475 + 477 + 528 + 418 + 490 + 535 + 449 + 443)/5000 = 0.9686$. In the case of perfect system performance, the diagonal of the table should contain the largest values, while all off-diagonal entries should be zero.

Table 1. Recognition results for 5,000 handwritten characters from the MNIST database obtained with the APCD-2024 augmented with a supplementary neural network with 90 input neurons

	0	1	2	3	4	5	6	7	8	9
0	476	0	0	0	0	1	1	1	0	2
1	0	552	4	2	1	0	0	1	3	0
2	0	1	475	3	1	2	0	1	4	1
3	1	1	2	477	0	3	0	3	4	2
4	0	0	0	0	528	1	1	1	0	4
5	0	1	1	5	0	418	3	1	4	1
6	0	0	1	0	6	4	490	0	1	0
7	0	1	3	3	2	0	0	535	1	5
8	0	1	1	3	2	3	0	0	449	3
9	4	2	1	4	15	2	1	17	6	443

The confusion matrix presented in Tab. 1 enables computation of the primary classification quality metrics for the digits. For instance, the precision (which indicates the proportion of digits predicted as class "X" that truly belong to class "X") for the digit "5" is 418/(0+1+1+5+0+418+3+1+4+1)=0.9631. The recall – representing the proportion of true class "X" digits that the classifier successfully identifies – for the digit "5" equals 418/(1+0+2+3+1+418+4+0+3+2)=0.9631. The F1-score, defined as the harmonic mean of precision and recall and considered the most informative metric for imbalanced data for the digit "5", equals

$$\mathrm{F1} = 2 \times (\mathrm{Precision} \times \mathrm{Recall})/(\mathrm{Precision} + \mathrm{Recall}) = 0.9631. \qquad (1)$$

A compact auxiliary neural network enables the APCD to achieve substantially improved recognition performance within a time frame of approximately several tens of microns. A recognition rate error of 3.14% was achieved. The processing time required by the auxiliary neural network to analyze the output data is on the order of several tens of microseconds, which is substantially shorter than the 16.7 ms operational cycle of the APCD-2024. Thus, the proposed hybrid approach practically does not slow down the operation of APCD-2024.

3. Experiment with the APCD-2025

In the APCD-2025, images are input via phase modulation of the optical field. This results in a high level of background illumination around the images under recognition and hence a pronounced intensity peak at the center of the spatial spectrum in the Fourier plane. To suppress this on-axis intensity peak, the phase function of a binary phase grating with a high spatial frequency (Nyquist frequency) is superimposed on the DNN phase function on modulator M2 in the Fourier plane. This grating redirects the light originally concentrated in the central region away from the APCD-2025 detection region. Table 2 presents recognition results from experiments conducted with the APCD-2025 in combination with the aforementioned DNN when recognizing 10,000 characters from the MNIST test set.

Table 2. Recognition results for 10,000 handwritten digits from the MNIST database obtained with the APCD-2025 augmented by a neural network featuring 90 input neurons

	0	1	2	3	4	5	6	7	8	9
0	998	0	0	0	0	0	0	0	0	2
1	0	997	0	0	1	0	0	2	0	0
2	0	17	981	1	0	0	0	1	0	0
3	0	0	1	989	0	9	0	0	0	1
4	0	1	0	0	986	0	0	2	0	11
5	0	1	0	0	9	983	6	0	0	1
6	9	0	0	0	7	0	982	0	1	1
7	0	7	0	0	0	1	0	984	0	8
8	0	0	0	2	1	0	7	0	982	8
9	0	0	0	0	6	0	0	1	0	993

Employing phase-based image input together with post-processing of the intensity distribution in the target regions using a linear perceptron resulted in a 1.3% classification error rate for handwritten digits from the MNIST database. The fairly high classification accuracy achieved on tests images provided a sufficient basis for proceeding to the recognition of real-world objects. Meanwhile, it is worth noting that developers of computer neural networks for the MNIST database reported reaching an error rate of 0.21% as early as 2013 using DropConnect regularization [35], while researchers employing Random Multimodel Deep Learning (RMDL) achieved a 0.18% error rate in 2018 [36].

4. Assessment of the APCD Performance

The authors deem it incorrect to compare the performance of the APCD and a conventional computer by evaluating the time required for a computer to solve the mathematical relationships describing the APCD's operation (computing two Fourier integrals via the fast Fourier transform and so on). We propose comparing the image recognition speed of the APCD with that of a modern graphics processor employing a convolutional neural network, such as YOLO family [37].

Currently, the YOLO (You Only Look Once) family [37, 38] serves as a kind of a benchmark among computer networks for object recognition. This neural network was specifically developed to achieve maximum processing speed while ensuring guaranteed recognition accuracy. Studies have demonstrated that YOLO performs efficiently in real time and can reliably detect small objects in video streams [37]. The authors of [37, 38] conducted a comparative analysis of the inference times required for object recognition using different neural networks from the YOLO family. However, those results were achieved on outdated graphics cards, namely, NVIDIA RTX 2080 (≈4 TFLOPS) and NVidia Tesla T4 16 Gb (8.1 TFLOPS in FP32). Because of this, in this paper, the comparison is made with a contemporary GPU, the RTX 4090 (52.2 TFLOPS).

The APCD can process images with resolutions ranging from 512×512 up to 4096×2140 (the dimension of the UPOLabs HDSLM36R modulator used) at a 60Hz frame rate, i.e. in 16.7 ms. Importantly, the APCD-aided image processing time is independent of the input image dimensionality and determined by the slowest input/output component in the system.

In this paper, we examined the YOLO neural network in a regime where it does not downsample the input image but instead processes every pixel of the frame. This is motivated by the fact that digit images from the MNIST dataset are only 28×28 pixels in size [28]. Downsampling the entire frame would inevitably downsample the digit images as well, leading to a deteriorating recognition quality. When no downsampling is applied, the inference speed of the YOLO neural network depends on the input frame dimensionality. Additionally, the operating speed of the YOLO neural network also depends on the size (type) of the neural network itself. Table 3 gives the image recognition times for images of varying resolution on a GPU RTX 4090 using the YOLO v.8, which was selected as the fastest modification of this family. Three network variants were evaluated (medium, large and extra-large), each providing recognition accuracy comparable to that of the APCD-2025.

Table 3 shows that, considering the APCD's constant processing time of 16.7 ms per frame, YOLO exhibits a superior performance for image resolutions up to 1024×1024 pixels. However, for larger-dimension images, the APCD-2025 outperforms YOLO with a significant margin. This advantage of the APCD-2025 could be further extended by operating the SLM at a higher clock frequency.

A further metric for comparing the performance of the graphic card and the APCD is energy efficiency – specifically, the energy consumed per processed frame. The RTX 4090 GPU consumes about 420 W (WGPU) [39], whereas the APCD-2025 consumes about 56 W (WAPCD). The energy per frame for the GPU is given by EGPU=WGPU×YOLO. For the APCD-2025, the energy required per frame is $E_{APCD} = W_{APCD} \times 16.7 \times 10^{-3}\text{s} = 0.94J$. Table 3 presents the per-frame energy consumption of the RTX 4090 GPU. Here, the energy advantage of the APCD-2025 already starts to be seen at 512 × 512-pixel resolution, and beginning at 2000 × 2000 pixels, the APCD's energy consumption is an order of magnitude lower than that of the RTX 4090 GPU, even for the lightest YOLO modification, YOLO v.8 medium. Thus, the energy efficiency of optical computing with the APCD-2025 is an order of magnitude higher than that of graphics accelerators for frame resolutions of 2000 × 2000 pixels and above.

Table 3. Recognition results for 10,000 handwritten digits from the MNIST database obtained with the APCD-2025 augmented by a neural network featuring 90 input neurons

Frame size	YOLO v.8 medium (ms/J)	YOLO v.8 large (ms/J)	YOLO v.8 xlarge (ms/J)
512 × 512	5.5 / 2.3	6.6 / 2.8	7.8 / 3.3
1024 × 1024	13.8 / 5.8	13.1 / 5.5	17 / 7.1
2048 × 2048	30.3 / 12.7	47.6 / 20	58.8 / 24.7
4000 × 3000	100 /42	154 / 64.7	232 / 97.4

Conclusion

Summing up, we have analyzed two versions of an analog photonic computing device (APCD) and have demonstrated the advantages of the APCD-2025 over the APCD-2024, with the former implementing phase modulation for image input and featuring several design simplifications. The effectiveness of combining the APCD with a compact digital neural network for post-processing the optical recognition results has also been demonstrated. The APCD-2025 has shown clear advantages over GPU-based neural networks in both processing speed and energy efficiency for frame resolutions exceeding 1024 × 1024 pixels. Specifically, at resolutions of 2000 × 2000 and higher, the APCD-2025 outperforms a state-of-the-art RTX 4090 GPU by an order of magnitude in energy efficiency. Hence, the use of the APCD enables real-time object recognition in video streams.

Acknowledgements

This study was conducted within the framework of the scientific program of the National Center for Physics and Mathematics, section 1 "National Center for Research on Supercomputer Architectures". Stage 2026-2027.

References

1. Shen, Y., Harris, N., Skirlo, S., *et al.*: Deep learning with coherent nanophotonic circuits. Nature Photon. 11, 441–446 (2017). `https://doi.org/10.1038/nphoton.2017.93`

2. Harris, N.C., Carolan, J., Bunandar, D., *et al.*: Linear programmable nanophotonic processors. Optica 5(12), 1623–1631 (2018). `https://doi.org/10.1364/OPTICA.5.001623`

3. Zhu, H.H., Zou, J., Zhang, H., *et al.*: Space-efficient optical computing with an integrated chip diffractive neural network. Nat. Commun. 13(1), 1044 (2022). https://doi.org/10.1038/s41467-022-28702-0

4. Kazanskiy, N.L., Khorin, P.A., Golovastikov, N.V., Khonina, S.N.: Analog optical computing: principles, progress, and prospects. Opt. Laser Technol. 193(A), 114220 (2026). https://doi.org/10.1016/j.optlastec.2025.114220

5. Lin, X., Rivenson, Y., Yardimci, N.T., *et al.*: All-optical machine learning using diffractive deep neural networks. Science 361(6406), 1004–1008 (2018). https://doi.org/10.1126/science.aat8084

6. Chang, J., Sitzmann, V., Dun, X., *et al.*: Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. Sci. Rep. 8, 12324 (2018). https://doi.org/10.1038/s41598-018-30619-y

7. Yan, T., Wu, J., Zhou, T., *et al.*: Fourier-space diffractive deep neural network. Phys. Rev. Lett. 123(2), 023901 (2019). https://doi.org/10.1103/physrevlett.123.023901

8. Luo, Y., Mengu, D., Yardimci, N.T., *et al.*: Design of task-specific optical systems using broadband diffractive neural networks. Light Sci. Appl. 8, 112 (2019). https://doi.org/10.1038/s41377-019-0223-1

9. Zhou, T., Fang, L., Yan, T., *et al.*: *In situ* optical backpropagation training of diffractive optical neural networks. Photon. Res. 8(6), 940–953 (2020). https://doi.org/10.1364/PRJ.389553

10. Mengu, D., Zhao, Y., Yardimci, N.T., *et al.*: Misalignment resilient diffractive optical networks. Nanophotonics 9(13), 4207–4219 (2020). https://doi.org/10.1515/nanoph-2020-0291

11. Zhou, T., Lin, X., Wu, J., *et al.*: Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. Nat. Photonics 15(5), 367–373 (2021). https://doi.org/10.1038/s41566-021-00796-w

12. Chen, H., Feng, J., Jiang, M., *et al.*: Diffractive deep neural networks at visible wavelengths. Engineering 7(10), 1483–1491 (2021). https://doi.org/10.1016/j.eng.2020.07.032

13. Zheng, S., Xu, S., Fan, D.: Orthogonality of diffractive deep neural network. Opt. Lett. 47(7), 1798–1801 (2022). https://doi.org/10.1364/OL.449899

14. Motz, G.A., Doskolovich, L.L., Soshnikov, D.V., *et al.*: Design of diffractive neural networks for solving different classification problems at different wavelengths. Photonics 11(8), 780 (2024). https://doi.org/10.3390/photonics11080780

15. Soshnikov, D.V., Doskolovich, L.L., Motz, G.A., *et al.*: Designing robust diffractive neural networks with improved transverse shift tolerance. J. Opt. Soc. Am. A 42(5), 699–709 (2025). https://doi.org/10.1364/JOSAA.561906

16. Soshnikov, D.V., Doskolovich, L.L., Motz, G.A., *et al.*: Design of DOE robust to positioning errors for optical classification purposes. Computer Optics 50(1), 1693 (2026). https://doi.org/10.18287/COJ1693

17. Lohmann, A.W., Paris, D.P.: Binary Fraunhofer holograms, generated by computer. Appl. Opt. 6(10), 1739–1748 (1967). https://doi.org/10.1364/ao.6.001739

18. Fienup, J.R.: Phase retrieval algorithms: a comparison. Appl. Opt. 21(15), 2758–2769 (1982). https://doi.org/10.1364/ao.21.002758

19. Soifer, V.A., Kotlyar, V.V., Doskolovich L.L.: Iterative methods for diffractive optical elements computation. Taylor & Francis (1997). https://doi.org/10.1201/9781482272918

20. Ripoll, O., Kettunen V., Herzig, H.P.: Review of iterative Fourier transform algorithms for beam shaping applications. Opt. Eng. 43(11), 2549–2556 (2004). https://doi.org/10.1117/1.1804543

21. Glses A.A., Jenkins, B.K.: Cascaded diffractive optical elements for improved multiplane image reconstruction. Appl. Opt. 52(15), 3608–3616 (2013). https://doi.org/10.1364/AO.52.003608

22. Wang, H., Piestun, R.: Dynamic 2D implementation of 3D diffractive optics. Optica 5(10), 1220–1228 (2018). https://doi.org/10.1364/OPTICA.5.001220

23. Doskolovich, L.L., Mingazov, A.A., Byzov, E.V., *et al.*: Hybrid design of diffractive optical elements for optical beam shaping. Opt. Express 29(20), 31875–31890 (2021). https://doi.org/10.1364/oe.439641

24. Kazanskiy, N.L., Khonina, S.N., Butt, M.A.: Exploring the functional characteristics of diffractive optical element: A comprehensive review. Opt. Laser Technol. 183, 112383 (2025). https://doi.org/10.1016/j.optlastec.2024.112383

25. Zheng, M., Shi, L., Zi, J.: Optimize performance of a diffractive neural network by controlling the Fresnel number. Photonics Res. 10(11), 2667–2676 (2022). https://doi.org/10.1364/PRJ.474535

26. Lugt, A.V.: Signal detection by complex spatial filtering. IEEE Trans. Inf. Theory 10(2), 139–145 (1964). https://doi.org/10.1109/TIT.1964.1053650

27. Kussul, E., Baidyk, T.: Improved method of handwritten digit recognition tested on MNIST database. Image Vis. Comput. 22(12), 971–981 (2004). https://doi.org/10.1016/j.imavis.2004.03.008

28. LeCun, Y., Cortes, C., Burges, C.J.C.: The MNIST database of handwritten digits. https://web.archive.org/web/20210407152035/http://yann.lecun.com/exdb/mnist/, accessed: 2026-05-14

29. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791

30. Zhang, Y., *et al.*: Memory-less scattering imaging with ultrafast convolutional optical neural networks. Sci. Adv. 10(24), eadn2205 (2024). https://doi.org/10.1126/sciadv.adn2205

31. Zhang, Z., Feng, F., Gan, J., *et al.*: Space-time projection enabled ultrafast all-optical diffractive neural network. Laser Photonics Rev. 18(8), 2301367 (2024). https://doi.org/10.1002/lpor.202301367

32. Li, B., Zhu, Y., Fei, J., *et al.*: Multi-functional broadband diffractive neural network with a single spatial light modulator. APL Photonics 10(1), 016115 (2025). https://doi.org/10.1063/5.0245832

33. Goncharov, D.S., Starikov, R.S., Zlokazov, E.Y.: Pattern recognition system based on a coherent diffractive correlator with deep learned processing of downsampled correlation responses. Appl. Opt. 63(36), 9196–9204 (2024). https://doi.org/10.1364/AO.541305

34. Volkov, A.A., *et al.*: Optoelectronic spatial filtering system for neural network applications. Vestnik natsional'nogo issledovatel'skogo yadernogo universiteta 'MIFI' 15(2), 134–142 (2026). https://doi.org/10.26583/vestnik.2026.2.4

35. Wan, L., Zeiler, M., Zhang, S., *et al.*: Regularization of neural network using DropConnect. In: Dasgupta, S., McAllester, D. (eds.) Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, Georgia, USA, June 17-19, 2013. vol. 28, pp. 1058–1066 (2013).

36. Kowsari, K., Heidarysafa, M., Brown, D.E., *et al.*: RMDL: Random multimodel deep learning for classification. In: ACM International Conference Proceeding Series, ICISDM18, Lakeland, FL, USA, April 9-11, 2018. pp. 19–28 (2018). https://doi.org/10.1145/3206098.3206111

37. Wang, G., *et al.*: TridentYOLO: Improving the precision and speed of mobile device object detection. IET Image Process. 16(1), 145–157 (2022). https://doi.org/10.1049/ipr2.12340

38. Vychegzhanin, S.V., Tatarinova, A.G., Myshkin, R.E.: Comparative analysis of neural network models for detecting unmanned aerial vehicles. Computer Optics 50(2), 1728 (2026). https://doi.org/10.18287/COJ1728

39. https://technicalcity.b-cdn.net/ru/video/Tesla-V100-PCIe-32-GB-protiv-H100--PCIe-80-GB , accessed: 2025-03-04