







Article

Design of Diffractive Neural Networks for Solving Different Classification Problems at Different Wavelengths

Georgy A. Motz ^{1,2}, Leonid L. Doskolovich ^{1,2,*} , Daniil V. Soshnikov ^{1,2} , Egor V. Byzov ^{1,2} ,
Evgeni A. Bezus ^{1,2} , Nikita V. Golovastikov ^{1,2}  and Dmitry A. Bykov ^{1,2} 

¹ Samara National Research University, 34 Moskovskoye Shosse, 443086 Samara, Russia; motzga@mail.ru (G.A.M.)

² Image Processing Systems Institute, National Research Centre “Kurchatov Institute”, 151 Molodogvardeyskaya st., 443001 Samara, Russia

* Correspondence: leonid@ipsiras.ru

Abstract: We consider the problem of designing a diffractive neural network (DNN) consisting of a set of sequentially placed phase diffractive optical elements (DOEs) and intended for the optical solution of several given classification problems at different operating wavelengths, so that each classification problem is solved at the corresponding wavelength. The problem of calculating the DNN is formulated as the problem of minimizing a functional that depends on the functions of the diffractive microrelief height of the DOEs constituting the DNN and represents the error in solving the given classification problems at the operating wavelengths. We obtain explicit and compact expressions for the derivatives of this functional, and using them, we formulate a gradient method for the DNN calculation. Using this method, we design DNNs for solving the following three classification problems at three different wavelengths: the problem of classifying handwritten digits from the MNIST database, the problem of classifying fashion products from the Fashion MNIST database, and the problem of classifying ten handwritten letters from the EMNIST database. The presented simulation results of the designed DNNs demonstrate the high performance of the proposed method.



Citation: Motz, G.A.; Doskolovich, L.L.; Soshnikov, D.V.; Byzov, E.V.; Bezus, E.A.; Golovastikov, N.V.; Bykov, D.A. Design of Diffractive Neural Networks Solving Different Classification Problems at Different Wavelengths. *Photonics* **2024**, *11*, 780. <https://doi.org/10.3390/photonics11080780>

Received: 25 July 2024

Revised: 16 August 2024

Accepted: 20 August 2024

Published: 22 August 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: diffractive neural network; classification problem; cascaded diffractive optical element; gradient method; scalar diffraction theory

1. Introduction

In recent years, the design of photonic structures for optical computing and optical information processing has attracted significant interest. These structures are considered as a promising platform for the further development of computing systems and are intended for creating an alternative to electronic components or supplementing them [1–4]. Optical neural networks [5–9], and, in particular, diffractive neural networks (DNNs), comprising a cascade of sequentially placed phase diffractive optical elements (DOEs) [10–26], are considered as one of the most promising and rapidly developing areas in the field of optical information processing. It should be noted that DOEs (both single and cascaded) have a long history and are widely used for solving a large class of problems of steering laser radiation [27–33]. At the same time, the use of cascaded DOEs for the optical solution of machine learning problems was first demonstrated only in 2018 in ref. [10]. In this work, the authors pointed out several analogies between a cascade of DOEs and “conventional” artificial neural networks and introduced the term “diffractive deep neural network”. The possibility of the optical solution of classification problems using cascaded DOEs was theoretically and experimentally demonstrated in ref. [10]. Subsequent works considered the use of DNNs (cascaded DOEs) for solving various classification problems [11–15,17,25,26], object and video recognition [13,15], salient object detection [11], implementing multispectral imaging [22], and performing matrix multiplication, as well as implementing other linear operators [12,18,20,21]. The main method

for designing DNNs is the stochastic gradient descent method, as well as “improved” first-order methods based on it [34]. These methods have become widely used and have shown their high efficiency in beam shaping problems traditionally solved using DOEs [35,36].

In most works, DNNs are calculated to work with radiation of a single operating wavelength. At the same time, the problem of calculating DNNs designed to work with radiation of various wavelengths is of great scientific and practical interest. In the following text, we will refer to such DNNs as spectral DNNs (or cascaded spectral DOEs). Spectral DNNs can be used to process spectral data, carry out parallel computations by simultaneously solving several machine learning problems at different wavelengths, change their functionality (i.e., the problem being solved) depending on the wavelength of the incident radiation, etc. In particular, references [23,24] considered the calculation of DNNs for spectral filtering and spectral analysis of the incident radiation. In references [21,22], spectral DNNs were considered for the optical implementation of various linear transformations at different wavelengths (each transformation being carried out at its “own” wavelength), as well as for multispectral imaging. One of the main problems, which can be efficiently solved using DNNs, is the problem of optical image classification. However, to the best of our knowledge, the relevant problem of calculating spectral DNNs for solving several different classification problems at different wavelengths has not yet been studied. In particular, although the solution of classification problems using the radiation of several different wavelengths was considered in recent works [25,26], several wavelengths were used only to improve the quality of the solution of a single fixed classification problem. Thus, the solution of several different classification problems at different wavelengths has not been considered in [25,26] (as well as in the other existing works).

In this work, we consider the design of spectral DNNs (cascaded spectral DOEs) for solving several different classification problems at several different wavelengths. We formulate the problem of calculating a spectral DNN as the problem of minimizing a functional representing the error of solving the given classification problems at the operating wavelengths. This functional depends on the functions defining the diffractive microrelief height of the DOEs constituting the DNN. Explicit and compact expressions are obtained for the derivatives of the error functional, and on this basis, a gradient method for the DNN design is presented. Using the proposed gradient method, we calculate several examples of spectral DNNs for solving the following three problems: classification of handwritten digits from the MNIST database at a wavelength of 457 nm, classification of fashion products from the Fashion MNIST database at 532 nm, and classification of ten handwritten letters from A to J (lowercase and uppercase) from the EMNIST database at 633 nm. The presented numerical simulation results demonstrate good classification accuracies provided by the designed spectral DNNs.

2. Design of Spectral DNNs for Solving Several Classification Problems

Let us consider the problem of calculating a spectral DNN (a cascaded DOE) intended for solving several classification problems $P_q, q = 1, \dots, Q$ at different wavelengths $\lambda_q, q = 1, \dots, Q$, so that each classification problem P_q is solved at the corresponding wavelength λ_q . We assume that the cascaded DOE consists of n phase DOEs located in the planes $z = f_1, \dots, z = f_n$ ($0 < f_1 < \dots < f_n$) and defined by the functions of diffractive microrelief height $h_1(\mathbf{u}_1), \dots, h_n(\mathbf{u}_n)$, where $\mathbf{u}_j = (u_j, v_j)$ are Cartesian coordinates in the planes $z = f_j$ (Figure 1).

Let us first describe the required operation of the DNN at a certain single wavelength λ_q . We assume that in the input plane $z = 0$, amplitude images of objects from N_q different classes corresponding to the classification problem P_q are sequentially generated. Each generated image is illuminated by a plane wave with wavelength λ_q . Let us denote $w_{0,q,j}(\mathbf{u}_0)$ as the complex amplitude of the light field generated in this way in the input plane. In the following, the subscript of a certain complex amplitude of the field $w_{m,q,j}(\mathbf{u}_m)$ contains the index m of the plane in which this amplitude is defined, the wavelength index q (which is also the index of the corresponding classification problem), and the class number j of the input image.

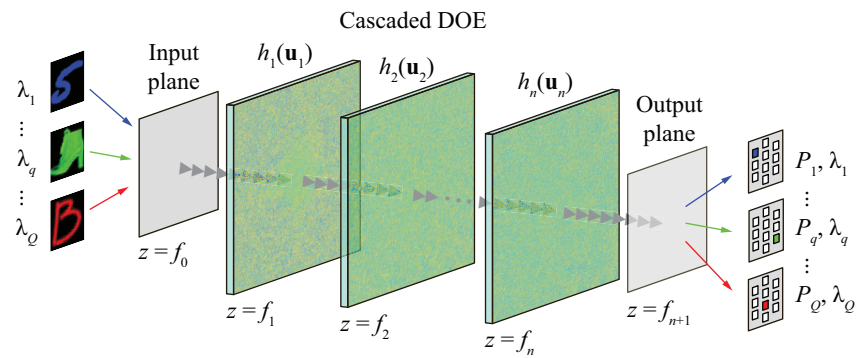


Figure 1. Geometry of the problem of calculating a DNN for solving different classification problems at different wavelengths.

The light field $w_{0,q,j}(\mathbf{u}_0)$ generated at $z = 0$ then propagates through the cascaded DOE to the output plane $z = f_{n+1}$. We assume that the light propagation in the free space (between the planes in which the DOEs are located) is described by the Fresnel–Kirchhoff diffraction integral, and that the transmission of the light field through a DOE can be described in the thin optical element approximation as the multiplication of the beam complex amplitude by the complex transmission function (CTF) of this DOE. The CTF of the m -th DOE is wavelength-dependent, and for the wavelength λ_q , it has the following form:

$$T_{m,q}(\mathbf{u}_m) = \exp\{i\varphi_{m,q}(\mathbf{u}_m)\} = \exp\left\{i\frac{2\pi}{\lambda_q}[n(\lambda_q) - 1]h_m(\mathbf{u}_m)\right\}, \quad (1)$$

where $\varphi_{m,q}(\mathbf{u}_m)$ is the phase function of the DOE (the phase shift introduced by the DOE) at the wavelength λ_q , and $n(\lambda_q)$ is the refractive index of the DOE material. Under these assumptions, the propagation of the input beam $w_{0,q,j}(\mathbf{u}_0)$ from the input plane $z = 0$ through the cascaded DOE to the output plane $z = f_{n+1}$ is described by the following recurrent formula:

$$w_{1,q,j}(\mathbf{u}_1) = C_{q,1} \iint w_{0,q,j}(\mathbf{u}_0) \exp\left\{i\frac{\pi}{\lambda_q d_1}(\mathbf{u}_1 - \mathbf{u}_0)^2\right\} d^2\mathbf{u}_0,$$

$$w_{m,q,j}(\mathbf{u}_m) = C_{q,m} \iint w_{m-1,q,j}(\mathbf{u}_{m-1}) T_{m-1,q}(\mathbf{u}_{m-1}) \cdot \exp\left\{i\frac{\pi}{\lambda_q d_m}(\mathbf{u}_m - \mathbf{u}_{m-1})^2\right\} d^2\mathbf{u}_{m-1}, \quad (2)$$

$$m = 2, \dots, n + 1,$$

where $w_{m,q,j}(\mathbf{u}_m)$, $m = 1, \dots, n$ are the complex amplitudes of the fields incident on the corresponding (m -th) DOEs having the CTFs $T_{m,q}(\mathbf{u}_m)$, $C_{q,m} = (i\lambda_q d_m)^{-1} \exp\{i2\pi d_m / \lambda_q\}$, and $d_m = f_m - f_{m-1}$ are the distances between the adjacent planes.

We assume that in the output plane $z = f_{n+1}$, N_q spatially separated target regions $G_{q,k}$, $k = 1, \dots, N_q$ are defined, which correspond to N_q different classes of the problem P_q (see Figure 1). At each input image, a certain “energy” distribution $E_{q,k}$, $k = 1, \dots, N_q$ is generated in these regions, which corresponds to the integrals of the generated intensity distribution $I_{n+1,q,j}(\mathbf{u}_{n+1}) = |w_{n+1,q,j}(\mathbf{u}_{n+1})|^2$ over the following regions:

$$E_{q,k} = \iint I_{n+1,q,j}(\mathbf{u}_{n+1}) \chi_{q,k}(\mathbf{u}_{n+1}) d^2\mathbf{u}_{n+1}, \quad k = 1, \dots, N_q, \quad (3)$$

where $\chi_{q,k}(\mathbf{u}_{n+1})$ is the indicator function of the region $G_{q,k}$. For solving the classification problem P_q , it is necessary for the cascaded DOE to generate such an intensity distribution in the output plane for the “input signal” of the j -th class $w_{0,q,j}(\mathbf{u}_0)$, so that the maximum of the generated energies $E_{q,k}$, $k = 1, \dots, N_q$ is reached in the corresponding target region $G_{q,j}$ [10,12].

Above, we described the required operation of the DNN at a single wavelength λ_q . The problem of designing a DNN for solving several different classification problems P_q , $q = 1, \dots, Q$ at different wavelengths λ_q can also be formulated as the problem of

calculating the functions of the diffractive microrelief height $h_1(\mathbf{u}_1), \dots, h_n(\mathbf{u}_n)$ of the cascaded DOE. In this case, these functions have to be found in such a way, so that at each wavelength λ_q , with an input signal being an image of a certain object of the problem P_q solved at this wavelength, the DNN provides the maximum energy in the target region corresponding to the class of the input image.

3. Gradient Method for Designing Spectral DNNs

For solving the described problem of calculating a spectral DNN, we will use a stochastic gradient descent method as it is commonly applied for training artificial neural networks. Let us first present a general description of the method. We assume that for the calculation (training) of the DNN (cascaded DOE), a training set $S = S_1 \cup \dots \cup S_Q$ is used, which consists of training subsets S_q for the considered classification problems $P_q, q = 1, \dots, Q$. Each training set S_q contains a number of input distributions (complex amplitudes of the fields) generated from the images of the objects of the problem P_q at the wavelength λ_q . At each step of the method, a set of distributions (referred to as a batch) is randomly chosen from the training set S . For this batch, we calculate the derivatives of a certain error functional $\varepsilon(h_1, \dots, h_n)$, which depends on the functions of the diffractive microrelief height and evaluates the DNN performance. Then, a step in the direction of the anti-gradient is performed, which gives the updated microrelief heights. Since the mathematical expectations of the derivatives calculated over a batch are proportional to the derivatives of the functional calculated for the whole training set, such an approach corresponds to the stochastic gradient descent method. Let us note that in contrast to the majority of the existing works on the spectral DNN design [21–23], below, we will present a detailed derivation of explicit expressions for the derivatives of the error functional.

Without losing generality, we will assume that the batch corresponds to the following set of input distributions: $w_{0,q,j}(\mathbf{u}_0), q = 1, \dots, Q, j = 1, \dots, N_q$. Thus, we assume that the batch contains $N_1 + N_2 + \dots + N_Q$ input distributions, and for each $q \in \{1, \dots, Q\}$, it includes N_q images of the objects of different classes from the training set S_q generated at the corresponding wavelength λ_q . In order to describe the calculations carried out for the batch, let us write the error functional in an explicit form. Let the classification error of an incident beam $w_{0,q,j}(\mathbf{u}_0)$ representing an object from the j -th class from the problem P_q be described by a certain error functional $\varepsilon_{q,j}(h_1, \dots, h_n)$. Since the classification is carried out by analyzing the energies $E_{q,k}$ in the regions $G_{q,k}$ [see Equation (3)], the functional $\varepsilon_{q,j}(h_1, \dots, h_n)$ in the general case has the following form:

$$\varepsilon_{q,j}(h_1, \dots, h_n) = D_{q,j}(E_{q,1}, \dots, E_{q,N_q}), \quad (4)$$

where $D_{q,j}$ is a certain function describing the deviation of the generated energy distribution (3) from the required distribution, in which the energy is concentrated in the required j -th target region. Then, the error functional for a batch containing the distributions $w_{0,q,j}(\mathbf{u}_0), q = 1, \dots, Q, j = 1, \dots, N_q$ can be represented as a sum of the presented functionals:

$$\varepsilon(h_1, \dots, h_n) = \sum_{q=1}^Q \sum_{j=1}^{N_q} \varepsilon_{q,j}(h_1, \dots, h_n). \quad (5)$$

For the functional (5), it is easy to find the Fréchet derivatives $\delta\varepsilon/\delta h_m$. Indeed, since the functional (5) is equal to the sum of functionals, its derivatives have the following form:

$$\frac{\delta\varepsilon(h_1, \dots, h_n)}{\delta h_m} = \sum_{q=1}^Q \sum_{j=1}^{N_q} \frac{\delta\varepsilon_{q,j}(h_1, \dots, h_n)}{\delta h_m}, \quad m = 1, \dots, n. \quad (6)$$

Let us consider the calculation of the derivative $\delta\varepsilon_{q,j}/\delta h_m$ in Equation (6) with respect to the function h_m . To do this, let us first denote the increment of this functional caused by an increment Δh_m of the microrelief height function h_m as follows:

$$\Delta_m\varepsilon_{q,j}(h_1, \dots, h_n) = \varepsilon_{q,j}(h_1, \dots, h_m + \Delta h_m, \dots, h_n) - \varepsilon_{q,j}(h_1, \dots, h_m, \dots, h_n) \quad (7)$$

According to Equations (3) and (4), this increment reads as follows:

$$\begin{aligned} \Delta_m\varepsilon_{q,j}(h_1, \dots, h_n) &= \sum_{k=1}^{N_q} \frac{\partial D_{q,j}}{\partial E_{q,k}} (\Delta_m E_{q,k}) \\ &= \sum_{k=1}^{N_q} \frac{\partial D_{q,j}}{\partial E_{q,k}} \iint [\Delta_m I_{n+1,q,j}(\mathbf{u}_{n+1})] \cdot \chi_{q,k}(\mathbf{u}_{n+1}) \, d^2\mathbf{u}_{n+1} \\ &= \sum_{k=1}^{N_q} \frac{\partial D_{q,j}}{\partial E_{q,k}} \iint \Delta_m [w_{n+1,q,j}(\mathbf{u}_{n+1})w_{n+1,q,j}^*(\mathbf{u}_{n+1})] \cdot \chi_{q,k}(\mathbf{u}_{n+1}) \, d^2\mathbf{u}_{n+1} \\ &= 2 \operatorname{Re} \iint [\Delta_m w_{n+1,q,j}(\mathbf{u}_{n+1})] \cdot F_{n+1,q,j}^*(\mathbf{u}_{n+1}) \, d^2\mathbf{u}_{n+1}, \end{aligned} \quad (8)$$

where $\Delta_m E_{q,k}$, $\Delta_m I_{n+1,q,j}(\mathbf{u}_{n+1})$, and $\Delta_m w_{n+1,q,j}(\mathbf{u}_{n+1})$ are the increments of the energy, intensity distribution, and complex amplitude, respectively, caused by an increment of the height Δh_m , and

$$F_{n+1,q,j}(\mathbf{u}_{n+1}) = w_{n+1,q,j}(\mathbf{u}_{n+1}) \cdot \sum_{k=1}^{N_q} \chi_{q,k}(\mathbf{u}_{n+1}) \frac{\partial D_{q,j}}{\partial E_{q,k}}. \quad (9)$$

By denoting the scalar product of complex functions with angled brackets, we arrive at the following:

$$\Delta_m\varepsilon_{q,j}(h_1, \dots, h_n) = 2 \operatorname{Re} \langle \Delta_m w_{n+1,q,j}(\mathbf{u}_{n+1}), F_{n+1,q,j}(\mathbf{u}_{n+1}) \rangle. \quad (10)$$

One can easily show that the operator describing the forward propagation of the light field through a set of phase DOEs [see Equation (2)], as well as the operator of the backpropagation of the field, are unitary and conserve the scalar product [16]. Using this conservation property, we can represent the increment of the error functional (10) as follows:

$$\Delta_m\varepsilon_{q,j}(\varphi_1, \dots, \varphi_n) = 2 \operatorname{Re} \left\langle \operatorname{Pr}_{f_{n+1} \rightarrow f_m^+} (\Delta_m w_{n+1,q,j}), \operatorname{Pr}_{f_{n+1} \rightarrow f_m^+} (F_{n+1,q,j}) \right\rangle, \quad (11)$$

where $\operatorname{Pr}_{f_{n+1} \rightarrow f_m^+}$ is the backpropagation operator of the field from the output plane $z = f_{n+1}$ to the plane $z = f_m^+$ located immediately after the plane of the m -th DOE $z = f_m$. Note that the backpropagation of the field in the free space is described by the same Fresnel–Kirchhoff integral, where the propagation distance is taken with a minus sign, in contrast to the forward propagation. The “backward propagation” of the beam through a phase DOE is described by the multiplication of the complex amplitude of the beam by the complex conjugate of the CTF of the DOE. Thus, at $m = n$, the field $F_{m,q,j}(\mathbf{u}_{n+1}) = \operatorname{Pr}_{f_{n+1} \rightarrow f_m^+} (F_{n+1,q,j})$ has the following form:

$$F_{n,q,j}(\mathbf{u}_n) = C_{q,n+1}^* \iint F_{n+1,q,j}(\mathbf{u}_{n+1}) \exp \left\{ i\pi \frac{(\mathbf{u}_n - \mathbf{u}_{n+1})^2}{\lambda_q \cdot (-d_{n+1})} \right\} \, d^2\mathbf{u}_{n+1}. \quad (12)$$

Then, at $m < n$, the field $F_{m,q,j}(\mathbf{u}_{n+1})$ is calculated recursively using the following formula:

$$F_{l-1,q,j}(\mathbf{u}_{l-1}) = C_{q,l}^* \iint F_{l,q,j}(\mathbf{u}_l) T_{q,j}^*(\mathbf{u}_l) \exp \left\{ i\pi \frac{(\mathbf{u}_{l-1} - \mathbf{u}_l)^2}{\lambda_q \cdot (-d_l)} \right\} \, d^2\mathbf{u}_l, \quad (13)$$

$$l = n, \dots, m + 1.$$

Let us note that since $\text{Pr}_{f_{n+1} \rightarrow f_n^+}(\Delta_m w_{n+1,q,j}) = \Delta_m(w_{m,q,j} T_{m,q})$, where $w_{m,q,j}(\mathbf{u}_m) T_{m,q}(\mathbf{u}_m)$ is the complex amplitude of the field immediately after the plane of the m -th DOE upon the forward propagation, the increment (11) can be transformed as follows:

$$\begin{aligned} \Delta_m \varepsilon_{q,j}(h_1, \dots, h_n) &= 2 \text{Re} \langle \Delta_m(w_{m,q,j} T_{m,q}), F_{m,q,j} \rangle \\ &= 2 \text{Re} \iint w_{m,q,j}(\mathbf{u}_m) \Delta T_{m,q}(\mathbf{u}_m) F_{m,q,j}^*(\mathbf{u}_m) d^2 \mathbf{u}_m. \end{aligned} \tag{14}$$

Since

$$\begin{aligned} \Delta T_{m,q} &= \exp\{i\gamma_q(h_m + \Delta h_m)\} - \exp\{i\gamma_q h_m\} \\ &= T_{m,q} i\gamma_q \Delta h_m + o(\Delta h_m), \end{aligned} \tag{15}$$

where $\gamma_q = 2\pi[n(\lambda_q) - 1]/\lambda_q$, then the principal linear part of the increment (14) can be written as the following scalar product:

$$\begin{aligned} \delta_m \varepsilon_{q,j}(h_1, \dots, h_n) &= -2\gamma_q \iint \Delta h_m(\mathbf{u}_m) \text{Im}[w_{m,q,j}(\mathbf{u}_m) T_{m,q}(\mathbf{u}_m) F_{m,q,j}^*(\mathbf{u}_m)] d^2 \mathbf{u}_m \\ &= -2\gamma_q \langle \Delta h_m, \text{Im}[w_{m,q,j} T_{m,q} F_{m,q,j}^*] \rangle. \end{aligned} \tag{16}$$

According to Equation (16), the Fréchet derivative of the functional (4) has the following form:

$$\frac{\delta \varepsilon_{q,j}(h_1, \dots, h_n)}{\delta h_m} = -2\gamma_q \text{Im}[w_{m,q,j}(\mathbf{u}_m) T_{m,q}(\mathbf{u}_m) F_{m,q,j}^*(\mathbf{u}_m)]. \tag{17}$$

Thus, the calculation of the gradient of the functional for a batch can be performed using Equations (6) and (17). It is worth noting that in the existing works on the design of spectral DNNs (see, e.g., refs. [21–23]), explicit expressions for the gradient of the error functionals are not presented, and their calculation is performed numerically using the standard PyTorch and TensorFlow frameworks. In this regard, we consider the obtained Equations (6) and (17) for the derivatives of the error functional as a new and important theoretical results.

Above, the functional $\varepsilon_{q,j}(h_1, \dots, h_n)$ describing the classification error of an object of the j -th class in the problem P_q was written in a general form (4), where $D_{q,j}(E_{q,1}, \dots, E_{q,N_q})$ is a certain error function depending on the energy distribution (3) generated at the functions h_1, \dots, h_n . Let us consider a particular example of the functional. For correct classification of an input image of the j -th class, it is necessary for the energy $E_{q,j}$ in the corresponding region $G_{q,j}$ to have a “large” value E_{\max} and for the energies in the other regions to be close to zero. Accordingly, as an error functional for recognizing an input distribution of the j -th class, one can, for example, use the following quadratic functional [19]:

$$\varepsilon_{q,j}(h_1, \dots, h_n) = \sum_{k=1}^{N_q} (E_{q,k} - E_{\max} \delta_{k,j})^2, \tag{18}$$

where $\delta_{k,j}$ is the Kronecker delta. The derivatives of the functional (18) are calculated using the general Formula (17), where, according to Equation (9), the function $F_{m,q,j}(\mathbf{u}_m)$ is calculated through the backpropagation of the field:

$$F_{n+1,q,j}(\mathbf{u}_{n+1}) = 2w_{n+1,q,j}(\mathbf{u}_{n+1}) \sum_{k=1}^{N_q} \chi_{q,k}(\mathbf{u}_{n+1}) \cdot (E_{q,k} - E_{\max} \delta_{k,j}). \tag{19}$$

Let us note that in the design of a cascaded DOE, the functions of the diffractive microrelief height $h_1(\mathbf{u}_1), \dots, h_n(\mathbf{u}_n)$ are usually assumed to be bounded and take values from a certain interval $[0, h_{\max}]$, where h_{\max} is the maximum microrelief height (the h_{\max} value is defined by the technology used for the DOE fabrication). The presence of constraints $0 \leq h_m(\mathbf{u}_1) \leq h_{\max}$, $i = 1, \dots, n$ makes the problem of designing a cascaded DOE

a conditional optimization problem. To take these constraints into account, it is necessary to introduce the following projection operator on the set of bounded height functions into the iterative calculation process:

$$P(h) = \begin{cases} 0, & h < 0, \\ h, & h \in [0, h_{\max}), \\ h_{\max}, & h \geq h_{\max}. \end{cases} \quad (20)$$

In particular, the introduction of this operator to the gradient method for designing cascaded DOEs leads to the gradient projection method, in which the height functions are updated as follows:

$$h_m^k(\mathbf{u}_m) = P \left[h_m^{k-1}(\mathbf{u}_m) - t \frac{\delta \varepsilon}{\delta h_m}(\mathbf{u}_m) \right], \quad m = 1, \dots, n, \quad (21)$$

where the superscript k denotes the iteration number and t is the step of the gradient method. Note that instead of the simplest version of the gradient method of Equation (21), one can utilize its various extensions, e.g., the widely used Adam method [34].

4. Design Examples of Spectral DNNs

Let us consider the calculation of a spectral DNN for solving three different classification problems $P_q, q = 1, 2, 3$ at the following three operating wavelengths: $\lambda_1 = 457$ nm, $\lambda_2 = 532$ nm, and $\lambda_3 = 633$ nm, which correspond to the solid-state lasers commonly used in optical design. Let us choose the following problems to solve: the problem of classifying handwritten digits from the MNIST dataset at the wavelength $\lambda_1 = 457$ nm (problem P_1), the problem of classifying fashion products from the Fashion MNIST dataset at $\lambda_2 = 532$ nm (problem P_2), and, finally, the problem of classifying handwritten letters from A to J (lowercase and uppercase) from the EMNIST dataset at $\lambda_3 = 633$ nm (problem P_3). Note that each of the chosen classification problems contains the objects of ten classes, i.e., $N_1 = N_2 = N_3 = 10$. Let us also note that it is these classification problems that are solved (by separately designed networks operating at a single wavelength) in the vast majority of the existing works on DNN design.

For the DNN design, let us use the following parameters. We assume the input images for the classification problems $P_q, q = 1, 2, 3$ in the input plane to be defined on a 56×56 square grid with a step size of $d = 10 \mu\text{m}$. The “interlayer” distances between the input plane to the first DOE, between the DOEs, and from the last DOE to the output plane are the same and equal $\Delta f = 160$ mm. The microrelief height functions in the DOE planes are defined on 512×512 square grids with a step size (pixel size) of $10 \mu\text{m}$. In this case, the side length of the DOE aperture amounts to 5.12 mm. We set the maximum height of the diffractive microrelief to be $h_{\max} = 6 \mu\text{m}$. Note that DOEs with such height and pixel size have a moderate aspect ratio of $h_{\max}/d = 0.6$ and thus can be fabricated using the standard direct laser writing technique [37,38]. Let us also note that the chosen parameters are in agreement with the results of ref. [17]. In that work, it was shown that a DNN operating at a wavelength of λ behaves like a fully connected neural network and can achieve a good performance if its Fresnel number, defined as $d^2/(\lambda \Delta f)$, lies in the range of $[10^{-4}, 10^{-2}]$. For our design parameters, the Fresnel numbers are of about 10^{-3} for all three operating wavelengths and thus belong to this “optimal” range. In addition, we verified that at the chosen parameters, the diffraction pattern from the input image formed on the first DOE approximately covered the DOE aperture, i.e., it was neither “concentrated” in its central part (which would make the peripheral parts of the DOE not operational) nor noticeably exceeded the DOE boundaries (which would result in a loss of information). For the sake of simplicity, as the refractive indices of the DOE material, we will use the same value $n(\lambda_1) = n(\lambda_2) = n(\lambda_3) = 1.46$, which, nevertheless, is quite close to the refractive index of fused silica at the operating wavelengths.

4.1. Sequential Solution of the Classification Problems

Let us first assume that the images of the objects from different classification problems $P_q, q = 1, 2, 3$ are generated in the input plane $z = 0$ sequentially, and that each image from the problem P_q is illuminated by a normally incident plane wave (propagating along the z axis) with wavelength λ_q . Since each of the considered problems P_q contains the objects of 10 classes, which, as we assumed, are generated in the input plane $z = 0$ in a sequential way, it is sufficient to use a single set of 10 target regions $G_k, k = 1, \dots, 10$ in the output plane for all three problems. In this case, the DNN will change the classification problem being solved by changing the wavelength λ_q of the incident radiation. The target regions, in which energy maxima have to be generated for different classes, are shown in Figure 2a and have a square shape with sides of 0.25 mm. The classes numbered as $0, \dots, 9$ in Figure 2a correspond to the digits $0, \dots, 9$ in problem P_1 , different fashion products (T-shirt/top, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, ankle boot) in problem P_2 , and the letters from A to J in problem P_3 .

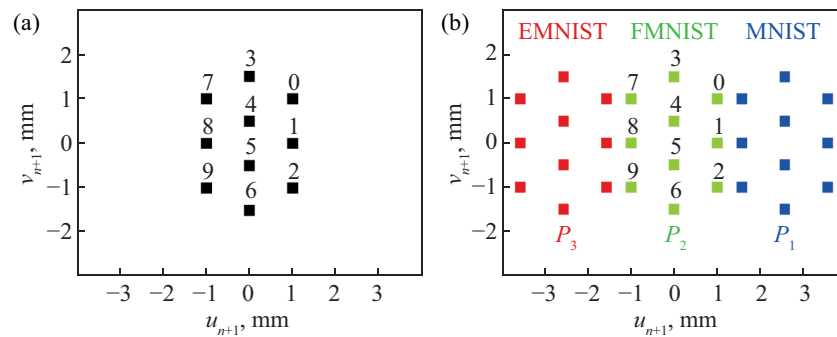


Figure 2. Target regions in the cases of sequential (a) and parallel (b) solution of the classification problems.

First, using the developed gradient method [Equations (5), (6), (17)–(21)], a DNN consisting of a single DOE was calculated. Let us note that the calculations of the derivatives of the error functionals were performed numerically using the angular spectrum method [39,40]. For the DOE calculation, we used a training set S containing 60,000 images of handwritten digits from the MNIST dataset, 60,000 images of fashion products from the Fashion MNIST dataset, and 48,000 images of handwritten letters from the EMNIST dataset. As the initial function of the microrelief height, a realization of white noise with a uniform distribution of values in the $[0, h_{\max}]$ range was used. Note that in addition to the random initial functions of the microrelief height, we also used constant initial functions, which led to close but slightly inferior results in terms of the DNN performance. The training, which was carried out until reasonable convergence of the value of the error functional, took approximately 4 h using an NVIDIA RTX 3060 12 Gb graphics card utilized for the computations. The obtained microrelief height function of the designed DOE is shown in Figure 3a.

After the training, the performance of the calculated DOE was evaluated using a test set containing 10,000 images for each of the problems P_1 and P_2 and 8000 images for problem P_3 (the images from the test set were not included in the training set). The obtained values of the classification accuracies of the objects from different classes (such values are often referred to as recall) for the three considered classification problems are shown with circles in Figure 4a, which are connected with solid lines as a guide to the eye. The overall classification accuracy (i.e., the ratio of the quantity of correctly recognized objects to the size of the test set) amounts to 96.41% for problem P_1 , 84.11% for problem P_2 , and 90.87% for problem P_3 . The full confusion matrices for the three considered classification problems describing the obtained results in more detail are given in the supplementary materials (see Figure S1). Let us note that a relatively low classification accuracy for the objects of the 6-th class of problem P_2 (shirt) in Figure 4a is caused by the fact that these objects are visually close to the objects of the classes 0, 2, and 4 (T-shirt/top, pullover, coat) (see Figure S1). This effect is also present for the latter classes, albeit in this case, it is not as

pronounced (see Figures 4a and S1). Note that this feature is in agreement with the results of other works in which the FMNIST classification problem was considered [10,17].

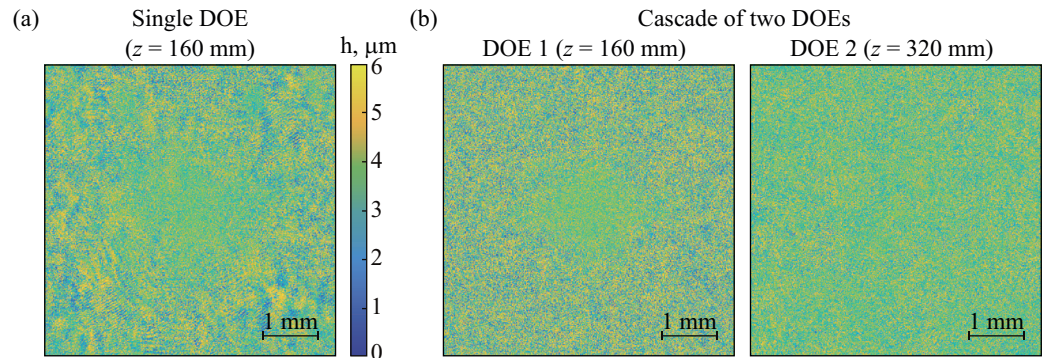


Figure 3. Microrelief height functions of the designed DNNs consisting of a single DOE (a) and a cascade of two DOEs (b) for sequential solution of three classification problems at three wavelengths.

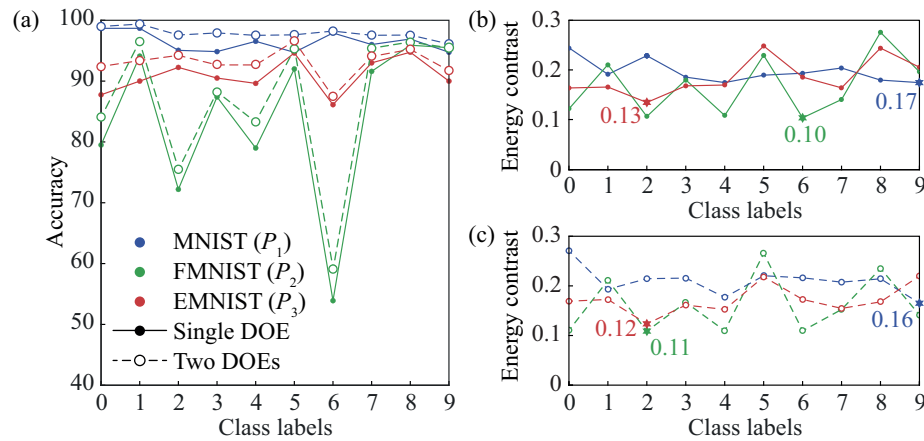


Figure 4. (a) Classification accuracy for single-DOE (solid lines) and two-DOE (dashed lines) DNNs in the case of sequential solution of three classification problems at three wavelengths. (b,c) Contrast for single-DOE (b) and two-DOE (c) DNNs. The stars show the minimum contrast values.

In addition to the classification accuracy, another important parameter is the energy distribution in the target regions generated by the DOE. Let us define $\bar{E}_{q,j \rightarrow k}$ as the average energy calculated for the test set, which is directed to the k -th target region for the input objects of the j -th class from the problem P_q . These average energy values are shown in Figure S1 in the supplementary material in the form of the so-called energy distribution matrices. From a practical point of view, an important characteristic is the contrast value, which shows how much the energy in the region of the class under consideration exceeds the energy in the regions corresponding to the other classes. Let us introduce the contrast for the objects of the j -th class in problem P_q as follows:

$$CR_{q,j} = \frac{\bar{E}_{q,j \rightarrow j} - \max_{k \neq j} \bar{E}_{q,j \rightarrow k}}{\bar{E}_{q,j \rightarrow j} + \max_{k \neq j} \bar{E}_{q,j \rightarrow k}}. \quad (22)$$

In the opinion of the authors, for robust identification of “true maxima” of the energy in the experimental implementation of the DNN, it is necessary for the theoretical values of $CR_{q,j}$ to exceed at least 0.1. The obtained contrast values for the three considered problems are shown in Figure 4b. The minimum contrast values $CR_{\min,q} = \min_j CR_{q,j}$ for problems $P_q, q = 1, 2, 3$ amount to 0.17, 0.10, and 0.13, respectively, and are not less than the chosen “critical” value of 0.1.

It is worth benchmarking the performance of the designed spectral single-DOE DNN solving three classification problems at three different wavelengths against separate DOEs, each of which solves a single classification problem P_q at the corresponding operating wavelength λ_q . These DOEs were calculated using the gradient method using the parameters given above. For the calculated DOEs (not shown here for the sake of brevity), the values of the overall accuracy and minimum contrast obtained using the corresponding test set amount to 96.88% and 0.19 (problem P_1), 86.64% and 0.11 (problem P_2), and 93.3% and 0.13 (problem P_3). As one would expect, the spectral DOE [Figure 3a], which enables solving all three classification problems, provides lower classification accuracies compared to “reference” DOEs designed separately for each of the problems. At the same time, the decrease in accuracy is relatively small, and for the considered problems $P_q, q = 1, 2, 3$, amounts to 0.47%, 2.53%, and 2.36%, respectively. The decrease in the minimum contrast for the three classification problems is also rather small.

It is also interesting to compare the performance of the calculated spectral DOE of Figure 3a with the performance of a DOE solving the same three classification problems, but at a single operating wavelength. This DOE was calculated using the gradient method for the wavelength $\lambda_1 = 457$ nm using the parameters given above. For this DOE (not presented for brevity), the overall accuracy and minimum contrast amount to 92.69% and 0.12 (problem P_1), 81.96% and 0.07 (problem P_2), and 84.9% and 0.10 (problem P_3). One can see that the single DOE solving three classification problems at the same wavelength exhibits inferior performance compared to the spectral DOE. The decrease in the overall classification accuracy occurring when a single operating wavelength is used instead of three different wavelengths amounts to 3.72% (problem P_1), 2.15% (problem P_2), and 5.97% (problem P_3). The better performance of the spectral DOE operating at three different wavelengths can be explained by the fact that the phase shifts introduced by the DOE at different wavelengths are different [see Equation (1)]. In comparison with a DOE designed for a single working wavelength, this provides additional degrees of freedom during the optimization.

Having discussed the properties of a single spectral DOE, let us now move to a DNN comprising two DOEs. The microrelief height functions of the calculated DOEs are shown in Figure 3b. The obtained values of the classification accuracy in the three considered problems for this DNN are shown in Figure 4a with circles connected by dashed lines. The corresponding contrast plots are shown in Figure 4c. The resulting values of the overall classification accuracy and minimum contrast for the designed cascade of two DOEs equal 97.86% and 0.16 (problem P_1), 86.93% and 0.11 (problem P_2), and 93.07% and 0.12 (problem P_3). Full confusion matrices and energy distribution matrices for this structure are given in the supplementary materials (Figure S2). It is evident that the cascade of two DOEs, compared to the single DOE, provides a better performance. In particular, the increase in the overall classification accuracy amounts to 1.45% (problem P_1), 2.82% (problem P_2), and 2.2% (problem P_3) at virtually the same contrast. In order to illustrate the operation of a DNN consisting of two DOEs, in Figure 5, particular examples of input images from the classification problems $P_q, q = 1, 2, 3$ are shown, as well as the corresponding energy distributions generated by the DNN in the output plane.

In addition, we also designed a spectral DNN containing three DOEs (microrelief height functions are not shown in the paper for the sake of brevity). For ease of comparison of the designed DNNs, Table 1 presents the values of the overall classification accuracy and minimum contrast for the single DOE and the cascades of two and three DOEs. One can see that the values of the overall accuracy and minimum contrast for the cascade of three DOEs are 97.89% and 0.20 (problem P_1), 89.75% and 0.11 (problem P_2), and 93.22% and 0.19 (problem P_3). In comparison with the cascade of two DOEs, the cascade of three DOEs provides better values of the minimum contrast for problems P_1 and P_3 and a noticeably higher overall classification accuracy for problem P_2 (the classification accuracy increases by almost 3%). At the same time, the classification accuracy values for problems P_1 and P_3 remain almost unchanged. Let us also note that the addition of a fourth DOE to the cascade leads to only a marginal increase in the classification accuracy and the contrast.

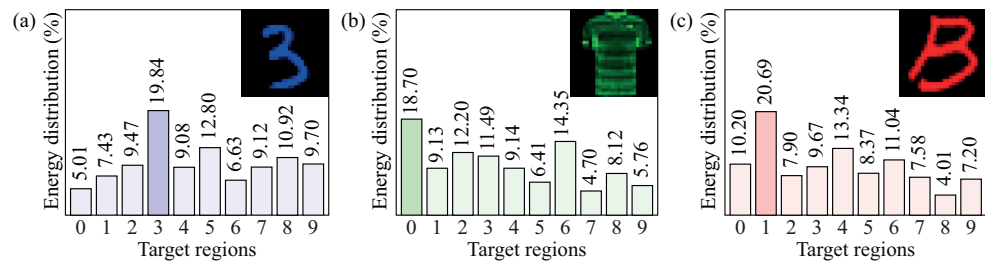


Figure 5. Examples of input images: digit “3” (a), object “T-shirt/top” (b), and letter “B” (c) from the classification problems $P_q, q = 1, 2, 3$ and generated energy distributions in the target regions for a DNN consisting of two DOEs.

To conclude this subsection, let us note that the achieved classification accuracy values are quite high, and even for the case of a single-DOE DNN, they exceed the values obtained in the other works applying DNNs operating at a single wavelength. For example, in refs. [10,12,14], the theoretical classification accuracies for the MNIST classification problem achieved by DNNs consisting of at least five DOEs and working at a single wavelength amounted to 91.75%, 92.28%, and 91.57%, respectively. These accuracy values are significantly lower than the value of 96.41% achieved by the designed single-DOE spectral DNN for the classification problem P_1 (MNIST).

Table 1. Overall accuracy and minimum contrast provided by spectral DNNs consisting of one, two, and three DOEs solving three classification problems in sequential and parallel regimes.

Number of DOEs	Classification Problem	Wavelength λ (nm)	Sequential Regime		Parallel Regime	
			Overall Accuracy (%)	Minimum Contrast	Overall Accuracy (%)	Minimum Contrast
One	P_1 : MNIST	457	96.41	0.17	96.25	0.18
	P_2 : FMNIST	532	84.11	0.10	83.71	0.11
	P_3 : EMNIST	633	90.87	0.13	90.56	0.14
Two	P_1 : MNIST	457	97.86	0.16	97.38	0.19
	P_2 : FMNIST	532	86.93	0.11	87.96	0.11
	P_3 : EMNIST	633	93.07	0.12	92.93	0.16
Three	P_1 : MNIST	457	97.89	0.20	97.41	0.21
	P_2 : FMNIST	532	89.75	0.11	89.10	0.13
	P_3 : EMNIST	633	93.22	0.19	92.95	0.17

4.2. Parallel Solution of the Classification Problems

In the previous subsection, we assumed that the input fields corresponding to objects from different classification problems $P_q, q = 1, 2, 3$ are generated in the input plane $z = 0$ one after another, so that the DNN solves the corresponding classification problems in a sequential way. In this case, it was sufficient to use one set of 10 target regions $G_k, k = 1, \dots, 10$ for all three classification problems [Figure 2a]. Let us now consider the case of parallel solution of the same classification problems $P_q, q = 1, 2, 3$. We will assume that at each moment, three input fields with wavelengths λ_q are simultaneously generated in the input plane. These fields correspond to certain objects from the considered problems of classifying handwritten digits (problem P_1), fashion products (problem P_2), and 10 handwritten letters (problem P_3). Since the problems P_q have to be solved simultaneously, it is necessary to define three spatially separated sets of target regions $G_{q,k}, k = 1, \dots, 10$ corresponding to the problems being solved. The geometry of the target regions used in the present example is shown in Figure 2b.

Let the images of the objects from the classification problems P_q in the input plane be defined on 56×56 grids with a pixel size of $d = 10 \mu\text{m}$, the centers of which for different problems P_q are shifted along the u_0 axis by different distances and are located at the

points $\mathbf{s}_1 = (-2.56, 0)$ mm (problem P_1), $\mathbf{s}_2 = (0, 0)$ (problem P_2), and $\mathbf{s}_3 = (2.56, 0)$ mm (problem P_3). These input images are schematically shown in Figure 1. We will assume that, in contrast to the previous case, the generated images are illuminated by obliquely incident plane waves with wavelengths λ_q and the propagation directions “aimed” from the points \mathbf{s}_q at the center of the first DOE. As before, in the considered DNN examples, the distances between the adjacent planes involved in the DNN design problem are the same and equal 160 mm.

For the considered geometry of parallel solution of the classification problems using the developed gradient method given in Equations (5), (6), (17)–(21), spectral DNN were calculated, consisting of a single DOE and cascades of two and three DOEs. As an example, Figure 6 shows the microrelief height functions of the designed single DOE and cascade of two DOEs. In Figure 7, the corresponding plots of the classification accuracy and contrast are shown. The full confusion and energy distribution matrices are shown in the supplementary materials in Figures S3 and S4. For ease of comparison of the performance of the designed DNNs operating in the sequential and parallel regimes, the values of the overall classification accuracy and minimum contrast for the parallel case are shown in the right part of Table 1. By comparing Figures 4 and 7 and the left and right parts of Table 1, one can see that the classification accuracy values in the sequential and parallel regimes are approximately the same. The rate of accuracy increase with increases in the number of DOEs constituting the DNN is also very similar for the sequential and parallel geometries.

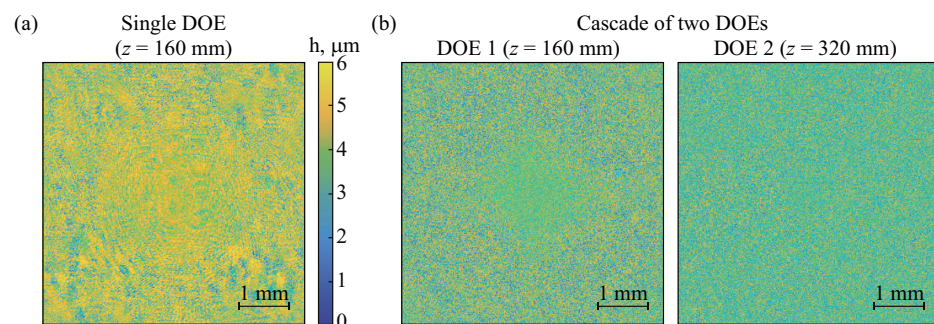


Figure 6. Microrelief height functions of the designed DNNs consisting of a single DOE (a) and a cascade of two DOEs (b) for parallel solution of three classification problems at three wavelengths.

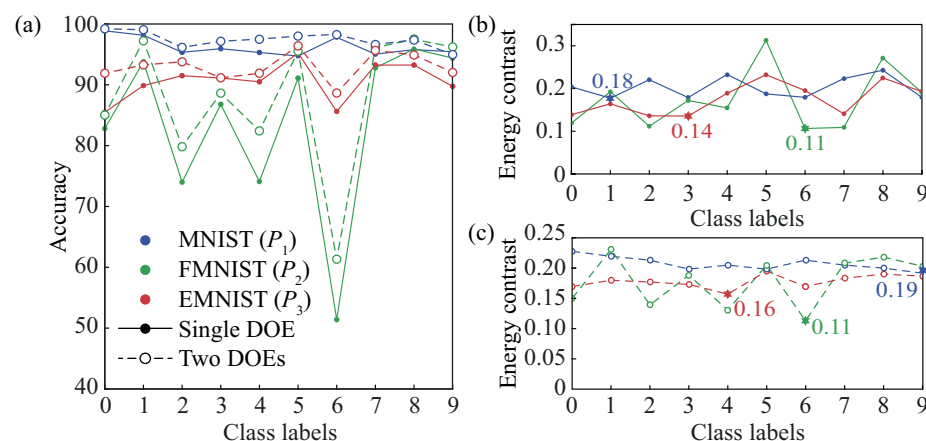


Figure 7. (a) Classification accuracy for single-DOE (solid lines) and two-DOE (dashed lines) DNNs in the case of parallel solution of three classification problems at three wavelengths. (b,c) Contrast for single-DOE (b) and two-DOE (c) DNNs. The stars show the minimum contrast values.

5. Discussion and Conclusions

We presented an approach for designing spectral DNNs (cascaded spectral DOEs) intended for solving several given classification problems at several different wavelengths, with each classification problem being solved at its “own” wavelength of the incident

radiation. In this approach, the problem of calculating the spectral DNN was formulated as the problem of minimizing a functional that depends on the functions of the diffraction microrelief height of the cascaded DOE, representing the error of solving the given classification problems at the design wavelengths. Explicit and compact expressions were obtained for the derivatives of the functional and were used for formulating a gradient method for the DNN calculation.

Using the proposed method, spectral DNNs were designed for solving the following three problems: the problem of classifying handwritten digits from the MNIST database at a wavelength of 457 nm (problem P_1), the problem of classifying fashion products from the Fashion MNIST database at a wavelength of 532 nm (problem P_2), and the problem of classifying ten handwritten letters from A to J (lowercase and uppercase) from the EMNIST database at a wavelength of 633 nm (problem P_3). DNNs were designed for two geometries, assuming sequential and parallel solution of different classification problems. In the first (sequential) geometry, the input beams are normally incident, and a single set of target regions is used for all the classification problems being solved. However, this configuration can also be applied to the case of parallel processing. In this case, similarly to ref. [21], it should be assumed that in the optical setup used to implement the solution of the classification problems, in addition to the DNN, there are additional optical elements that perform wavelength multiplexing of the incident beams in the input plane and wavelength demultiplexing of the resulting field distributions in the output plane. At the same time, the second (parallel) geometry does not require the use of additional multiplexing and demultiplexing devices due to the spatial separation of input and output fields with different wavelengths.

The presented numerical simulation results of the designed DNNs demonstrate the high performance of the proposed approach. In particular, in the parallel regime of solving the classification problems, a cascade of three DOEs provides the overall classification accuracy values of 97.41%, 89.1%, and 92.95% for the P_1 (MNIST), P_2 (Fashion MNIST), and P_3 (EMNIST) problems, respectively. It is important to note that these classification accuracy values (as well as the values achieved in the sequential regime) exceed the values obtained in the other works for DNNs designed for a single operating wavelength. For example, in refs. [10,12,14], the theoretical classification accuracies for the MNIST classification problem achieved by “single-wavelength” DNNs consisting of at least five DOEs amounted to 91.75%, 92.28%, and 91.57%, respectively. The classification accuracy values obtained in the seminal paper [10] for the Fashion MNIST classification problem were equal to 81.13% and 86.60% for DNNs consisting of five and ten DOEs, respectively. We do not present a comparison for the third classification problem (EMNIST), since, to the best of our knowledge, it has not been considered in the existing works dedicated to the DNN design.

An important problem, which is interesting from both theoretical and practical points of view, is the investigation of the achievable number of operating wavelengths (spectral channels) of the DNN and of the influence of this number on the DNN performance (namely, the classification accuracy and contrast). In this regard, it is worth mentioning recent work [21] in which the design of spectral DNNs was considered for the optical implementation of different linear transformations at different wavelengths. In the simulations, the authors claimed to implement more than 180 transformations at different wavelengths; however, their proof-of-concept experiment was carried out at two wavelengths only for very simple linear transformations (permutations of 3×3 matrices), which highlights the complexity of this problem. Such an investigation for spectral DNNs solving different classification problems will be the subject of further research.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/photonics11080780/s1>, Figure S1: Confusion and energy distribution matrices for the problems P_1, P_2, P_3 for a single DOE in the sequential geometry. Figure S2: Confusion and energy distribution matrices for the problems P_1, P_2, P_3 for a cascade of two DOEs in the sequential geometry. Figure S3: Confusion and energy distribution matrices for

the problems P_1, P_2, P_3 for a single DOE in the parallel geometry. Figure S4: Confusion and energy distribution matrices for the problems P_1, P_2, P_3 for a cascade of two DOEs in the parallel geometry.

Author Contributions: Conceptualization, L.L.D.; methodology, L.L.D.; software, G.A.M., D.V.S. and E.V.B.; validation, L.L.D., G.A.M., D.V.S., E.A.B. and D.A.B.; investigation, G.A.M., D.V.S. and L.L.D.; formal analysis, G.A.M., D.V.S., L.L.D., E.A.B., D.A.B. and N.V.G.; writing—original draft preparation, L.L.D., G.A.M. and E.A.B.; writing—review and editing, L.L.D., E.A.B., D.A.B. and N.V.G.; visualization, E.V.B., G.A.M. and D.A.B.; supervision, L.L.D.; project administration, N.V.G. and L.L.D.; funding acquisition, N.V.G. and L.L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Ministry of Science and Higher Education of the Russian Federation (State assignment to Samara University, project FSSS-2024-0016, development of a gradient method for calculating spectral DNNs and its application for solving different classification problems); State assignment of NRC “Kurchatov Institute” (software development for simulating the operation of cascaded DOEs); Russian Science Foundation (project 24-19-00080, general methodology for calculating the Fréchet derivatives of the error functionals based on the unitarity property of light propagation operators).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the presented results are available from the corresponding author upon reasonable request.

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

References

1. Silva, A.; Monticone, F.; Castaldi, G.; Galdi, V.; Alù, A.; Engheta, N. Performing Mathematical Operations with Metamaterials. *Science* **2014**, *343*, 160–163. [[CrossRef](#)] [[PubMed](#)]
2. Zhou, Y.; Zheng, H.; Kravchenko, I.I.; Valentine, J. Flat optics for image differentiation. *Nat. Photonics* **2020**, *14*, 316–323. [[CrossRef](#)]
3. Estakhri, N.M.; Edwards, B.; Engheta, N. Inverse-designed metastructures that solve equations. *Science* **2019**, *363*, 1333–1338. [[CrossRef](#)]
4. Kitayama, K.i.; 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](#)]
5. Shen, Y.; Harris, N.C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.; Laroche, H.; Englund, D.; et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **2017**, *11*, 441–446. [[CrossRef](#)]
6. Harris, N.C.; Carolan, J.; Bunandar, D.; Prabhu, M.; Hochberg, M.; Baehr-Jones, T.; Fanto, M.L.; Smith, A.M.; Tison, C.C.; Alsing, P.M.; et al. Linear programmable nanophotonic processors. *Optica* **2018**, *5*, 1623–1631. [[CrossRef](#)]
7. Zhu, H.H.; Zou, J.; Zhang, H.; Shi, Y.Z.; Luo, S.B.; Wang, N.; Cai, H.; Wan, L.X.; Wang, B.; Jiang, X.D.; et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **2022**, *13*, 1044. [[CrossRef](#)]
8. Zhang, H.; Gu, M.; Jiang, X.D.; Thompson, J.; Cai, H.; Paesani, S.; Santagati, R.; Laing, A.; Zhang, Y.; Yung, M.H.; et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **2021**, *12*, 457. [[CrossRef](#)] [[PubMed](#)]
9. Zhang, J.; Wu, B.; Cheng, J.; Dong, J.; Zhang, X. Compact, efficient, and scalable nanobeam core for photonic matrix-vector multiplication. *Optica* **2024**, *11*, 190–196. [[CrossRef](#)]
10. 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](#)]
11. 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](#)]
12. Zhou, T.; Fang, L.; Yan, T.; Wu, J.; Li, Y.; Fan, J.; Wu, H.; Lin, X.; Dai, Q. In situ optical backpropagation training of diffractive optical neural networks. *Photon. Res.* **2020**, *8*, 940–953. [[CrossRef](#)]
13. Zhou, T.; Lin, X.; Wu, J.; Chen, Y.; Xie, H.; Li, Y.; Fan, J.; Wu, H.; Fang, L.; Dai, Q. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **2021**, *15*, 367–373. [[CrossRef](#)]
14. 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](#)]
15. Ferdman, B.; Saguy, A.; Xiao, D.; Shechtman, Y. Diffractive optical system design by cascaded propagation. *Opt. Express* **2022**, *30*, 27509–27530. [[CrossRef](#)]
16. Zheng, S.; Xu, S.; Fan, D. Orthogonality of diffractive deep neural network. *Opt. Lett.* **2022**, *47*, 1798–1801. [[CrossRef](#)]
17. Zheng, M.; Shi, L.; Zi, J. Optimize performance of a diffractive neural network by controlling the Fresnel number. *Photon. Res.* **2022**, *10*, 2667–2676.
18. Wang, T.; Ma, S.Y.; Wright, L.G.; Onodera, T.; Richard, B.C.; McMahon, P.L. An optical neural network using less than 1 photon per multiplication. *Nat. Commun.* **2022**, *13*, 123. [[CrossRef](#)] [[PubMed](#)]

19. Soshnikov, D.V.; Doskolovich, L.L.; Motz, G.A.; Byzov, E.V.; Bezus, E.A.; Bykov, D.A.; Mingazov, A.A. Design of cascaded diffractive optical elements for optical beam shaping and image classification using a gradient method. *Photonics* **2023**, *10*, 766. [[CrossRef](#)]
20. Kulce, O.; Mengu, D.; Rivenson, Y.; Ozcan, A. All-optical synthesis of an arbitrary linear transformation using diffractive surfaces. *Light. Sci. Appl.* **2021**, *10*, 196. [[CrossRef](#)]
21. Li, J.; Gan, T.; Bai, B.; Luo, Y.; Jarrahi, M.; Ozcan, A. Massively parallel universal linear transformations using a wavelength-multiplexed diffractive optical network. *Adv. Photonics* **2023**, *5*, 016003. [[CrossRef](#)]
22. Mengu, D.; Tabassum, A.; Jarrahi, M.; Ozcan, A. Snapshot multispectral imaging using a diffractive optical network. *Light. Sci. Appl.* **2023**, *12*, 86. [[CrossRef](#)]
23. 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](#)]
24. Zhu, Y.; Chen, Y.; Negro, L.D. Design of ultracompact broadband focusing spectrometers based on diffractive optical networks. *Opt. Lett.* **2022**, *47*, 6309–6312. [[CrossRef](#)]
25. Shi, J.; Chen, Y.; Zhang, X. Broad-spectrum diffractive network via ensemble learning. *Opt. Lett.* **2022**, *47*, 605–608. [[CrossRef](#)]
26. Feng, J.; Chen, H.; Yang, D.; Hao, J.; Lin, J.; Jin, P. Multi-wavelength diffractive neural network with the weighting method. *Opt. Express* **2023**, *31*, 33113–33122. [[CrossRef](#)]
27. Fienup, J.R. Phase retrieval algorithms: A comparison. *Appl. Opt.* **1982**, *21*, 2758–2769. [[CrossRef](#)]
28. Soifer, V.A.; Kotlyar, V.; Doskolovich, L. *Iterative Methods for Diffractive Optical Elements Computation*; CRC Press: Boca Raton, FL, USA, 1997.
29. Ripoll, O.; Kettunen, V.; Herzig, H.P. Review of iterative Fourier-transform algorithms for beam shaping applications. *Opt. Eng.* **2004**, *43*, 2549–2556.
30. Lатышевская, Т. Iterative phase retrieval in coherent diffractive imaging: Practical issues. *Appl. Opt.* **2018**, *57*, 7187–7197. [[CrossRef](#)]
31. Deng, X.; Chen, R.T. Design of cascaded diffractive phase elements for three-dimensional multiwavelength optical interconnects. *Opt. Lett.* **2000**, *25*, 1046–1048. [[CrossRef](#)]
32. Gülses, A.A.; Jenkins, B.K. Cascaded diffractive optical elements for improved multiplane image reconstruction. *Appl. Opt.* **2013**, *52*, 3608–3616. [[CrossRef](#)]
33. Wang, H.; Piestun, R. Dynamic 2D implementation of 3D diffractive optics. *Optica* **2018**, *5*, 1220–1228. [[CrossRef](#)]
34. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Shi, J.; Wei, D.; Hu, C.; Chen, M.; Liu, K.; Luo, J.; Zhang, X. Robust light beam diffractive shaping based on a kind of compact all-optical neural network. *Opt. Express* **2021**, *29*, 7084–7099. [[CrossRef](#)] [[PubMed](#)]
36. Buske, P.; Völl, A.; Eisebitt, M.; Stollenwerk, J.; Holly, C. Advanced beam shaping for laser materials processing based on diffractive neural networks. *Opt. Express* **2022**, *30*, 22798–22816. [[CrossRef](#)]
37. Doskolovich, L.L.; Mingazov, A.A.; Byzov, E.V.; Skidanov, R.V.; Ganchevskaya, S.V.; Bykov, D.A.; Bezus, E.A.; Podlipnov, V.V.; Porfirev, A.P.; Kazanskiy, N.L. Hybrid design of diffractive optical elements for optical beam shaping. *Opt. Express* **2021**, *29*, 31875–31890. [[CrossRef](#)]
38. Doskolovich, L.L.; Skidanov, R.V.; Bezus, E.A.; Ganchevskaya, S.V.; Bykov, D.A.; Kazanskiy, N.L. Design of diffractive lenses operating at several wavelengths. *Opt. Express* **2020**, *28*, 11705–11720. [[CrossRef](#)]
39. Schmidt, J.D. *Numerical Simulation of Optical Wave Propagation with Examples in MATLAB*; SPIE: Bellingham, WA, USA, 2010.
40. Cubillos, M.; Jimenez, E. Numerical simulation of optical propagation using sinc approximation. *J. Opt. Soc. Am. A* **2022**, *39*, 1403–1413. [[CrossRef](#)]

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.