Deep Learning: A Rapid and Efficient Adaptive Design Approach for Phase-Type Optical Needle Modulators

Simo Wang, Jiangyong Zhang, Fanxing Li, Jupu Yang, Jixiao Liu, and Wei Yan

***Abstract*—The radially polarized beams are modulated by phase-type optical needle modulators can be tightly focused to create needle-like focused beams, which are called optical needles. The use of optical needles with different resolutions and focal depths as direct writing heads for laser direct lithography enables periodic, cross-scale processing of high aspect ratio micro-nano structures with different line widths. The design of the phase-type optical needle modulators is the key to obtain optical needles with different resolutions and focal depths. However, the existing conventional methods for designing phase-type optical needle modulators rely on the physical model for generating optical needles and the defined fitness function, which makes their design time long and not adaptive. Based on the deep learning, a novel phase-type optical needle modulator design (PONMD) approach is proposed in this paper. The results show that the PONMD method takes 0.5526ms to design a phase-type optical needle modulator, and the similarity between the designed and target values is 96.73%. Compared with the conventional methods, the time consumption is reduced by about 8 orders of magnitude, and the similarity is improved by 11.19%. The PONMD approach has the advantages of adaptability, more efficient, less time-consuming, and less computational resource-consuming.**

***Index Terms*—Optical needles, phase-type optical needle modulators, deep learning, laser direct lithography, high aspect ratio micro-nano structures.**

I. Introduction

OPTICAL needles play an important role in many applications such as particle acceleration [1]–[4], optical trapping [5], [6], optical data storage [7], super-resolution microscopy [8], and lithography [9]. The optical needle can be obtained near the focal point by focusing a radially polarized incident beam with a combination of an optical needle modulator and a high numerical aperture (NA) focusing objective. At present, there are three main types of optical needle modulators [10]–[15], and the phase-type optical needle modulators are slightly better due to their high conversion efficiency and low processing difficulty. Moreover, the five-ring band phase-type optical needle modulator is a trade-off between the depth of focus, resolution and optical efficiency of the optical needle [15]. Therefore, the design of a five-ring band phase-type optical needle modulator is used as an example in this paper. For the convenience of expression, all the modulators refer to five-ring band phase-type optical needle modulators in the following.

The optical needle can be regarded as the extension of the subwavelength focused spot along the axial direction, which is a symmetric structure. The structure of the optical needle can be constrained by the intensities along the $z = 0$ and $r = 0$ directions. Moreover, the depth of focus (DOF) and resolution (FWHM) are used to measure the length of the optical needle along the axial direction and the size along the radial direction, which are respectively defined as the full width at half maximum of the normalized intensity profiles along the $z = 0$ and $r = 0$ directions. At present, the conventional methods represented by simulated annealing algorithm and genetic algorithm are the mainstream methods to design modulators. Such methods usually include the definition of fitness function, parameter scanning and other processes [13], [15]–[18], which are not only time-consuming and computational resource-consuming, but also not adaptive.

Machine learning has been widely used as a powerful tool to study the interaction between complex light fields and matter, and it has been used to design structures such as nano-photonic devices [19], metasurfaces [20], and chiral metamaterials [21]. Recently, some scholars have also applied deep learning to the optimal design of optical needles. In 2020, Min Gu *et al.* introduced artificial neural networks into the design of optical needles for the first time [22]. They reported a two-layer fully connected neural network with advanced performance in solving nonlinear problems based on the dipole array principle, enabling the design of optical needles with a central zero intensity point. In 2021, a physically assisted artificial neural network scheme was proposed by Rui Chen et al [23]. They built a four-layer fully connected neural network as a transmission coefficient

Manuscript received 6 April 2022; revised 15 June 2022; accepted 30 June 2022. Date of publication 5 July 2022; date of current version 25 July 2022. This work was supported in part by the Instrument Development of Chinese Academy of Sciences under Grant YJKYYQ20180008, YJKYYQ20200060, YJKYYQ20210041, and YJKYYQ20180006 and in part by the Sichuan Science and Technology Program under Grants 2020JDJQ0007, 2021JDRC0089, 2021JDRC0084, 2022YFG0249, 2022YFG0223 and 2022102. *(Corresponding author: Dr. Wei Yan).*

Simo Wang, Jupu Yang, and Jixiao Liu are with the Institute of Optics and Electronics, Chinese Academy Sciences, Chengdu, Sichuan 610209, China, and also with School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: wangsimo18@mails.ucas.ac.cn; yangjupu20@mails.ucas.ac.cn; liujixiao20@mails.ucas.ac.cn).

Jiangyong Zhang is with the 54th Research Institute of CETC, Shijiazhuang 050000, China (e-mail: zhangjiangyong18@mails.ucas.ac.cn).

Fanxing Li and Wei Yan are with the Institute of Optics and Electronics, Chinese Academy Sciences, Chengdu, Sichuan 610209, China (e-mail: lifanxing15@mails.ucas.ac.cn; yanwei@ioe.ac.cn).

Digital Object Identifier 10.1109/JPHOT.2022.3188303

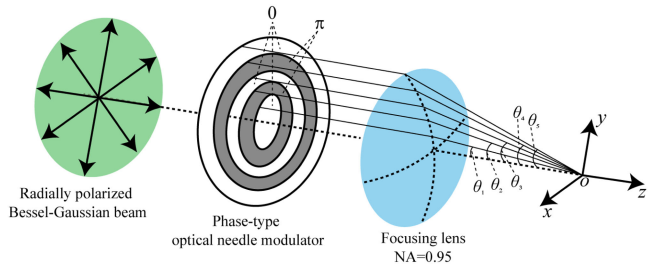

Fig. 1. Schematic diagram of the principle of generating optical needle.

generator to achieve the selection of initial parameters. The method is essentially the same as the existing conventional methods, it only improves the selection of the initial parameters using artificial neural networks and still relies on the physical model that generates the optical needles.

Inspired by this, we proposed a rapid, efficient and adaptive phase-type optical needle modulator design method named PONMD method in this paper. The PONMD method built a PONMD model based on deep learning, which established a mapping relationship between the normalized intensities of the optical needle to the structural parameters of the phase-type optical needle modulator. That is, the PONMD method can design the required structural parameters of the modulator based on the normalized intensities of the target optical needle. It is found that the PONMD method takes only 0.5526ms to design a modulator, and the similarity between the design and target values can reach 96.73%. In addition, considering practical applications, the mathematical model of the ideal optical needle is established by ignoring the side lobes of the optical needle. The normalized intensities of the ideal optical needle is fed into the trained PONMD model, and the desired structural design of the modulator is also achieved. Compared with conventional methods, the PONMD method proposed in this paper is applicable to the rapid and efficient design of modulators required for optical needles with different resolutions and focal depths. It is important for achieving periodic, cross-scale processing of high aspect ratio micro-nano structures with different line widths.

II. Theory and Design

A. Physical Model for Generating Optical Needles

The schematic diagram of the radially polarized Bessel Gaussian beam tightly focused to generate an optical needle after modulation by a modulator is shown in Fig. 1. According to the Richards-Wolf vector diffraction theory, the radial and longitudinal electric field components of the optical needle generated near the focal point of the focusing objective can be expressed as (1) and (2), respectively [15].

$$E_r(r,z) = A\int_0^{\alpha} \cos^{1/2}\theta \sin(2\theta) l(\theta) T(\theta) J_1(kr\sin\theta) \exp(ikz\cos\theta) d\theta \quad (1)$$

$$E_z(r,z) = 2iA\int_0^{\alpha} \cos^{1/2}\theta \sin^2\theta l(\theta) T(\theta) J_0(kr\sin\theta) \exp(ikz\cos\theta) d\theta \quad (2)$$

The numerical aperture of the focusing lens is NA = 0.95, α = arcsin(NA) is the maximum convergence angle of the focusing objective, n is the refractive index of the immersion liquid ($n = 1$), A is the coefficient of power ($A = 1$), $k = 2\pi n/\lambda$ is the wavenumber, J_0 and J_1 denote Bessel functions, and λ is the wavelength of the incident beam. Considering that $E_r(r,z)$ and $E_z(r,z)$ are orthogonal, the total electric energy density can be expressed as $I_t = |E_r(r,z)|^2 + |E_z(r,z)|^2$. The electric field distribution of the Bessel Gaussian beam at the entrance pupil can be expressed as [15]:

$$l(\theta) = \exp\left[-\beta^2\left(\frac{\sin\theta}{\sin\alpha}\right)^2\right] J_1\left(2\gamma\frac{\sin\theta}{\sin\alpha}\right) \quad (3)$$

The modulator can be considered as a multi-ring band binary optical element with alternating phases of 0 and π. The transmittance function of the five-ring band phase-type optical needle modulator can be expressed as:

$$T(\theta) = \begin{cases} 1, \ for\, 0 \le \theta < \theta_1, \theta_2 \le \theta < \theta_3, \theta_4 \le \theta < \alpha, \\ -1, \ for\, \theta_1 \le \theta < \theta_2, \theta_3 \le \theta < \theta_4. \end{cases} \quad (4)$$

θ_i (i = 1–4) are the structural parameters of the five-ring phase optical needle modulator, and θ_i are normalized and expressed as $Q_i = \theta_i/\alpha$.

It is apparent that the structural parameters of each modulator correspond to a set of Q parameters, and a set of Q parameters corresponds to an optical needle of a specific DOF and FWHM. However, an optical needle with a specific DOF and FWHM may correspond to multiple sets of Q parameters, and it is difficult to directly obtain the Q parameters from the optical needle. In addition, the number of combinations of Q parameters is quite large, and it is almost an impossible task to traverse all combinations under the current computational conditions. The following is a description of the methods for designing modulators.

B. Design Methods of Phase-Type Optical Needle Modulators

Fig. 2 is a flow chart of the comparison between the PONMD method and the conventional methods to design modulators. After the PONMD model training is completed, the structural parameters of the modulator can be designed only by inputting the normalized intensities of the target optical needle into the model. In contrast, the design process of conventional methods requires steps such as initialization, definition of fitness functions, parameter sweeps, and calculation of physical models for generating optical needles.

In the following, we compared the two methods in terms of time and computational resources consumption, data collection, and self-adaptability. From the perspective of time-consuming and computational resource-consuming, the integral calculation introduced by the physical model can take up a lot of computational time and computational resources. Conventional methods require hundreds or even thousands of parameter sweeping, while the number of integration calculations in the physical model is T times the number of parameter sweeping, where T is the number of sampling points for the normalized intensities of the optical needle. On the contrary, the PONMD method only needs to focus on the design target, and the structural design of

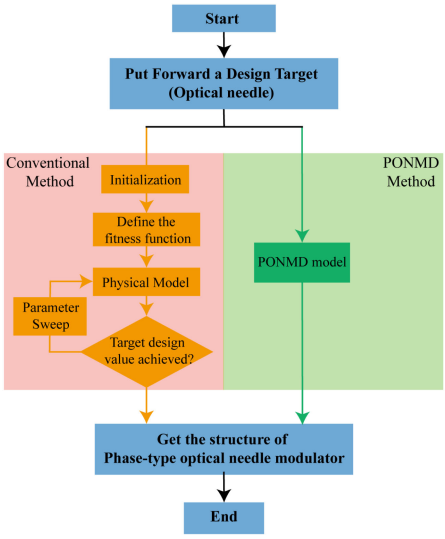

Fig. 2. The contrastive flowchart of design process of PONMD method and conventional method.

the modulator can be completed after one iteration. Therefore, the PONMD approach can significantly improve efficiency.

From the perspective of data collection, the PONMD approach is based on a data-driven, which requires collecting a large amount of data that can describe the problem in advance, and then using some of the data as a training set to train the established PONMD model. In contrast, conventional methods passively generate data during iterative optimization and then determine the strategy for generating new data in the next iteration by means of a defined fitness function, and do not end the generation of new data until an approximate optimal solution is found. Therefore, the data collection of the PONMD method is a one-time cost, while the data collection of the conventional methods is a continuous cost.

In terms of self-adaptability, the conventional methods depend on the fitness function defined according to the characteristics of the target optical needle such as FWHM and DOF, which relies on the physical model for generating optical needles. The fitness function directly determines the optimization result. When the target optical needles are different, the designer needs to change the fitness function and re-optimize the design. Therefore, the conventional methods are not adaptive. However, the PONMD approach is data-driven. Once the model is trained, it can be adapted to the design of modulators required for optical needles with different FWHM and DOF.

In summary, the PONMD method is highly competitive in periodic, cross-scale laser direct writing lithography with the optical needle as the direct writing head, which is important for realizing the processing of high aspect ratio micro-nano structures with different line widths and depths.

III. Theories and Methodology of PONMD Method

A. Data Gathering

The training of the PONMD model requires the collection of the normalized intensities of the optical needles along the $r = 0$ and $z = 0$ directions and the structural parameters of the modulator required for each optical needle to build the data set. We first used the "rand" function in MATLAB to generate 50,000 sets of random matrices representing the Q parameters of the modulator, and then input these matrices into the physical model shown in Fig. 1 to calculate the corresponding focused spot and save the normalized intensities of the focused spot along the $r = 0$ and $z = 0$ directions. We filtered 29,984 sets of data from all the data as a dataset, and all of them were in the shape of optical needles. The dataset was divided by the hold-out method. Among them, the training set and test set account for 20000 and 9984, respectively. The intensities of the optical needles along the $r = 0$ and $z = 0$ directions are extracted and expressed as I_{DOF_ln} and I_{FWHM_ln}, respectively. l is the lth set of data. n is the nth sampling point of each set of data, and the value range of n is [1, 256]. The intensities of the optical needle along the $r = 0$ and $z = 0$ directions are stitched by rows and represented by $I_{lj} = [I_{DOF\_ln}, I_{FWHM\_ln}]$, and the value range of j is [1, 512]. The structural parameters of the phase-type optical needle modulators in the dataset are denoted by Q_{li}, where i denotes the i-th sampling point of each data set, and the value range of i is [1], [4]. At this point, the collection of the data set is completed, and the inputs and outputs of the PONMD model are I_{lj} and Q_{li}, respectively.

B. Test and Comparison

After the training of the PONMD model is completed, Eq. (5) is used to evaluate the PONMD model. It represents the mean absolute error between the designed and target values of the Q parameters of the modulator over the range of the test set.

$$MAE\_Q = \frac{\sum_{l=1}^{L} \frac{\sum_{i=1}^{4} |Q_{li} - Q'_{li}|}{4}}{L}, i = 1, 2, 3, 4 \tag{5}$$

Where Q_{li} and Q'_{li} are the target and designed value of the i-th sampling point of the l-th sample, respectively. L is the total number of samples in the test set. The similarity function S_Q as shown in (6) is defined to measure the similarity between the designed Q parameters and the target Q parameters. A higher similarity S_Q indicates that the designed value is closer to the target value.

$$S\_Q = \frac{1}{1 + MAE\_Q} \times 100\% \tag{6}$$

MAE_Q and *S_Q* can characterize the overall level of the model in designing modulators. However, it is not clear how small *MAE_Q* should be smaller to meet the design requirements. Therefore, we established the test and comparison process shown in Fig. 3 to verify whether the designed modulator meets the requirements.

The testing and comparison process is divided into three steps. The first step is to use the trained PONMD model to design the Q parameters of the modulator through the intensities of the target optical needle. In the second step, the designed Q parameters are substituted into the physical model according to the Richards-Wolf vector diffraction theory to obtain intensities of the designed optical needle. In the third step, the intensities of the target and designed optical needle are compared.

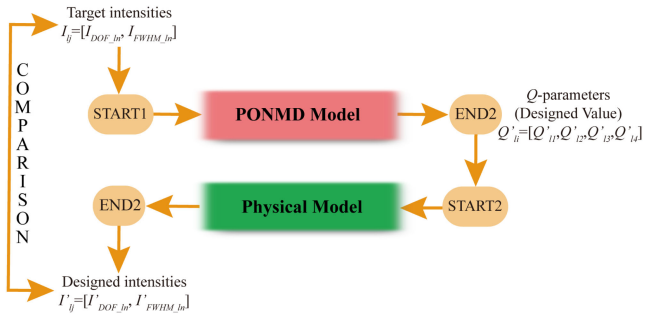

Fig. 3. The test and comparison process.

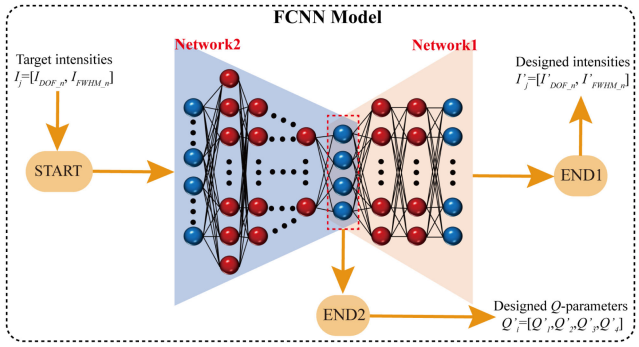

Fig. 4. The structural of the FCNN model.

The mean absolute error of the intensities of the optical needle is defined as:

$$MAE\_I = \frac{\sum_{l=1}^{L} \frac{\sum_{j=1}^{512} |I_{lj} - I'_{lj}|}{512}}{L}, j = 1, 2, \ldots, 512 \quad (7)$$

Where I_{lj} and I'_{lj} are the target and designed value of the j-th sampling point of the l-th sample, respectively. Similarly, define the similarity function S_I as shown in (8) to calculate the similarity of the normalized intensities of the designed optical needle and the target optical needle.

$$S\_I = \frac{1}{1 + MAE\_I} \times 100\% \quad (8)$$

IV. Model Building and Training

This section introduced the building, training, evaluating, and testing and comparing process of the PONMD models. Before that, in order to illustrate the design idea of the PONMD model, the FCNN model based on fully connected neural network built by ourselves was introduced first. Both models were built under a Dell laptop. The image processor, central processor, and operating system are GTX1060 Max-Q, Core i5-8300H, and Ubuntu 16.04, respectively. The integrated development environment is Pycharm, and the framework is Pytorch.

A. Related Work

Fully connected neural networks are widely used because of their simple structure and advanced performance in solving nonlinear problems. Inspired by the research of Min Gu and Rui Chen *et al.* [22], [23], we first established a fully connected neural network (FCNN) model as shown in Fig. 4 to realize the design of the modulator. The FCNN model is composed of Network1 and Network2, and the optimal network structure of the FCNN model is finally obtained through tuning hyperparameters. Network1 is used to replace the physical model for generating the optical needle, and its input and output are the Q parameters of the modulator and the normalized intensities of the optical needle along the $r = 0$ and $z = 0$ directions, respectively. Network1 is a three-layer fully connected neural network with 512 neurons in both hidden layers. Since a modulator corresponds to only one optical needle, the network with a one-to-one mapping relationship is easy to converge in the training process. Network1 avoids the integral operation when calculating the optical needle corresponding to the designed modulator by using the physical model for generating optical needle, which greatly reduces the calculation time and resource consumption. In addition, Network2 is a six-layer fully connected neural network whose input is the normalized intensities of the optical needle and the output is the Q parameters of the modulator. The number of neurons in the five hidden layers are 1024, 666, 512, 200 and 100, respectively.

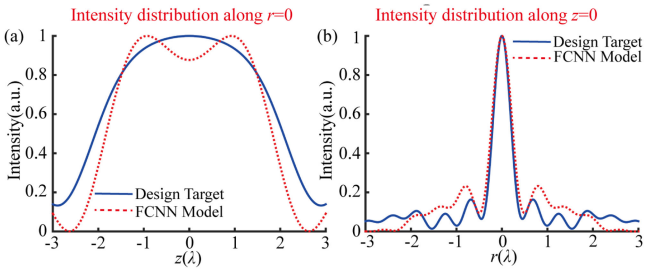

Fig. 5. FCNN Method: The comparison of the intensity distributions along (a) $r = 0$ and (b) $z = 0$ directions.

The experiment found that the MAE_I of Network1 on the test set is 0.0364, and the S_I is 96.49%. The MAE_Q of Network2 on the test set is 0.2773 and the S_Q is 78.29%. The S_Q of Network2 characterizes the ability of the FCNN model to design modulators.

The test and comparison process shown in Fig. 3 is used to determine whether this FCNN model meets the design requirements of the modulator. As shown in Fig. 5, a randomly selected optical needle from the test set is input to the trained FCNN model as the target optical needle, and the intensity distributions of the optical needle corresponding to the designed modulator are shown as the red dashed line. The results show that the FCNN model has a large error in the design of modulators. That is, the FCNN model with MAE_Q of 0.2773 does not satisfy the design requirements of modulators, which mainly because the expressive power of fully connected neural networks is limited when the number of input parameters is too large. For the input data, the connection between each sample point and its surrounding neighboring sample points is relatively tight, and the connection with the far away sample points is small. However, the fully connected neural network connects one neuron to all neurons in the previous layer, and treats each sampling point of the input

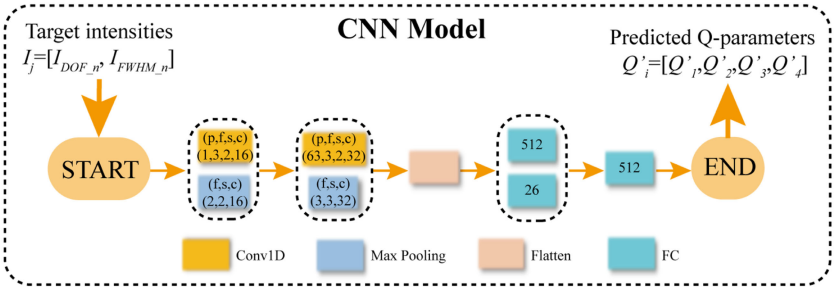

Fig. 6. The structural of the CNN model. The hyperparameters p, f, s and c denote the convolution kernel size, step size, padding and number of channels, respectively. The number on the fully connected layer represents the number of neurons in that layer.

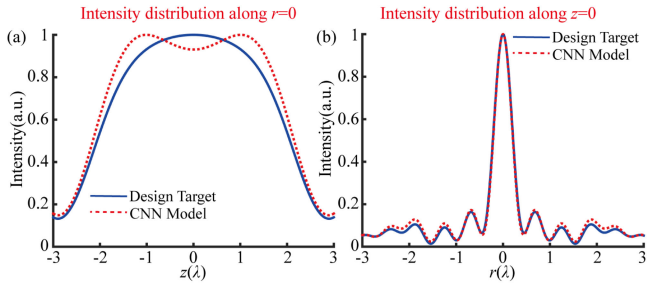

Fig. 7. CNN Method: The comparison of the intensity distributions along (a) $r = 0$ and (b) $z = 0$ directions.

data equally, which is contrary to the characteristics of the input data. Therefore, the FCNN model cannot learn the relationship between the data.

B. PONMD Model

Based on the above research, we found that the design of the modulator can be regarded as a regression problem, which maps a vector of size 1×512 to a vector of size 1×4. Convolutional neural networks have local correlation and stronger expressiveness. We first proposed a CNN model with convolutional neural network as the basic element to design the structure of modulators, whose network structure and hyperparameters are shown in Fig. 6. The input and output are the normalized intensities of the optical needle along the $r = 0$ and $z = 0$ directions and the Q parameters of the modulator, respectively. The experiment found that the *MAE_Q* of the CNN model on the test set is 0.0772, and the *S_Q* is 92.83%.

The CNN model was tested using the intensity of the target optical needle in Fig. 5 as input. The results of the comparison between the intensity of the optical needle corresponding to the phase-type optical needle modulator designed by the CNN model and the intensity of the target optical needle are shown in Fig. 7. The results show that the phase-type optical needle modulator designed by the CNN model has some improvement in similarity compared with the FCNN model, but the predicted optical needle still has a certain gap compared with the target optical needle.

To further improve the design performance of the model, a PONMD model consists of NetworkA and NetworkB is proposed, whose network structure and hyperparameters are shown in Fig. 8. The hyperparameters tuning process of the network is not the focus of this paper, so it will not be repeated here. Only the structure and function of the network will be introduced below.

NetworkA is a convolutional autoencoder [24], [25], which consists of an encoder and a decoder. It enables feature downscaling and feature extraction, and uses the extracted high-level abstract features as input to NetworkB to improve the performance of the whole deep learning model. The input and output of NetworkA are the normalized intensities of the target and designed optical needles, respectively. The encoder maps the original data to the hidden layer feature space and compresses the hidden feature data into a low-dimensional encoding space to extract the most relevant information. Among them, the convolutional layer can extract the characteristics of the input data and reduce the parameters of the network. The max pooling layer can achieve spatial dimensionality reduction and prevent over-fitting to a certain extent. The decoder decompresses the encoded data and reconstructed the same output as the encoder input. The encoder in NetworkA not only summarizes the inherent characteristics of high-dimensional input data, but also provides low-dimensional input data for NetworkB.

The encoder of NetworkA and NetworkB are cascaded using the partial stacking strategy [26]. The input of NetworkB is the output of the encoder of NetworkA, which is a vector of size 1×26. NetworkB consists of a convolution module and a fully connected module. The role of the convolutional layer in the convolution module is to capture the key data content from the input data, and the role of the pooling layer is to compress the convolutional results to obtain more important features and effectively control overfitting. The Batch Normalization layer is added to solve the problem of the internal covariate shift of the convolutional neural network. The Flatten layer stretches the feature map obtained from the last convolutional layer into a vector, and the fully connected module completes the feature-to-target mapping. The output of NetworkB is the Q parameters of the modulator, which is a 1×4 vector. Therefore, the NetworkB is a dimensionality reduction process. Slowly reducing the dimensionality by increasing the number of network layers helps to retain valid information. However, too many layers not only lead to information redundancy, but are also harmful to the training of the model. Therefore, under the premise of

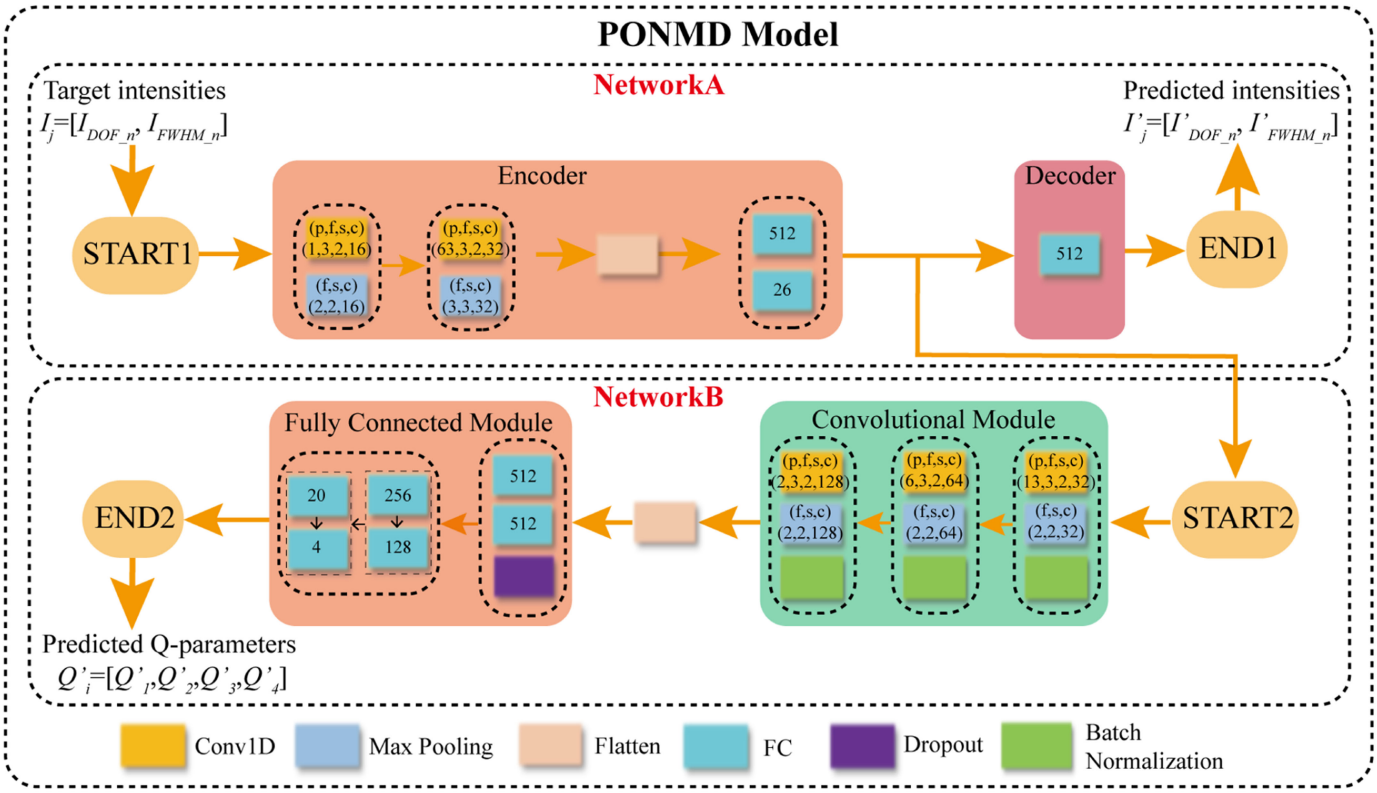

Fig. 8. The structure of PONMD Model.

ensuring design similarity, we reduced the number of network layers of NetworkB. Finally, three convolutional layers and six hidden layers are used for the convolutional module and the fully connected module, respectively.

1) Training and Evaluation: NetworkA and NetworkB in the PONMD model needed to be individually trained with the same batch size of 16 for 500 iterations, which employed an Adam optimization [27] with a learning rate of 0.001. In order to reduce overfitting, we used the dropout to increase regularization, with initial values set at 0.5. After every 50 iterations, the dropout factor, and learning rate were reduced by 0.5, and 0.2, respectively. The smooth L1 loss was use as the cost function in the experiment. The loss functions of NetworkA and NetworkB are shown in (9) and (10), respectively.

$$Loss\_I = \begin{cases} |I_j - I'_j| - 0.5, |I_j - I'_j| > 1 \\ 0.5(I_j - I'_j)^2, |I_j - I'_j| < 1 \end{cases} \tag{9}$$

$$Loss\_Q = \begin{cases} |Q_i - Q'_i| - 0.5, |Q_i - Q'_i| > 1 \\ 0.5(Q_i - Q'_i)^2, |Q_i - Q'_i| < 1 \end{cases} \tag{10}$$

I_j and I_j' denote the intensity of the optical needles, Q_j and Q_j' denote the structural parameters of phase-type optical needle modulators. I_j and Q_j are the actual values, and I_j' and Q_j' are the predictions values of the data labels. Based on the defined loss function, the difference between the model prediction and the data label value can be obtained. Then, the weights of the convolutional, pooling, and fully connected layers in NetworkA and NetworkB are updated and corrected according to the difference using the back propagation algorithm, and the process is repeated until the model meets the demand.

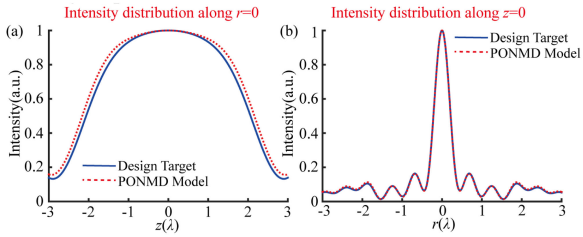

Fig. 9. PONMD Method: The comparison of the intensity distributions along (a) $r = 0$ and (b) $z = 0$ directions.

The experiment found that the MAE_*I* of NetworkA on the test set is 0.0012, and the *S_I* is 99.88%. The *MAE_Q* of NetworkB on the test set is 0.0338 and the *S_Q* is 96.73%. The *S_Q* of NetworkB characterizes the ability of the PONMD model to design phase-type optical needle modulators.

2) Test and Comparison: The test and comparison process shown in Fig. 3 is also used to determine whether the PONMD model meets the design requirements of the modulator. The normalized intensities of the target optical needle shown in Fig. 5 was used as the input, and the results were shown in Fig. 9. It is easy to find that the PONMD model can accurately design the structure of the modulator required by the target optical needle.

3) Design of Modulators Corresponding to Optical Needles With Different DOF and FWHM: After the PONMD model

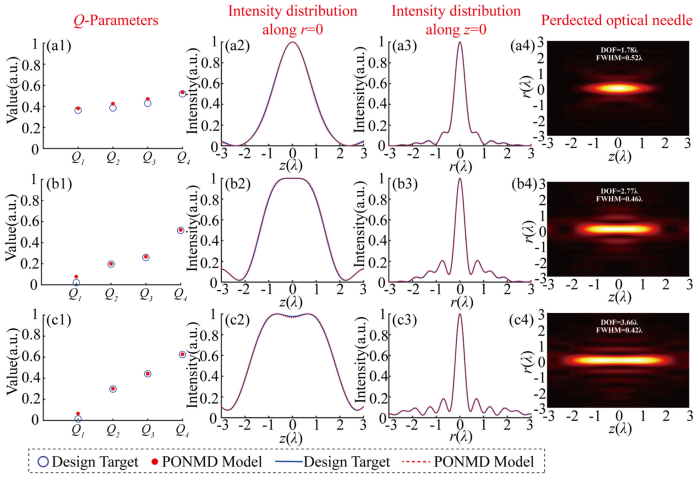

Fig. 10. The design of modulators corresponding to different DOF and FWHM of the optical needles.

TABLE I
LIST OF THE DESIGNED PHASE-TYPE OPTICAL NEEDLE MODULATORS BY PONMD METHOD

Sample	[Q_1, Q_2, Q_3, Q_4]	[Q'_1, Q'_2, Q'_3, Q'_4]	DOF (λ)	FWHM (λ)
1	[0.3636, 0.3859, 0.4289, 0.5188]	[0.3813, 0.4254, 0.4706, 0.5323]	1.78	0.52
2	[0.0208, 0.1965, 0.2596, 0.5163]	[0.0772, 0.1994, 0.2664, 0.5212]	2.77	0.46
3	[0.0145, 0.2976, 0.4421, 0.6255]	[0.0652, 0.3011, 0.4438, 0.6276]	3.66	0.42

was trained, three target optical needle with different DOF and FWHM were arbitrarily taken in the test set to design the structure of the required modulators. The normalized intensity distributions of the target optical needles along the $r = 0$ and $z = 0$ directions are shown by the solid blue line. Through the testing and comparison process shown in Fig. 3, the corresponding optical needles of the modulators designed by the PONMD method are shown in Fig. 10(a4)–(c4), and the normalized intensity distributions of each optical needle along the $r = 0$ and $z = 0$ directions are indicated by the red dashed lines. Comparing the red dashed lines with the blue solid lines, it can be found that the modulators designed by the PONMD model can meet the requirements of the optical needles that with different DOF and FWHM. In addition, it also shows that the PONMD method is adaptive. Applying the PONMD method to laser direct lithography with an optical needle as the direct writing head can improve the processing speed and efficiency of periodic, cross-scale, high aspect ratio micro-nano structures.

From the comparison in Table I, it can be found that the designed values Q'_2–Q'_4 of the modulators are very close to the target values, while Q'_1 may have certain errors in some cases. However, it does not affect the designed optical needle. This is mainly because the effective numerical aperture is proportional to the radius of the pupil at the exit of the objective lens. From the structure of the modulator, Q_1 is close to the center of the focusing objective, while Q_4 is close to the edge. Therefore, the effective numerical apertures corresponding to Q_1, Q_2, Q_3, and Q_4 are increased sequentially. The incident beam at a position with a high effective numerical aperture has a great influence on the optical needle. Therefore, the influence of Q_4 on the optical needle is the greatest, and the influence of Q_1 on the optical needle is the smallest.

V. DESIGN OF MODULATORS CORRESPONDING TO IDEAL OPTICAL NEEDLE

The required for the optical needle in the test set is designed above, but the specific intensity distributions of the target optical needle in the actual design is usually unknown. Therefore, in this section, the optical needle after ignoring the side lobes is regarded as an ideal optical needle. The mathematical model of the ideal optical needle is established, and the required modulator is designed with the normalized intensities of the ideal optical needle as the input of the PONMD model.

A. Mathematical Model of the Ideal Optical Needle

The intensities of the ideal optical needle along the $z = 0$ direction can be approximated as a first-order Gaussian function, and the intensities of the ideal optical needle along the $r = 0$ direction can be regarded as a piecewise function with a flat top and a first-order Gaussian function on both sides. Therefore, the mathematical model of the ideal optical needle is defined as (11) and (12) to fit the normalized intensity distributions along the $z = 0$ direction and the $r = 0$ direction, respectively.

$$I_{Radial} = \exp\left[-\frac{(r - b_R)^2}{2c_R^2}\right] \tag{11}$$

$$I_{axial3} = \begin{cases} \exp\left[-\left(\frac{z+b_A}{c_A}\right)^2\right], z < -b_A \\ 1, -b_A \le z \le b_A \\ \exp\left[-\left(\frac{z-b_A}{c_A}\right)^2\right], z < b_A \end{cases} \tag{12}$$

Where b_R and b_A represent the center coordinates of the peak of the Gaussian curve. c_R and c_A are the width of the Gaussian curve, which directly affects the slope of the edge curve. b_R, b_A, c_R, and c_A are kept in the same units as z and r, and they are all quantified in wavelength λ.

B. Design Example of a Modulator Corresponding to an Ideal Optical Needle

In the following, an ideal optical needle is used as the target optical needle, and the structure of the required modulator is designed by the conventional method and the PONMD method, respectively. The intensity distributions along $r = 0$ and $z = 0$ directions of the target ideal optical needle are shown in Fig. 11(a1)–(a2). The Q parameters of the modulator designed by the conventional method are $[Q'_1, Q'_2, Q'_3, Q'_4] = [0.0691, 0.3035, 0.477, 0.6527]$. Substituting the designed Q parameters of the modulator into the physical model for generating optical needle, the intensity distributions shown by the green dotted line in Fig. 11(b1) and (b2) are obtained. The results show that there is a 0.84λ deviation of the DOF between the designed and target optical needles. Using the same design target, the Q parameters of the modulator designed by the PONMD method

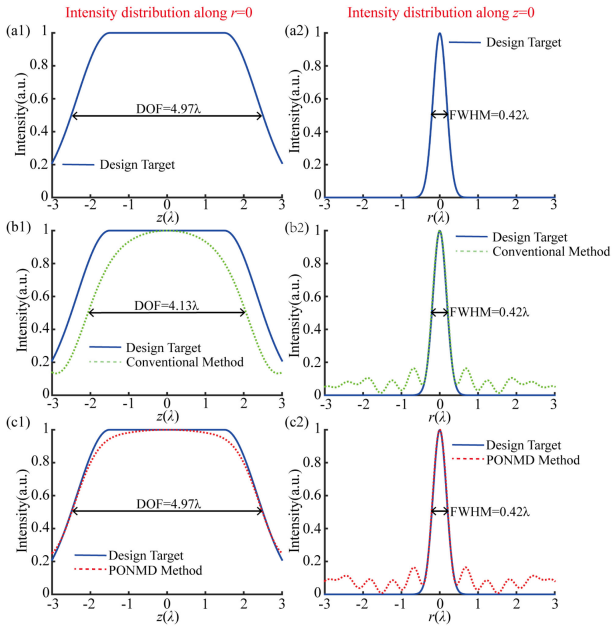

Fig. 11. The comparison of the intensity distributions along $r = 0$ and $z = 0$ directions. (a) Design target, (b) Conventional method, (c) PONMD method.

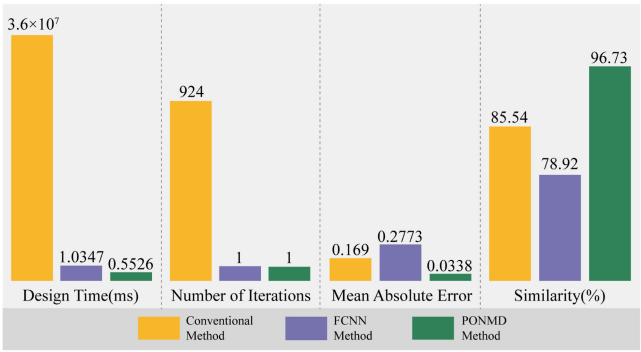

Fig. 12. Comparison of three design methods in terms of design time, number of iterations, mean squared error, and similarity.

are $[Q_1', Q_2', Q_3', Q_4'] = [0.1715, 0.3468, 0.5169, 0.6862]$, and the normalized intensity distributions of the corresponding optical needle are shown in the red curves in Fig. 11(c1)–(c2). Obviously, the designed optical needle obtained by the PONMD method is closer to the target ideal optical needle than the conventional method.

In order to more intuitively illustrate the advantages of the PONMD method proposed in this paper. As shown in Fig. 12, we compared the conventional method, the FCNN method and the PONMD method from four aspects: design time, iterations, mean absolute error and similarity. It can be seen that the time to design a modulator by the PONMD method is nearly 8 orders of magnitude faster than that of the conventional method. In addition, compared with the conventional method and the FCNN method, the mean absolute error of the PONMD method is the smallest, and the similarity is the highest.

This is mainly because the conventional method relies on an adaptation function defined according to the resolution and depth of focus of the target optical needle, without a constraint on the intensity of the optical needle. In contrast, the inputs to both the FCNN model and the PONMD model are 512 neurons, with each neuron representing one sampling point. The first 256 sampling points and the last 256 sampling points represent the normalized intensity distribution of the optical needle along the $z = 0$ and $r = 0$ directions, respectively. From the physical meaning of the sampling points, it can be found that the curve formed by these sampling points can be regarded as a continuous function, and the closer the distance between two sampling points, the stronger the connection. The FCNN model based on fully connected neural network treats all sampling points equally and ignores the local connection relationship between individual sampling points. While the PONMD model based on convolutional neural network is more concerned with the extraction of local features, it can synthesize the local information at a higher level to get the global information. Therefore, the PONMD model fits the intensity distribution characteristics of optical needles more closely. In addition, Network A, consisting of encoders and decoders, can extract high-level abstract features from the input data and use the high-level abstract features as another deep learning model input, finally improving the performance of the whole deep learning model. In conclusion, both the optical physical properties of the optical needle and the experimental results demonstrate that the PONMD method proposed in this paper can achieve rapid, efficient and adaptive design of the modulators.

VI. Conclusion

In this paper, a new method named PONMD is proposed to design phase-type optical needle modulators. The PONMD model enables the design of phase-type optical needle modulators required for optical needles with different focal depths and resolutions. In addition, ignoring the side lobes of the optical needle, the mathematical model of the ideal optical needle is established. It is found by experiments that the PONMD model can also realize the design of the phase-type optical needle modulator required by the ideal optical needle. Compared with the conventional method and the FCNN method, the PONMD method proposed in this paper has the advantages of rapidity, high efficiency and self-adaptation. It only takes 0.5526ms to design a phase-type optical needle modulator required for a target optical needle, and the similarity between the design value and the target value is as high as 96.73%. In addition, the PONMD method can be applied to periodic and cross-scale laser direct writing lithography with optical needles as direct writing heads, which provides a powerful technology to achieve rapid and efficient processing of high aspect ratio micro-nano structures with different line widths and depths.

References

[1] D. Richard, R. D. Romea, and W. D. Kimura, "Modeling of inverse erenkov laser acceleration with axicon laser-beam focusing," *Phys. Rev. D Part Fields*, vol. 42, no. 5, pp. 1807–1818, 1990, doi: 10.1103/physrevd.42.1807.

[2] S. C. Tidwell, G. H. Kim, and W. D. Kimura, "Efficient radially polarized laser beam generation with a double interferometer," *Appl. Opt.*, vol. 32, no. 27, pp. 5222–5229, 1993, doi: 10.1364/AO.32.005222.

[3] J. R. Fontana and R. H. Pantell, "A high-energy, laser accelerator for electrons using the inverse Cherenkov effect," *J. Appl. Phys.*, vol. 54, no. 8, pp. 4285–4288, 1983, doi: 10.1063/1.332684.

[4] L. Cicchitelli, H. Hora, and R. Postle, "Longitudinal field components for laser beams in vacuum," *Phys. Rev. A*, vol. 41, no. 7, pp. 3727–3732, 2008, doi: 10.1103/PhysRevA.41.3727.

[5] Q. Zhan, "Trapping metallic Rayleigh particles with radial polarization," *Opt. Exp.*, vol. 12, no. 15, pp. 3377–3382, 2004, doi: 10.1364/OPEX.12.003377.

[6] Q. Zhan, "Trapping metallic Rayleigh particles with radial polarization: Reply to comment," *Opt. Exp.*, vol. 20, no. 6, pp. 6058–6059, 2012, doi: 10.1364/OE.20.006058.

[7] Y. Zhang and J. Bai, "Improving the recording ability of a near-field optical storage system by higher-order radially polarized beams," *Opt. Exp.*, vol. 17, no. 5, pp. 3698–3706, 2009, doi: 10.1364/OE.17.003698.

[8] K. S. Youngworth and T. G. Brown, "Focusing of high numerical aperture cylindrical-vector beams," *Opt. Exp.*, vol. 7, no. 2, pp. 77–87, 2000, doi: 10.1364/OE.7.000077.

[9] M. Meier, V. Romano, and T. Feurer, "Material processing with pulsed radially and azimuthally polarized laser radiation," *Appl. Phys. A*, vol. 86, no. 3, pp. 329–334, 2007, doi: 10.1007/s00339-006-3784-9.

[10] Y. Zha, J. Wei, H. Wang, and F. Gan, "Creation of an ultra-long depth of focus super-resolution longitudinally polarized beam with a ternary optical element," *J. Opt.*, vol. 15, no. 7, 2013, Art. no. 075703, doi: 10.1088/2040-8978/15/7/075703.

[11] C.-C. Sun and C.-K. Liu, "Ultrasmall focusing spot with a long depth of focus based on polarization and phase modulation," *Opt. Lett.*, vol. 28, no. 2, pp. 99–101, 2003, doi: 10.1364/OL.28.000099.

[12] K. Kitamura, K. Sakai, and S. Noda, "Sub-wavelength focal spot with long depth of focus generated by radially polarized, narrow-width annular beam," *Opt. Exp.*, vol. 18, no. 5, pp. 4518–4525, 2010, doi: 10.1364/OE.18.004518.

[13] K. Huang, P. Shi, X. Kang, X. Zhang, and Y. Li, "Design of DOE for generating a needle of a strong longitudinally polarized field," *Opt. Lett.*, vol. 35, no. 7, pp. 965–967, 2010, doi: 10.1364/OL.35.000965.

[14] L. Jie *et al.*, "Achievement of longitudinally polarized focusing with long focal depth by amplitude modulation," *Opt. Lett.*, vol. 36, no. 7, pp. 1185–1187, 2011, doi: 10.1364/OL.36.001185.

[15] H. Wang, L. Shi, B. Luk'yanchuk, C. Sheppard, and C. T. Chong, "Creation of a needle of longitudinally polarized light in vacuum using binary optics," *Nature Photon.*, vol. 2, no. 8, pp. 501–505, 2008, doi: 10.1038/nphoton.2008.127.

[16] J. Yu, C. Zhou, and W. Jia, "Transverse superresolution with extended depth of focus using binary phase filters for optical storage system," *Opt. Commun.*, vol. 283, no. 21, pp. 4171–4177, 2010, doi: 10.1016/j.optcom.2010.06.055.

[17] C. K. Shi, Z. Q. Nie, Y. T. Tian, C. Liu, Y. C. Zhao, and B. H. Jia, "Super-resolution longitudinally polarized light needle achieved by tightly focusing radially polarized beams," *Optoelectron. Lett.*, vol. 14, no. 1, pp. 1–5, 2018, doi: 10.1007/s11801-018-7162-6.

[18] J. Lin, H. Zhao, Y. Ma, J. Tan, and P. Jin, "New hybrid genetic particle swarm optimization algorithm to design multi-zone binary filter," *Opt. Exp.*, vol. 24, no. 10, pp. 10748–10758, 2016, doi: 10.1364/OE.24.010748.

[19] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, "Deep learning for the design of photonic structures," *Nat. Photon.*, vol. 15, no. 2, pp. 77–90, 2021, doi: 10.1038/s41566-020-0685-y.

[20] Z. Liu, D. Zhu, S. P. Rodrigues, K. T. Lee, and W. Cai, "Generative model for the inverse design of metasurfaces," *Nano Lett.*, vol. 18, no. 10, pp. 6570–6576, 2018, doi: 10.1021/acs.nanolett.8b03171.

[21] W. Ma, F. Cheng, and Y. Liu, "Deep-learning-enabled on-demand design of chiral metamaterials," *ACS Nano*, vol. 12, no. 6, pp. 6326–6334, 2018, doi: 10.1021/acsnano.8b03569.

[22] W. Xin, Q. Zhang, and M. Gu, "Inverse design of optical needles with central zero-intensity points by artificial neural networks," *Opt. Exp.*, vol. 28, no. 26, pp. 38718–38732, 2020, doi: 10.1364/oe.410073.

[23] Z.-Y. Chen, Z. Wei, R. Chen, and J.-W. Dong, "Focus shaping of high numerical aperture lens using physics-assisted artificial neural networks," *Opt. Exp.*, vol. 29, no. 9, pp. 13011–13024, 2021, doi: 10.1364/oe.421354.

[24] E. Kodirov, T. Xiang, and S. Gong, "Semantic autoencoder for zero-shot learning," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit* ., 2017, pp. 4447–4456, doi: 10.1109/CVPR.2017.473.

[25] J. Deng, Z. Zhang, F. Eyben, and B. Schuller, "Autoencoder-based unsupervised domain adaptation for speech emotion recognition," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1068–1072, Sep. 2014, doi: 10.1109/LSP.2014.2324759.

[26] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. BT-3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, May 7–9, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980