



Design of Cascaded Diffractive Optical Elements for Optical Beam Shaping and Image Classification Using a Gradient Method

Daniil V. Soshnikov^{1,2}, Leonid L. Doskolovich^{1,2,*}, Georgy A. Motz², Egor V. Byzov^{1,2}, Evgeni A. Bezus^{1,2}, Dmitry A. Bykov^{1,2} and Albert A. Mingazov^{1,2}

- ¹ Image Processing Systems Institute, Branch of the Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, 151 Molodogvardeyskaya Street, 443001 Samara, Russia
- ² Samara National Research University, 34 Moskovskoye Shosse, 443086 Samara, Russia
- * Correspondence: leonid@ipsiras.ru

Abstract: We present a gradient method for designing cascaded diffractive optical elements (DOEs) consisting of several sequentially located phase DOEs. Using the unitarity property of the operator of light propagation through the cascaded DOE, we obtain explicit expressions for the derivatives of the error functional with respect to the phase functions of the cascaded DOE. We consider the application of the gradient method to the problem of focusing different incident beams to regions with different intensity distributions and to the problem of optical image classification. The presented description of the gradient method treats the problems of designing cascaded DOEs for focusing laser radiation and for image classification within a unified approach reducing the calculation of single and cascaded DOEs for focusing different incident beams to different regions and for classifying handwritten digits, which demonstrate the high performance of the proposed method. The presented results may find application in the design of diffractive neural networks and systems for focusing laser radiation.

Keywords: diffractive optical element; phase function; scalar diffraction theory; gradient method; image classification

1. Introduction

Nowadays, the design and investigation of diffractive optical elements (DOEs) are the subjects of active research [1–7]. The main reasons for the interest in this research field are the compactness, manufacturability, and efficiency of using DOEs for solving a wide range of problems of transforming and focusing optical radiation. As a rule, the DOE design is carried out within the framework of the scalar diffraction theory. The problem of calculating a DOE belongs to the class of ill-posed inverse problems and consists in determining the shape of the "phase" diffractive microrelief, which ensures the formation of a light field with specified parameters (usually, with a required intensity distribution) in a given region of space. Since the height of the DOE microrelief is proportional to the phase function of the light field formed by the DOE, the problem of DOE design is usually considered as the problem of calculating a phase function, which ensures the generation of the required intensity distribution. For the calculation of the phase function, various iterative algorithms are traditionally used including the "classical" Gerchberg–Saxton algorithm, the error-reduction algorithm, and a wide range of their modifications [8–14].

In addition to single DOEs, the so-called cascaded DOEs are widely used, consisting of several sequentially located phase DOEs. Such DOEs possess a wider functionality and make it possible to solve more complex problems, e.g., the problem of focusing different incident beams (in particular, with different propagation directions or different wavelengths) into different regions [2,15,16]. For the calculation of cascaded DOEs, iterative algorithms



Citation: 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. https://doi.org/10.3390/ photonics10070766

Received: 23 May 2023 Revised: 19 June 2023 Accepted: 29 June 2023 Published: 3 July 2023



Copyright: © 2023 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/). are also used, which are a generalization of the existing algorithms for designing single DOEs. At the same time, most of these algorithms are heuristic and do not have a strict theoretical justification. In particular, for the iterative algorithms used in [2,15,16], there is no analysis of the error decrease property, which the Gerchberg–Saxton algorithm and error-reduction algorithm possess. In this regard, it appears relevant to develop gradient methods for calculating cascaded DOEs, which, at least, have a clear theoretical justification.

In addition to the utilization of cascaded DOEs in complex problems of focusing laser radiation, in recent years, they have found wide application in solving various problems of machine learning (in particular, the problems of image classification) [3,5,17–22], as well as in the problems of implementing various mathematical transformations described by linear operators [23]. In these problems, cascaded DOEs are often referred to as diffractive neural networks (DNNs). The main method for designing DNNs is the stochastic gradient method, as well as the "improved" first-order methods based on it [24]. In a number of works, expressions were obtained for the gradients of the error functions with respect to the parameters of the phase functions [2,5,20]. However, these expressions were obtained for a discrete form of the operators of light propagation between the DOEs. As a result, the expressions for the gradients have a complex and cumbersome form, which, in the opinion of the present authors, complicates their understanding and implementation.

In this work, we present a gradient method for calculating cascaded DOEs. Using the unitarity property of the operator of light propagation through a cascaded DOE, we obtain explicit and compact expressions for the derivatives of the error functional with respect to the phase functions of the cascaded DOE. We consider the application of the gradient method to the problem of focusing different incident beams to regions with different intensity distributions and to the problem of image classification. The presented description of the gradient method "combines" the problems of designing cascaded DOEs for focusing laser radiation and for classifying images within a single methodological approach, in which the calculation of the derivatives of the error functionals is reduced to the same formula. Using the proposed gradient method, single and cascaded DOEs are calculated for solving the problem of focusing several incident beams on different regions and the problem of classifying handwritten digits. The presented numerical simulation results demonstrate the high performance of the proposed method.

2. Problem Statement

Let the complex amplitude of the "input" field $w_0(\mathbf{u}_0)$ be defined in the input plane $z = f_0 = 0$, where $\mathbf{u}_0 = (u_0, v_0)$ are the Cartesian coordinates in this plane. We will assume that the light field with the wavelength λ propagates from the plane z = 0 through a set of n phase DOEs located in the planes $z = f_1, \ldots, z = f_n$ ($0 < f_1 < \ldots < f_n$), and finally comes to the output plane $z = f_{n+1} > f_n$ (Figure 1).

Let us denote by $\varphi_1(\mathbf{u}_1), \ldots, \varphi_n(\mathbf{u}_n)$ the phase functions of the DOEs, where $\mathbf{u}_j = (u_j, v_j)$ are the Cartesian coordinates in the planes $z = f_1, \ldots, z = f_n$. We will assume that the propagation of light between the planes $z = f_i$, $i = 1, \ldots, n + 1$ is described by the Fresnel–Kirchhoff integral of the scalar diffraction theory. We will describe the propagation of light through a DOE in the thin optical element approximation as the multiplication of the complex amplitude of the incident beam by the complex transmission function of the DOE $\exp\{i\varphi_m(\mathbf{u}_m)\}, m = 1, \ldots n$.



Figure 1. Geometry of the problem of the design of a cascaded DOE.

In this case, the propagation of light through a cascaded DOE is described by the following expressions:

$$w_{1}(\mathbf{u}_{1}) = \frac{e^{ikd_{1}}}{\lambda id_{1}} \iint w_{0}(\mathbf{u}_{0}) \exp\left\{i\frac{\pi}{\lambda d_{1}}(\mathbf{u}_{1}-\mathbf{u}_{0})^{2}\right\} d^{2}\mathbf{u}_{0},$$

$$w_{m}(\mathbf{u}_{m}) = \frac{e^{ikd_{m}}}{\lambda id_{m}} \iint w_{m-1}(\mathbf{u}_{m-1}) e^{i\varphi_{m-1}(\mathbf{u}_{m-1})} \exp\left\{i\frac{\pi}{\lambda d_{m}}(\mathbf{u}_{m}-\mathbf{u}_{m-1})^{2}\right\} d^{2}\mathbf{u}_{m-1}, \quad m = 2, \dots, n+1,$$
(1)

where $w_m(\mathbf{u}_m)$, m = 1, ..., n are the complex amplitudes of the fields incident on DOEs located in the planes $z = f_m$, and $d_m = f_m - f_{m-1}$ are the distances between these planes. According to Equation (1), the calculation of the complex amplitude of the output field $w_{n+1}(\mathbf{u}_{n+1})$ is carried out recursively. For the following analysis, it will be convenient to consider Equation (1) as a representation of linear operators describing the propagation of light from the input plane $z = f_0$ to the planes $z = f_m$, m = 1, ..., n + 1.

Under the inverse problem, we will understand the problem of calculating the phase functions $\varphi_1(\mathbf{u}_1), \ldots, \varphi_n(\mathbf{u}_n)$ from the condition of generating a light field with the required intensity distribution $I(\mathbf{u}_{n+1})$ in the output plane. Let us describe the error of generating the required distribution using an integral criterion

$$\varepsilon(\varphi_1,\ldots,\varphi_n) = \iint D[I_{n+1}(\mathbf{u}_{n+1}),I(\mathbf{u}_{n+1})]d^2\mathbf{u}_{n+1},$$
(2)

where $I_{n+1}(\mathbf{u}_{n+1}) = |w_{n+1}(\mathbf{u}_{n+1})|^2$ is the intensity distribution generated by a cascaded DOE with phase functions $\varphi_1(\mathbf{u}_1), \ldots, \varphi_n(\mathbf{u}_n)$, and *D* is a certain function representing the difference between the generated and required distributions at the current point.

In what follows, we will consider the inverse problem of designing a cascaded DOE as the problem of minimizing the functional of Equation (2), which we will refer to as the error functional:

$$\varepsilon(\varphi_1,\ldots,\varphi_n) \to \min_{\varphi_1,\ldots,\varphi_n}.$$
 (3)

3. Gradient Method for Calculating a Cascaded DOE

For solving the problem of Equation (3), we will utilize a gradient method. Let us consider the calculation of the derivative of the functional of Equation (3) with respect to the function φ_m . Let us denote by

$$\Delta_m \varepsilon(\varphi_1, \dots, \varphi_n) = \varepsilon(\varphi_1, \dots, \varphi_m + \Delta \varphi_m, \dots, \varphi_n) - \varepsilon(\varphi_1, \dots, \varphi_m, \dots, \varphi_n), \qquad (4)$$

the increment of the error functional caused by the increment $\Delta \varphi_m$ of the function φ_m . According to Equation (2), this increment has the form

$$\Delta_{m}\varepsilon(\varphi_{1},\ldots,\varphi_{n}) = \iint \frac{\partial D[I_{n+1}(\mathbf{u}_{n+1}),I(\mathbf{u}_{n+1})]}{\partial I_{n+1}} \Delta_{m} (w_{n+1}(\mathbf{u}_{n+1})w_{n+1}^{*}(\mathbf{u}_{n+1}))d^{2}\mathbf{u}_{n+1} = \iint \frac{\partial D[I_{n+1}(\mathbf{u}_{n+1}),I(\mathbf{u}_{n+1})]}{\partial I_{n+1}} 2\operatorname{Re}[(\Delta_{m}w_{n+1}(\mathbf{u}_{n+1}))w_{n+1}^{*}(\mathbf{u}_{n+1})]d^{2}\mathbf{u}_{n+1} = 2\operatorname{Re}[\iint (\Delta_{m}w_{n+1}(\mathbf{u}_{n+1}))F_{n+1}^{*}(\mathbf{u}_{n+1})d^{2}\mathbf{u}_{n+1}] = 2\operatorname{Re}\langle\Delta_{m}w_{n+1}(\mathbf{u}_{n+1}),F_{n+1}(\mathbf{u}_{n+1})\rangle,$$
(5)

where $\Delta_m I_{n+1}(\mathbf{u}_{n+1})$ and $\Delta_m w_{n+1}(\mathbf{u}_{n+1})$ are the increments of the intensity and complex amplitude caused by the phase increment $\Delta \varphi_m$, the angle brackets denote the scalar product of functions, and the function $F_{n+1}(\mathbf{u}_{n+1})$ has the form

$$F_{n+1}(\mathbf{u}_{n+1}) = \frac{\partial D[I_{n+1}(\mathbf{u}_{n+1}), I(\mathbf{u}_{n+1})]}{\partial I_{n+1}} w_{n+1}(\mathbf{u}_{n+1}).$$
(6)

For the following derivations, let us introduce the operator $\Pr_{f_{n+1} \to f_m^+}$ of the "backpropagation" of light from the output plane $z = f_{n+1}$ to the plane $z = f_m^+$ located immediately after the plane of the *m*-th DOE $z = f_m$. In this plane, the complex amplitude of the field in the case of "direct" propagation reads as $w_m e^{i\varphi_m}$. Let us present formulas for calculating this operator by starting with the field $F_{n+1}(\mathbf{u}_{n+1})$ (see Figure 1). At m = n, the backpropagation operator corresponds to the Fresnel–Kirchhoff integral, in which the propagation distance d_{n+1} is taken with a negative sign:

$$F_{n}(\mathbf{u}_{n}) = \Pr_{f_{n+1} \to f_{n}^{+}}(F_{n+1}) = \frac{e^{-ikd_{n+1}}}{-\lambda i d_{n+1}} \iint F_{n+1}(\mathbf{u}_{n+1}) \exp\left\{-i\frac{\pi}{\lambda d_{n+1}}(\mathbf{u}_{n} - \mathbf{u}_{n+1})^{2}\right\} d^{2}\mathbf{u}_{n+1}.$$
(7)

At m < n, the operator $\Pr_{f_{n+1} \to f_m^+}(F_{n+1})$ is calculated recursively using the following expression:

$$F_{j-1}(\mathbf{u}_{j-1}) = \frac{e^{-ikd_j}}{-\lambda i d_j} \iint F_j(\mathbf{u}_j) e^{-i\varphi_j(\mathbf{u}_j)} \exp\left\{-i\frac{\pi}{\lambda d_j} (\mathbf{u}_{j-1} - \mathbf{u}_j)^2\right\} d^2 \mathbf{u}_j, \quad j = n, \dots, m+1.$$
(8)

One can easily show that the operators of direct propagation and backpropagation of light through a set of phase DOEs are unitary and conserve the scalar product [18]. Due to the scalar product conservation, the increment of the criterion of Equation (5) can be represented as

$$\Delta_m \varepsilon(\varphi_1, \dots, \varphi_n) = 2\operatorname{Re}\langle \Delta_m w_{n+1}, F_{n+1} \rangle = 2\operatorname{Re}\langle \operatorname{Pr}_{f_{n+1} \to f_m^+}(\Delta_m w_{n+1}), \operatorname{Pr}_{f_{n+1} \to f_m^+}(F_{n+1}) \rangle.$$
(9)

Since $\Pr_{f_{n+1}\to f_m^+}(\Delta_m w_{n+1}) = \Delta_m(w_m e^{i\varphi_m})$, where $w_m e^{i\varphi_m}$ is the complex amplitude of the field immediately after the plane of the *m*-th DOE at direct propagation and $\Pr_{f_{n+1}\to f_m^+}(F_{n+1}) = F_m$. Let us transform Equation (9) to the form

$$\Delta_m \varepsilon(\varphi_1, \dots, \varphi_n) = 2\operatorname{Re}\left\langle \Delta_m(w_m e^{i\varphi_m}), F_m \right\rangle = 2\operatorname{Re} \iint w_m(\mathbf{u}_m) \Delta e^{i\varphi_m(\mathbf{u}_m)} F_m^*(\mathbf{u}_m) d^2 \mathbf{u}_m.$$
(10)

By substituting the increment $\Delta e^{i\varphi_m}$ as its Taylor series expansion up to the linear term $\Delta e^{i\varphi_m} = e^{i\varphi_m + i\Delta\varphi_m} - e^{i\varphi_m} \approx i\Delta\varphi_m e^{i\varphi_m}$, let us write the main (linear) part of the increment of the functional (10) as

$$\delta_m \varepsilon(\varphi_1, \dots, \varphi_n) = -2 \iint \Delta \varphi_m(\mathbf{u}_m) \operatorname{Im}[w_m(\mathbf{u}_m) e^{i\varphi_m(\mathbf{u}_m)} F_m^*(\mathbf{u}_m)] d^2 \mathbf{u}_m.$$
(11)

According to Equation (11), the derivative of the functional has the form

$$\frac{\delta\varepsilon(\varphi_1,\ldots,\varphi_n)}{\delta\varphi_m} = -2\mathrm{Im}[w_m(\mathbf{u}_m)e^{i\varphi_m(\mathbf{u}_m)}F_m^*(\mathbf{u}_m)].$$
(12)

When solving the problem of minimizing the functional (3) using the gradient method, the calculation of the phase functions of the DOEs is carried out iteratively. Let us describe the calculations performed at each iteration of the method. Let $\varphi_1^k(\mathbf{u}_1), \ldots, \varphi_n^k(\mathbf{u}_n)$ be the phase functions of the cascaded DOE obtained at the *k*-th iteration. Then, for the calculation of the next approximations of the phase functions, the following steps are performed:

(1) Using Equation (1) describing the direct propagation of the field, complex amplitudes of the fields $w_m(\mathbf{u}_m)e^{i\varphi_m(\mathbf{u}_m)}$ in the planes $z = f_m$, m = 1, ..., n and in the output plane $z = f_{n+1}$ are calculated.

(2) For the chosen optimization criterion, the function $F_{n+1}(\mathbf{u}_{n+1})$ of Equation (6) is calculated, and, using Equations (7) and (8) describing the backpropagation of the field, the functions $F_j(\mathbf{u}_j)$, j = n, n - 1, ..., 1 are found.

(3) Using Equation (12), the derivatives of the functional $\frac{\delta \varepsilon}{\delta \varphi_m}$, m = 1, ..., n are calculated.

(4) New approximations of the phases are found as

$$\varphi_m^{k+1}(\mathbf{u}_m) = \varphi_m^k(\mathbf{u}_m) - t_k \frac{\delta\varepsilon}{\delta\varphi_m}(\mathbf{u}_m), \ m = 1, \dots, n,$$
(13)

where t_k is the step of the gradient method.

4. Application of the Gradient Method to the Case of Several Incident Beams

The presented gradient method can be easily generalized to the problem, in which there are defined K > 1 different input distributions $w_{0,j}(\mathbf{u}_0), j = 1, ..., K$ (different incident beams), and, for each input distribution, the cascaded DOE has to generate a corresponding output intensity distribution $I_j(\mathbf{u}_{n+1})$. In this case, the following sum of functionals can be used as the error functional:

$$\varepsilon_{set}(\varphi_1,\ldots,\varphi_n) = \sum_{j=1}^K \varepsilon^j(\varphi_1,\ldots,\varphi_n), \tag{14}$$

where the functionals $\varepsilon^{j}(\varphi_{1}, ..., \varphi_{n})$ describe the difference of the intensity distributions $I_{n+1,j}(\mathbf{u}_{n+1})$ generated for the input distributions $w_{0,j}(\mathbf{u}_{0})$ from the required distributions $I_{j}(\mathbf{u}_{n+1})$. Without loss of generality, we can assume that these functionals are defined by Equation (2). Since the derivatives of the sum of functionals (14) simply equal the sum of the derivatives of these functionals

$$\frac{\delta\varepsilon_{set}(\varphi_1,\ldots,\varphi_n)}{\delta\varphi_m} = \sum_{j=1}^K \frac{\delta\varepsilon^j(\varphi_1,\ldots,\varphi_n)}{\delta\varphi_m}, \ m = 1,\ldots,n,$$
(15)

the calculation of the derivatives of the functional of Equation (14) is also reduced to Equation (12). The calculation of new approximations of the phase functions at each iteration is carried out using a formula similar to Equation (13). The investigation of the performance of the presented method in the problem of the design of cascaded DOEs

generating different intensity distributions at different incident beams is presented below in Section 6.1.

Let us note that the considered gradient method can also be easily generalized to the problem, in which the required intensity distributions $I_j(\mathbf{u}_{n+1})$ are defined in different output planes located at different distances from the output DOE.

5. Application of the Gradient Method to the Image Classification Problem

Let us now consider the application of the developed gradient method to the design of cascaded DOEs performing optical image classification. Let the amplitude images of objects belonging to *L* different classes (for example, the images of handwritten digits) be generated in the input plane z = 0. The generated light field then propagates through the cascaded DOE and comes to the output plane $z = f_{n+1}$. Let *L* spatially separate regions G_i corresponding to the images of different classes be defined in the output plane. At each input distribution (image), certain energy distribution E_i , i = 1, ..., L is generated in these regions, where E_i corresponds to the integral of the generated intensity distribution over the region G_i . In this case, the problem of designing a cascaded DOE for classifying images can be formulated as a problem of calculating such phase functions $\varphi_1(\mathbf{u}_1), ..., \varphi_n(\mathbf{u}_n)$ of the cascaded DOE, so that for an "input signal" corresponding to an image of a certain *j*-th class, the maximum of the generated energies E_i , i = 1, ..., L is achieved in the corresponding region G_i [3,5].

In the problems of calculating cascaded DOEs for image classification, approaches typical for the design of artificial neural networks are used [3,5,19–21]. In this case, for the design (training) of the cascaded DOE, a training data set is used, which contains a number of input distributions corresponding to the images of objects belonging to different classes. Due to the large size of the training set, usually, for performing a training step, a smaller set (batch) of distributions is randomly chosen from the whole set, for which the derivatives of the error functional are calculated. One can show that the expectations of the derivatives calculated over a batch are proportional to the derivatives calculated using the whole training set, which enables considering this approach as a stochastic gradient method.

The training of a cascaded DOE on a particular batch corresponds to the gradient method in the case of several incident beams. Indeed, the error functional in the case of training on a batch can be defined in the form of Equation (14), where *K* is the batch size, and the functionals $\varepsilon^{j}(\varphi_{1}, \ldots, \varphi_{n})$ represent the classification errors of different classes included in the batch. The difference of the image classification problem from the problem of generating different intensity distributions for different incident beams consists mainly in the form of the functionals $\varepsilon^{j}(\varphi_{1}, \ldots, \varphi_{n})$. In the following subsections, we will consider two error functionals used for the solution of the classification problem and will show that the calculation of the derivatives of these functionals is also reduced to the general Formula (12).

5.1. Quadratic Error Functional

Let $w_{0,j}(\mathbf{u}_0)$ be an input distribution corresponding to some image of the *j*-th class. The energy values in the regions G_k of the output plane in this case have the form

$$E_{k} = \iint I_{n+1}(\mathbf{u}_{n+1})\chi_{k}(\mathbf{u}_{n+1})d^{2}\mathbf{u}_{n+1}, \ k = 1, \dots, L.,$$
(16)

where $\chi_k(\mathbf{u}_{n+1})$ is the indicator function of the region G_k . For recognizing (i.e., correctly classifying) the input image $w_{0,j}(\mathbf{u}_0)$, it is necessary for the energy E_j in the corresponding region G_j to be "large", with the energies in the rest of the regions being close to zero. Accordingly, the following quadratic functional can be utilized as the error functional for recognizing an input image of the *j*-th class:

$$\varepsilon^{j}(\varphi_{1},\ldots,\varphi_{n})=\sum_{k=1}^{L}\left(E_{k}-E_{\max}\delta_{k,j}\right)^{2},$$
(17)

where $\delta_{k,j}$ is the Kronecker delta and E_{\max} is the maximum possible energy value. As for the E_{\max} value, one can, for example, use the total energy of the input distribution $E_{\max} = \iint |w_{0,j}(\mathbf{u}_0)|^2 d^2 \mathbf{u}_0$.

Let us demonstrate that the calculation of the derivatives of the functional $\varepsilon^{j}(\varphi_{1},...,\varphi_{n})$ is very similar to the calculation of the derivatives of the "general" functional of Equation (2) considered in Section 3. Indeed, let $\Delta_{m}\varepsilon^{j}(\varphi_{1},...,\varphi_{n})$ be the increment of the functional of Equation (17) caused by an increment $\Delta\varphi_{m}$ of the function φ_{m} . According to Equation (17), this increment has the form

$$\Delta_{m}\varepsilon^{j}(\varphi_{1},\ldots,\varphi_{n}) = 2\sum_{k=1}^{L} \left(E_{k} - E_{\max}\delta_{k,j} \right) (\Delta_{m}E_{k}) = 2\sum_{k=1}^{L} \left(E_{k} - E_{\max}\delta_{k,j} \right) \iint \Delta_{m} (w_{n+1}(\mathbf{u}_{n+1})w_{n+1}^{*}(\mathbf{u}_{n+1})) \chi_{k}(\mathbf{u}_{n+1})d^{2}\mathbf{u}_{n+1} = 4\operatorname{Re} \left[\iint \left(\Delta_{m}w_{n+1}(\mathbf{u}_{n+1}) \right) w_{n+1}^{*}(\mathbf{u}_{n+1}) \sum_{k=1}^{L} \chi_{k}(\mathbf{u}_{n+1}) \left(E_{k} - E_{\max}\chi_{j}(\mathbf{u}_{n+1}) \right) d^{2}\mathbf{u}_{n+1} \right] = 4\operatorname{Re} \langle \Delta_{m}w_{n+1}(\mathbf{u}_{n+1}), F_{n+1}(\mathbf{u}_{n+1}) \rangle,$$
(18)

where

$$F_{n+1}(\mathbf{u}_{n+1}) = w_{n+1}(\mathbf{u}_{n+1}) \sum_{k=1}^{L} \chi_k(\mathbf{u}_{n+1}) \Big(E_k - E_{\max} \chi_j(\mathbf{u}_{n+1}) \Big).$$
(19)

Similarly to Equation (5), we obtained the increment of the functional $\Delta_m \varepsilon^j(\varphi_1, \ldots, \varphi_n)$ in the form of a scalar product. Thus, the derivatives of the functional $\varepsilon^j(\varphi_1, \ldots, \varphi_n)$ are also defined by Equation (12), where the functions $F_m(\mathbf{u}_m)$ are calculated through the backpropagation of the field of Equation (19).

5.2. Cross Entropy Functional

In the classification problems, the so-called softmax cross entropy is used as a criterion [19,20]. In this case, the following error functional is used for recognizing an input distribution belonging to the *j*-th class $w_{0,j}(\mathbf{u}_0)$:

$$\varepsilon^{j}(\varphi_{1},\ldots,\varphi_{n}) = -\ln\left[\frac{\exp(E_{j})}{\sum\limits_{k=1}^{L}\exp(E_{k})}\right],$$
(20)

where E_k are the energies in the regions G_k defined by Equation (16). Let us note that Equation (20) is close to zero when the energy in the required region G_j is much greater than the energies in the other regions.

Let us consider the increment of the functional (20) caused by an increment $\Delta \varphi_m$ of the function φ_m . By carrying out transformations similar to those presented above, it is easy to obtain the increment of the functional as

$$\Delta_m \varepsilon^j(\varphi_1, \dots, \varphi_n) = 4 \operatorname{Re} \left[\iint (\Delta_m w_{n+1}(\mathbf{u}_{n+1})) F_{n+1}^*(\mathbf{u}_{n+1}) d^2 \mathbf{u}_{n+1} \right] = 4 \operatorname{Re} \langle \Delta_m w_{n+1}(\mathbf{u}_{n+1}), F_{n+1}(\mathbf{u}_{n+1}) \rangle, \quad (21)$$
where

$$F_{n+1}(\mathbf{u}_{n+1}) = 2w_{n+1}(\mathbf{u}_{n+1}) \left[\sum_{k=1}^{L} \exp(E_k) \chi_k(\mathbf{u}_{n+1}) - \chi_j(\mathbf{u}_{n+1}) \sum_{k=1}^{L} \exp(E_k) \right] \left(\sum_{k=1}^{L} \exp(E_k) \right)^{-1}.$$
 (22)

As in the previous case, we obtained the increment of the functional $\Delta_m \varepsilon^j(\varphi_1, ..., \varphi_n)$ as a scalar product. Accordingly, the derivatives of the functional (20) are also defined by Equation (12), where the functions $F_m(\mathbf{u}_m)$ are calculated through the backpropagation of the field of Equation (22).

Thus, the calculation of the phase functions of the cascaded DOE in the problem of image classification consists of the following. For the current batch, the gradient of the functional (14) is calculated, where the calculation of the derivatives of the terms is carried out using Equations (12) and (19) or Equations (12) and (22) depending on the chosen criterion. After calculating the derivatives of the functional (14), the phase functions are corrected using a formula similar to Equation (13). Then, the next batch is considered and the process is repeated. The investigation of the performance of the proposed method in the problem of classifying handwritten digits is presented below in Section 6.2.

6. Numerical Examples of Cascaded DOE Design

In the previous two sections, we considered the application of the proposed gradient method to the problems of calculating cascaded DOEs for the generation of required intensity distributions (for several incident beams) and for optical image classification. In the present section, numerical examples illustrating the performance of the method in the indicated problems are presented. In Section 6.1, we discuss the design of cascaded DOEs for generating different intensity distributions at different angles of incidence of the input beam, and Section 6.2 is dedicated to the design of DOEs for classifying handwritten digits.

6.1. Design of Cascaded DOEs for Focusing Different Incident Beams to Different Regions

Let in the input plane of the cascaded DOE, four input distributions $w_{0,j}(\mathbf{u}_0)$, j = 1, 2, 3, 4 be defined, which correspond to Gaussian beams with the radius at the $1/e^2$ level equal to $2\sigma = 2.3$ mm and the wavelength $\lambda = 532$ nm, incident on this plane from different directions. Let the vectors defining the propagation directions of the beams $w_{0,1}(\mathbf{u}_0)$ and $w_{0,2}(\mathbf{u}_0)$ lie in the plane u_0z and make angles $\pm \theta = \pm 0.16^\circ$ with the z axis, and the corresponding vectors of the beams $w_{0,3}(\mathbf{u}_0)$ and $w_{0,4}(\mathbf{u}_0)$ lie in the plane v_0z and also make angles $\pm \theta$ with the z axis. The complex amplitudes of these beams in the plane z = 0 have the form

$$w_{0,1}(\mathbf{u}_0) = \exp(-\frac{\mathbf{u}_0^2}{2\sigma^2}) \exp(+i\frac{2\pi}{\lambda}u_0\sin\theta), \quad w_{0,2}(\mathbf{u}_0) = \exp(-\frac{\mathbf{u}_0^2}{2\sigma^2}) \exp(-i\frac{2\pi}{\lambda}u_0\sin\theta), \\ w_{0,3}(\mathbf{u}_0) = \exp(-\frac{\mathbf{u}_0^2}{2\sigma^2}) \exp(-i\frac{2\pi}{\lambda}v_0\sin\theta), \quad w_{0,4}(\mathbf{u}_0) = \exp(-\frac{\mathbf{u}_0^2}{2\sigma^2}) \exp(+i\frac{2\pi}{\lambda}v_0\sin\theta).$$
(23)

Let us consider the calculation of cascaded DOEs generating in the output plane z = 600 mm different uniform intensity distributions $I_j(\mathbf{u}_{n+1})$, j = 1, 2, 3, 4 for the incident beams of Equation (23). The four output distributions are centered at the origin of coordinates and correspond to a circle with the diameter of 2.3 mm, contour of a square with the side of 2.3 mm, a cross consisting of two perpendicular segments with the length of 2.3 mm, and a "rotated cross" consisting of two diagonals of the square with the side of 2.3 mm (Figure 2). The thickness of the lines of the required output intensity distributions amounts to 0.2 mm.



Figure 2. Required intensity distributions in the output plane for the incident beams of Equation (14).

We will consider three design examples: a single DOE (located in the plane $z = f_1 = 300 \text{ mm}$) and cascaded DOEs consisting of two DOEs (located at $z = f_1 = 200 \text{ mm}$) and $z = f_2 = 400 \text{ mm}$) and three DOEs (located at $z = f_1 = 150 \text{ mm}$, $z = f_2 = 300 \text{ mm}$, and $z = f_3 = 450 \text{ mm}$). We will define the phase functions in the DOE planes on

 512×512 grids with the step of $d = 18 \mu m$ (these parameters correspond to some of the available spatial light modulators, which can be used as DOEs). In this case, the side length of the square aperture of each DOE amounts to 9.216 mm.

Let us note that at the chosen parameters, the incident beams strongly overlap in the planes of the DOEs. For example, after the propagation to the plane z = 300 mm, the centers of the beams are displaced from the optical axis (the *z* axis) only by 300 $tg\theta \approx 0.83$ mm, which is significantly smaller than the radius of the beams. This overlap of the incident beams significantly complicates the problem of calculating the cascaded DOE.

The calculation of the phase functions of the DOE was carried out using the gradient method described above. As the error functional, the sum of functionals (14) was used, where the functionals $\varepsilon^{j}(\varphi_{1},...,\varphi_{n})$ representing the difference between the required distributions and the ones generated at the input fields $w_{0,i}(\mathbf{u}_{0})$ were defined as

$$\varepsilon^{j}(\varphi_{1},\ldots,\varphi_{n}) = \iint \left[I_{n+1,j}(\mathbf{u}_{n+1}) - I_{j}(\mathbf{u}_{n+1}) \right]^{2} d^{2} \mathbf{u}_{n+1}.$$
(24)

At each iteration, the derivatives of the error functional were calculated, which, according to Equation (17), correspond to the sum of derivatives of the functionals $\varepsilon^{j}(\varphi_{1}, ..., \varphi_{n})$. The calculation of the derivatives of the functionals $\varepsilon^{j}(\varphi_{1}, ..., \varphi_{n})$ was carried out using Equation (12), where $w_{m}(\mathbf{u}_{m}) = w_{m,j}(\mathbf{u}_{m})$ is the complex amplitude of the field incident on the *m*-th DOE in the case of the direct propagation of the incident beam $w_{0,j}(\mathbf{u}_{0})$, and the function $F_{m}(\mathbf{u}_{m}) = F_{m,j}(\mathbf{u}_{m})$ is calculated through the backpropagation of the field

$$F_{n+1,j}(\mathbf{u}_{n+1}) = 2 [I_{n+1,j}(\mathbf{u}_{n+1}) - I_j(\mathbf{u}_{n+1})] w_{n+1,j}(\mathbf{u}_{n+1}),$$
(25)

where $w_{n+1,j}(\mathbf{u}_{n+1})$ is the complex amplitude of the field in the output plane. In the optimization, the calculation of the functions $w_{m,j}(\mathbf{u}_m)$ and $F_{m,j}(\mathbf{u}_m)$ featured in the expressions for the derivatives of the functionals was based on the numerical calculation of the Fresnel–Kirchhoff integrals using the fast Fourier transform routine.

Figure 3 shows the calculated phase functions of one, two, and three DOEs. For the calculation of each example, 8000 iterations with an exponentially decreasing step were performed (such a number of iterations turned out to be sufficient for the convergence of the method). As initial values, phases equal to zero at the whole aperture were used. The calculation time on a standard PC (Intel Core i9 10920X CPU, 3.50 GHz) was from 30 min for the single DOE to approximately one hour for the cascade of three DOEs.

One can see that the calculated phase functions of the single DOE and of the first DOEs in the cascaded structures are close to zero (to the initial phase value) near the edges of the aperture. This is caused by the fact that the amplitude of the fields generated in the plane of the first DOE in the case of the input beams of Equation (23) is close to zero in the peripheral regions of the aperture. Since the derivatives of the error functional are close to zero in the regions with a small amplitude of the field, the phase functions changed only weakly in these regions and remained close to the initial zero value.

Figure 4 shows the calculated intensity distributions generated by the calculated single and cascaded DOEs at different incident beams of Equation (23). In order to characterize the quality of the generated distributions, let us use the energy efficiencies Eff_j and rootmean-square errors δ_j . The energy efficiencies

$$Eff_{j} = \frac{1}{E_{0,j}} \iint_{G_{j}} I_{n+1,j}(\mathbf{u}_{n+1}) d^{2}\mathbf{u}_{n+1}$$
(26)

describe the fraction of the energy $E_{0,j} = \iint |w_{0,j}(\mathbf{u}_0)|^2 d^2 \mathbf{u}_0$ of the *j*-th incident beam, which arrives to the required region $G_j = \{\mathbf{u}_{n+1} | I_j(\mathbf{u}_{n+1}) \neq 0\}$. The root-mean-square errors

$$\delta_{j} = \frac{1}{M_{j}} \sqrt{\frac{1}{\|G\|} \iint_{G} \left[I_{n+1,j}(\mathbf{u}_{n+1}) - Eff_{j}I_{j}(\mathbf{u}_{n+1}) \right]^{2} d^{2} \mathbf{u}_{n+1}}$$
(27)

Cascade of three DOEs

describe the root-mean-square deviation of the distribution $I_{n+1,j}(\mathbf{u}_{n+1})$ generated for the *j*-th incident beam from the required distribution $I_j(\mathbf{u}_{n+1})$ in the region *G* covering all the required regions G_j and corresponding to a square with the side of 3 mm centered at the origin of coordinates. Here, ||G|| is the area of the region *G* and $M_j = \frac{1}{||G||} \iint_{G} I_{n+1,j}(\mathbf{u}_{n+1}) d^2 \mathbf{u}_{n+1}$ is

the average intensity in this region. The values of the energy efficiencies and root-meansquare errors for the designed DOE examples are presented in Figure 4 above each of the calculated intensity distributions.



Figure 3. The calculated phase functions of the single DOE (first row) and of cascaded structures containing two (second row) and three (third row) DOEs.

From Figure 4, it is evident that the quality of the generated distributions increases with an increase in the number of DOEs. In particular, for the single DOE, the required distributions are generated with extremely large root-mean-square errors (being close to or even exceeding 100%) and at relatively low energy efficiencies (less than 54%). For a cascaded structure containing three DOEs, the root-mean-square error significantly decreases (the maximum error, which corresponds to the distribution $I_1(\mathbf{u}_{n+1})$, amounts to 9.8%), and the energy efficiency exceeds 87%.

Thus, the presented examples demonstrate the advantages of cascaded DOEs over single ones in the problem of generating different required intensity distributions for different incident beams and confirm the high performance of the proposed design method.

6.2. Design of Cascaded DOEs for Classifying Handwritten Digits

In this subsection, we will consider the design of DOEs for classifying handwritten digits from the MNIST database [25]. Let us start by considering the case of a single DOE. In the calculations, the input images of the digits were defined on a 56 × 56 grid with the step of $d = 18 \mu m$. The phase function of the DOE was defined on a 512 × 512 grid with the same step. Let the DOE and the output plane be located at $z = f_1 = 300 \text{ mm}$ and $z = f_2 = 600 \text{ mm}$, respectively. Let us note that at the design wavelength $\lambda = 532 \text{ nm}$, the diffraction angle at a pixel of the input distribution amounts to $\varphi = \arcsin(\lambda/d) \approx 1.7^\circ$. In this case, the diffraction pattern from the pixel (with respect to the first minimum) at the distance $f_1 = 300 \text{ mm}$ roughly covers the DOE aperture. In this regard, the chosen parameters ensure the "connection" of each pixel of the input image with all the pixels (grid nodes), at which the phase function of the DOE is defined.

In accordance with the design method, in the output plane, 10 spatially separated square regions G_j with the side length of 0.5 mm were defined, in which maximum energies for different input images of different digits have to be generated (see Figure 5).



Figure 4. Calculated intensity distributions generated for each of the incident beams of Equation (23) by the designed single DOE (first row) and cascaded structures consisting of two (second row) and three (third row) DOEs.



Figure 5. Regions G_j in the output plane, in which maximum energies for the input images of different digits have to be generated.

In the calculation, a training set containing 60,000 images of digits from the MNIST database was used. The DOE was calculated using batch training, with each batch containing 60 randomly chosen digits. As the error functionals, the quadratic error (QE) functional of Equations (14) and (17) and the softmax cross entropy (SCE) functional of Equations (14) and (20) were used. As the initial approximation for the DOE phase function, a random phase from the range $[0, 2\pi)$ was chosen. In the DOE calculation, 10 epochs were performed, which takes approximately 7 min on a NVIDIA GTX 1070 8 Gb graphics card. Under an epoch, we understand the training of the DOE on 1000 batches containing all the images from the training set. The phase functions of the DOEs calculated using the QE and SCE criteria are shown in Figure 6.



Figure 6. Phase functions of DOEs calculated using the quadratic error criterion (**a**) and the softmax cross entropy criterion (**b**).

After training, "blind" testing of the performance of the calculated DOEs was performed using a test set consisting of 10,000 images not included in the training set. For each image from the test set, the generated intensity distribution was simulated, the energies (16) in the regions G_j were calculated, and then the input digit was determined using the maximum energy value. The testing results represented as confusion matrices and energy distribution matrices are represented in Figure 7. The element (*i*,*j*) of the confusion matrix contains the percentage of cases, in which an input image of the digit *j* was recognized as the digit *i*. Accordingly, the diagonal elements of these matrices contain the percentage of the correct classifications. Similarly, the element (i,j) of the energy distribution matrix contains the averaged energy (in percent) in the region G_i at an input image of the digit *j*. The diagonal elements of this matrix correspond to mean energies (in percent) in the "correct" regions corresponding to each digit.

For the DOE calculated using the QE criterion (Figure 6a), the accuracy of the digit recognition varies from 93.9% for the digit "9" to 99.2% for the digit "1". The overall classification accuracy (i.e., the ratio of the number of correctly recognized digits to the total amount of digits in the test set) amounts to 97.2%. For the DOE calculated using the SCE criterion (Figure 6b), the accuracy varies from 91.9% for the digit "8" to 99.5% for the digit "0", and the overall accuracy equals 96.8%. Let us note that the achieved classification accuracy values are quite high for single DOEs. For the sake of comparison, the overall classification accuracies in Refs. [3,5,21] achieved using cascaded structures containing 5–10 DOEs vary from 91.8% to 93.4%.

	Confusion Matrix											Energy Distribution Matrix (%)											
	Overall accuracy = 97.2%												Minimum contrast = 0.11										
QE		0	98.6	0.0	0.6	0.0	0.1	0.3	0.8	0.0	0.3	0.8	0	15.6	8.4	9.2	8.2	8.4	8.6	10.1	8.3	9.2	8.3
		1	0.0	99.2	0.0	0.0	0.1	0.2	0.3	0.1	0.0	0.8	1	8.1	16.6	8.1	8.1	9.0	6.6	8.2	8.8	7.7	7.4
		2	0.2	0.2	96.5	0.3	0.2	0.1	0.1	1.1	0.4	0.1	2 and	10.4	8.9	16.2	10.8	8.6	8.6	10.2	10.0	10.1	8.4
	its	3	0.0	0.3	0.3	97.9	0.0	1.8	0.0	0.5	0.9	1.1	but b	8.6	9.2	11.7	16.2	7.7	11.8	7.8	10.6	11.0	10.0
	Dig	4	0.3	0.0	0.4	0.1	97.9	0.0	0.5	0.0	0.6	1.0	lino 4	8.4	11.1	8.3	7.9	16.1	8.3	11.3	10.0	9.3	11.7
	cted	5	0.0	0.0	0.0	0.3	0.0	95.7	0.3	0.1	0.3	0.3	ons in the	9.2	7.9	8.6	11.5	8.0	16.3	9.5	9.1	10.2	9.0
	redi	6	0.4	0.1	0.2	0.0	0.9	0.8	97.7	0.0	0.6	0.3		10.9	9.1	8.9	7.5	10.1	9.7	16.3	6.0	9.4	8.3
	Р	7	0.1	0.2	1.3	0.7	0.1	0.3	0.0	98.0	0.6	1.4	egio	9.0	11.9	9.7	8.9	10.9	8.7	7.7	17.4	7.9	11.7
		8	0.3	0.1	0.8	0.7	0.2	0.6	0.2	0.3	95.9	0.4	≈ 8	9.9	9.4	10.5	10.7	9.7	11.7	10.1	8.5	14.7	10.6
		0	0.1	0.0	0.0	0.0	0.5	0.1	0.0	0.0	03	93.9	9	9.9	7.7	8 7	10.7	11.5	0.7	8 0	11.4	10.6	14.5
		,	0.1	1	o.o 2	3	4	5	6	7	8	9		9.0	1	o.7	3	4	9.1 5	6.9	7	8	9
	(a))	v	•	-	U	Input	Digit	s	,	Ū	-	(b)	v		-	Ŭ	Input	: Digit	s	,	Ū	-
									Ov	erall aco	curacy =	= 96.8%			Minimum contrast = 0.1								= 0.17
		0	99.5	0.1	0.5	0.2	0.1	0.4	0.8	0.3	2.3	1.6	0	16.4	9.7	9.7	9.5	9.7	9.4	10.0	9.4	9.8	9.4
		1	0.0	98.9	0.3	0.0	0.2	0.0	0.2	0.4	0.2	0.6	1	9.5	15.8	9.4	9.5	9.6	8.9	9.5	9.2	9.2	8.7
		2	0.1	0.4	97.1	1.1	0.1	0.1	0.2	0.9	0.6	0.2	put plane	9.8	9.6	17.0	9.9	8.8	8.5	9.6	10.1	9.6	8.7
	gits	3	0.0	0.1	0.7	95.8	0.0	0.4	0.1	0.1	1.3	0.7		9.1	9.3	10.3	15.7	8.5	10.1	9.2	9.5	9.8	9.8
[+]	d Dig	4	0.0	0.0	0.0	0.1	97.1	0.0	0.2	0.2	0.8	1.0	tno a	8.5	9.5	8.4	8.6	15.9	8.4	9.7	8.9	9.1	10.3
SCI	licte	5	0.0	0.1	0.0	1.3	0.0	98.0	0.2	0.2	1.0	0.7	the 2	9.3	9.4	8.0	10.8	9.2	16.8	9.9	9.3	9.7	9.7
	Pred	6	0.2	0.3	0.3	0.2	0.9	0.7	98.1	0.0	1.2	0.2	i suc	10.1	9.9	9.4	8.5	10.0	10.1	16.7	8.5	9.9	9.4
		7	0.1	0.1	0.5	1.0	0.2	0.2	0.0	97.0	0.4	0.6		9.4	10.3	10.2	9.1	9.6	9.3	8.5	16.7	8.8	10.2
		8	0.1	0.1	0.7	0.3	0.3	0.1	0.1	0.1	91.9	0.2	8	8.5	9.2	9.3	9.0	8.6	9.6	8.7	8.1	14.3	9.2
		9	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.9	0.2	94.3	9	9.3	7.2	8.4	9.4	10.1	9.1	8.2	10.1	9.7	14.6
	ഹ		0	1	2	3	4	5	6	7	8	9	(d)	0	1	2	3	4	5	6	7	8	9
	(1)	·			Input Digits							(u)	Input Digits										

Figure 7. Confusion matrices and energy distribution matrices for the DOEs calculated using the quadratic error criterion (**a**,**b**) and the softmax cross entropy criterion (**c**,**d**).

As it was noted above, for the DOE calculated using the SCE criterion, the overall classification accuracy turned out to be 0.4% lower. At the same time, the energy distribution matrix for this DOE is better. Indeed, from the practical point of view, an important

parameter is the contrast value, which shows, how much the energy in the required region exceeds the energy values in the other regions. Let us define the contrast for the digit *i* as

$$\gamma_{\min,i} = \frac{I_{i,i} - \max_{j \neq i} I_{j,i}}{I_{i,i} + \max_{j \neq i} I_{j,i}},$$
(28)

where $I_{i,j}$, i, j = 0, ..., 9 are the elements of the energy distribution matrix. For robust determination of the "true maxima", it is necessary for the contrast values $\gamma_{\min,i}$ to exceed 0.1. According to the energy distribution matrix shown in Figure 7b and corresponding to the DOE calculated using the QE criterion, the minimum contrast is achieved for the digit "9" and amounts to $\gamma_{\min} \approx 0.11$. For the energy distribution matrix of Figure 7d corresponding to the DOE calculated using the SCE criterion, the minimum contrast is also achieved for the digit "9" but is somewhat greater: $\gamma_{\min} \approx 0.17$.

As an example, Figure 8 shows a typical input image of the digit "3" and the corresponding energy distribution demonstrating a correct digit recognition.



Figure 8. Input image of the digit "3" (a) and the corresponding energy distribution (b).

Then, using the QE and SCE criteria, we designed cascaded DOEs comprising two DOEs located in the planes $z = f_1 = 300$ mm and $z = f_2 = 600$ mm. The output plane was located at z = 900 mm. All the other parameters (discretization, wavelength, and aperture sizes) coincide with the parameters of the examples considered above. The phase functions of the cascaded DOEs calculated after 10 epochs are shown in Figure 9.

The confusion matrices and the energy distribution matrices for the designed cascaded DOEs are presented in Figure 10. As before, the DOE performance was evaluated on a test set containing 10,000 images not included in the training set. By comparing the confusion matrices for single and cascaded DOEs (Figures 7a,c and 10a,c), one can see an increase in the classification accuracy. The overall accuracy values for the cascaded DOEs calculated using the QE and SCE criteria amount to 98.0% and 97.6%, respectively. Thus, for the considered example, the increase in the classification accuracy achieved by using a cascaded structure containing two DOEs equals 0.8%. The energy distribution matrices for the cascaded DOEs (Figure 10b,d) are also improved. In particular, minimum contrast values for the cascaded DOEs, which are also achieved for the digit "9", amount to 0.19 and 0.31 for the QE and SCE criteria, respectively. These contrast values are more than 1.7 times greater than those for single DOEs.



Figure 9. Phase functions of the cascaded DOEs calculated using the quadratic error criterion (**a**,**b**) and the softmax cross entropy criterion (**c**,**d**).

Let us note that a further increase in the number of DOEs leads to only a marginal increase in the classification accuracy but enables improving the contrast values. In particular, for a cascaded structure consisting of three DOEs calculated using the SCE criterion, the minimum contrast amounts to 0.55, which is significantly greater than the value of 0.31 provided by the cascaded structure of two DOEs.

Another way to increase the DOE performance consists in increasing the number of the optimized parameters, which can be achieved by decreasing the step of the grid, at which the phase functions of the DOEs are defined. For example, a single DOE with the step size $d = 4 \,\mu\text{m}$ (and the rest of the parameters coinciding with those of the single DOE examples considered above) calculated using the QE criterion provides the overall accuracy of 97.9% and minimum contrast of 0.16, which is considerably better than in the case of a single DOE with the larger step size of 18 μm [see Figure 7a,b]. It is worth noting that this result is comparable with the performance of the cascaded structure of two DOEs with the 18 μm step size [see Figure 10a,b].

From the practical point of view, it is important to discuss the misalignment issues, which will inevitably occur when implementing cascaded DOEs (DNNs). It is known that alignment errors smaller than the neuron (DOE pixel) size show a minor influence on the DNN performance [3,21]. When the alignment error is just bigger than the neuron size, the classification accuracy can be drastically reduced. It should also be noted that the longitudinal misalignment usually influences the performance of a DNN much less than the lateral (transverse) one [21].

The cascaded DOEs studied in this work are no exception. In order to estimate the influence of DOE misalignment, as an example, let us consider the cascaded DOE comprising two DOEs and designed using the SCE criterion (Figure 9c,d). The simulation results demonstrate that when the first DOE is laterally displaced by the vectors $\mathbf{\Delta} = (\Delta u_1, \Delta v_1) = (\pm 18, \pm 18) \,\mu\text{m}$ (in the case of a fixed position of the second DOE), the overall classification accuracy remains greater than 95% (i.e., the decrease in the overall

accuracy does not exceed 3%). The minimum contrast in this case also stays acceptable and exceeds 0.12. At further increase of the lateral displacement, the accuracy and contrast decrease more significantly: for example, at the lateral displacement $\Delta = (36, 36) \mu m$, the overall accuracy and minimum contrast amount to 90.9% and 0.09, respectively. The lateral displacement of the second DOE influences the performance somewhat less, e.g., at the (36, 36) μm displacement, the overall accuracy equals 96.3%, whereas the minimum contrast is 0.13. Similar to the results presented in [21], the longitudinal misalignment is much less critical: for example, the displacement of each of the DOEs along the optical axis by 200 μm leads to a decrease in the overall efficiency not exceeding 0.1% at virtually the same contrast.

	Confusion Matrix											Energy Distribution Matrix (%)											
	Overall accuracy = 98.0%												Minimum contrast = 0.19										
QE		0	98.6	0.0	0.1	0.0	0.1	0.2	0.4	0.0	0.3	0.2	0	17.5	7.4	8.1	6.9	7.1	7.6	9.3	7.3	7.8	7.3
		1	0.0	99.6	0.2	0.0	0.0	0.0	0.3	0.3	0.1	0.4	1	7.9	19.6	8.2	7.6	9.0	6.2	8.6	8.0	7.6	7.1
		2	0.1	0.2	98.9	0.9	0.3	0.2	0.0	0.9	0.5	0.2	2 all	11.1	8.9	20.3	10.7	7.8	8.3	10.7	12.3	10.4	7.7
	its	3	0.1	0.1	0.4	97.2	0.0	0.6	0.1	0.3	0.8	0.5	but b	8.2	8.5	11.8	19.3	7.5	12.1	8.0	10.7	11.0	10.4
	l Dig	4	0.1	0.0	0.1	0.0	97.6	0.0	0.4	0.0	0.7	0.7	Regions in the out 8 6 7 9 5 9 5 4 1 1 1 1 1 1 1 1 1 1 1 1 1	7.7	11.8	7.4	6.8	19.5	7.5	11.4	8.2	8.8	12.2
	icted	5	0.4	0.0	0.0	0.9	0.0	98.7	0.6	0.0	0.6	0.4		9.7	6.5	8.0	12.2	7.2	20.3	9.9	8.2	10.3	8.8
	Pred	6	0.3	0.1	0.0	0.0	0.4	0.2	98.1	0.0	0.4	0.2		10.5	8.4	9.1	6.7	9.4	8.8	18.6	4.3	8.8	6.6
		7	0.1	0.0	0.2	0.3	0.1	0.1	0.0	98.0	0.6	0.3		9.2	11.5	9.4	8.5	10.5	8.3	6.4	20.7	7.5	11.7
		8	0.3	0.0	0.1	0.3	0.2	0.0	0.0	0.1	95.2	0.2		9.2	9.6	9.6	10.2	8.8	11.0	8.9	7.7	16.6	10.1
		9	0.0	0.0	0.0	0.4	1.3	0.0	0.0	0.5	0.7	96.9		9.1	7.8	8.2	11.0	13.1	10.0	8.1	12.7	11.3	17.9
	(a))	0	1	2	3	4 Input	5 Digit	6	7	8	9	(b)	0	1	2	3	4 Input	5 Digit	6	7	8	9
											= 97.6%												
		0	98.9	0.0	0.3	0.0	0.1	0.4	0.7	0.0	0.6	0.3	0	22.3	8.9	8.5	8.3	8.5	8.3	9.4	8.6	9.0	8.5
		1	0.1	99.1	0.0	0.0	0.2	0.0	0.2	0.5	0.0	0.4	1 5 output plane 7 contraction of the second	8.9	21.8	8.5	8.4	8.8	7.8	8.4	8.8	8.5	8.2
		2	0.2	0.4	97.8	0.6	0.4	0.3	0.0	1.2	0.5	0.3		9.6	8.2	23.7	9.8	7.9	8.4	9.3	10.0	9.4	7.7
	its	3	0.0	0.0	0.6	97.6	0.0	1.1	0.1	0.2	0.5	1.2		8.2	9.2	10.1	22.5	8.1	10.7	8.4	8.6	9.7	9.4
	l Dig	4	0.0	0.1	0.1	0.0	97.1	0.0	0.2	0.0	0.6	0.6		8.1	9.3	8.1	8.0	21.7	7.4	9.4	7.9	8.7	10.4
SCF	licted	5	0.0	0.0	0.0	0.3	0.0	97.2	0.3	0.0	0.3	0.1	n the	8.2	8.0	7.1	9.5	7.6	21.5	9.2	7.7	9.0	8.2
	Pred	6	0.3	0.2	0.1	0.0	0.6	0.2	98.3	0.0	0.3	0.2	i suo	9.3	9.1	8.6	8.0	9.5	9.2	22.3	8.4	9.0	8.4
		7	0.2	0.0	0.5	0.8	0.1	0.2	0.0	97.2	0.5	0.6	Regi	8.6	9.4	9.1	7.9	8.9	8.1	7.8	22.9	7.5	10.3
		8	0.2	0.2	0.7	0.4	0.2	0.1	0.1	0.3	96.1	0.3	8	8.1	8.3	8.5	8.5	8.2	9.3	8.0	7.0	19.7	9.0
		9	0.1	0.0	0.0	0.3	1.2	0.3	0.0	0.7	0.5	96.0	9	8.6	7.7	7.6	8.9	10.7	9.2	7.9	10.1	9.7	20.0
	(c))	0	1	2	3	4 Input	5 Digit	6 s	7	8	9	(d)	0	1	2	3	4 Input	5 t Digit	6 .s	7	8	9

Figure 10. Confusion matrices and energy distribution matrices for cascaded DOEs calculated using the quadratic error criterion (**a**,**b**) and the softmax cross entropy criterion (**c**,**d**).

7. Conclusions

In this work, we presented a gradient method for calculating cascaded DOEs. Using the unitarity property of the operator of light propagation through the cascaded DOE, we obtained explicit expressions for the derivatives of the general error functional with respect to the phase functions of the cascaded DOE. We considered the application of the gradient method to the problem of focusing different incident beams to regions with different intensity distributions and to the problem of image classification. The presented description of the gradient method unifies the problems of designing cascaded DOEs for focusing laser radiation and for image classification within a general methodological approach, in which the calculation of the derivatives of the error functionals is reduced to the same formula.

Using the proposed gradient method, we designed single and cascaded DOEs for solving the problem of focusing different incident beams on different regions and the problem of classifying handwritten images. The presented numerical simulation results demonstrate the high performance of the proposed method. In particular, it was shown that a single DOE enables solving the classification problem with an accuracy of approximately 97%, whereas a cascaded structure containing two DOEs provides a 98% accuracy.

In the opinion of the authors, the presented gradient method can be generalized to the case of generating required intensity distributions in the case of several incident beams with different wavelengths. The design and investigation of such DOEs operating at several wavelengths will be the subject of further research.

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

Funding: This work was performed within the State assignment of Federal Scientific Research Center "Crystallography and Photonics" of Russian Academy of Sciences (development of a gradient method for designing cascaded DOEs) and was supported by the Ministry of Science and Higher Education of the Russian Federation (State assignment for research to Samara University (laboratory "Photonics for Smart Home and Smart City", project FSSS-2021-0016), design and investigation of DOEs for the problem of classifying images of handwritten digits).

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 conflict of interest.

References

- Zhang, J.; Pégard, N.; Zhong, J.; Adesnik, H.; Waller, L. 3D computer-generated holography by non-convex optimization. *Optica* 2017, 4, 1306–1313. [CrossRef]
- 2. Wang, H.; Piestun, R. Dynamic 2D implementation of 3D diffractive optics. Optica 2018, 5, 1220–1228. [CrossRef]
- 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] [PubMed]
- 4. Schmidt, S.; Thiele, S.; Toulouse, A.; Bösel, C.; Tiess, T.; Herkommer, A.; Gross, H.; Giessen, H. Tailored micro-optical freeform holograms for integrated complex beam shaping. *Optica* **2020**, *7*, 1279–1286. [CrossRef]
- 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]
- Banerji, S.; Meem, M.; Majumder, A.; Vasquez, F.G.; Sensale-Rodriguez, B.; Menon, R. Imaging with flat optics: Metalenses or diffractive lenses? *Optica* 2019, 6, 805–810. [CrossRef]
- Banerji, S.; Sensale-Rodriguez, B. Computational design framework for efficient, fabrication error-tolerant, planar THz diffractive optical elements. *Sci. Rep.* 2019, 9, 5801. [CrossRef]
- 8. Gerchberg, R.; Saxton, W. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **1972**, *35*, 237.
- 9. Fienup, J.R. Phase retrieval algorithms: A comparison. Appl. Opt. 1982, 21, 2758–2769. [CrossRef]
- 10. Soifer, V.A.; Kotlyar, V.V.; Doskolovich, L.L. Iterative Methods for Diffractive Optical Elements Computation; Taylor & Francis: London, UK; Bristol, UK, 1997.
- 11. Shechtman, Y.; Eldar, Y.C.; Cohen, O.; Chapman, H.N.; Miao, J.W.; Segev, M. Phase retrieval with application to optical imaging. *IEEE Signal Process. Mag.* 2015, *32*, 87–109. [CrossRef]
- 12. Latychevskaia, T. Iterative phase retrieval in coherent diffractive imaging: Practical issues. *Appl. Opt.* **2018**, *57*, 7187–7197. [CrossRef] [PubMed]

- 13. Ripoll, O.; Kettunen, V.; Herzig, H.P. Review of iterative Fourier transform algorithms for beam shaping applications. *Opt. Eng.* **2004**, *43*, 2549–2556.
- 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]
- 15. Gülses, A.A.; Jenkins, B.K. Cascaded diffractive optical elements for improved multiplane image reconstruction. *Appl. Opt.* **2013**, 52, 3608–3616. [CrossRef]
- 16. Deng, X.; Chen, R.T. Design of cascaded diffractive phase elements for three-dimensional multiwavelength optical interconnects. *Opt. Lett.* **2000**, *25*, 1046–1048. [CrossRef] [PubMed]
- 17. Yan, T.; Wu, J.; Zhou, T.; Xie, H.; Xu, F.; Fan, J.O.; Fang, L.; Lin, X.; Dai, Q. Fourier-space Diffractive Deep Neural Network. *Phys. Rev. Lett.* **2019**, *123*, 023901. [CrossRef] [PubMed]
- 18. Zheng, S.; Xu, S.; Fan, D. Orthogonality of diffractive deep neural network. Opt. Lett. 2022, 47, 1798–1801. [CrossRef]
- 19. Chang, J.; Sitzmann, V.; Dun, X.; Heidrich, W.; Wetzstein, G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* **2018**, *8*, 12324. [CrossRef]
- Liu, C.; Ma, Q.; Luo, Z.J.; Hong, Q.R.; Xiao, Q.; Zhang, H.C.; Miao, L.; Yu, W.M.; Cheng, Q.; Li, L.; et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat. Electron.* 2022, 5, 113–122. [CrossRef]
- 21. 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]
- Ferdman, B.; Saguy, A.; Xiao, D.; Shechtman, Y. Diffractive optical system design by cascaded propagation. *Opt. Express* 2022, 30, 27509–27530. [CrossRef]
- 23. 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] [PubMed]
- 24. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv 2015, arXiv:1412.6980.
- 25. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [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.