



Channeled spectroscopic ellipsometry enabled by physics-informed tandem untrained neural networks

Shilong Yang^a, Xiuguo Chen^{a,b,*}, Wenlong Chen^a, Jing Hu^a, Yifu Wang^a, Shuo Liu^a, Shiyuan Liu^{a,b}

^a State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science & Technology, Wuhan 430074, China

^b Optics Valley Laboratory, Wuhan 430074, China

ARTICLE INFO

Keywords:

Ellipsometry
Channeled spectroscopic ellipsometry
Physics-informed machine learning
Untrained neural networks

ABSTRACT

Ellipsometry is a powerful metrology technique for characterizing the optical properties of various materials. Channeled spectroscopic ellipsometry (CSE) has shown great promise among the different types of ellipsometry due to its simple setup and rapid performance. Furthermore, CSE modulates the polarization parameters of thin films into a spectrum, thus transforming the measurement process into a demodulation problem. However, conventional CSE faces challenges in measurement accuracy and computational efficiency, with strict hardware and calibration requirements. Inspired by physics-informed machine learning, we propose CSE enabled by the physics-informed tandem untrained neural networks (PITUNN), which does not require training, exhibits high computational efficiency and partially alleviates the strict requirements for hardware and calibration accuracy. We also demonstrate the effectiveness of CSE enabled by the PITUNN and its ability to handle system errors and random noise through simulations and experiments on thin films of different thickness and materials.

1. Introduction

Ellipsometry is a noncontact, nondestructive, and widely used metrology technique for characterizing the optical properties of various thin films ranging in thickness from sub-nanometers to several microns intended for a broad array of scientific and industrial applications [1–6]. Among the diverse types of ellipsometry with rotating and active polarization components [4–8], channeled spectroscopic ellipsometry (CSE) can realize measurement in a snapshot manner with a compact size and no active component, thus holding great promise in various applications, including material characterization [9–13], biomedical imaging [14,15], and remote sensing [16,17], among others. The principle of CSE relies on an approximately linear response of the retardance of some multi-order birefringent crystals (e.g., quartz, calcite, and magnesium fluoride), to the wavenumber, which modulates sample polarization parameters (e.g., Mueller matrix elements) into the output intensity spectrum [18–31]. Consequently, the CSE measurement becomes a demodulation problem aimed at recovering sample polarization parameters from the modulated intensity spectrum. Thus, finding ways to accurately demodulate polarization parameters from the modulated

intensity spectrum is a critical issue in CSE and is currently an active research field [23–31]. Thus far, measurement accuracy and efficiency related to demodulation methods have been considered barriers in the practical measurement of CSE. Hence, any advancement in demodulation methods would benefit CSE and expand its potential applications.

The most commonly used demodulation algorithm for CSE, known as “Fourier reconstruction” (FR), uses the Fourier transform to recover sample polarization parameters by separating them into different channels based on their carrier frequencies [23]. Although FR is straightforward and fast, it often suffers from measurement noise, band limitation, and channel cross-talk, resulting in considerable restrictions on the final measurement accuracy [24–31]. Many methods have recently been proposed to address the limitations of FR. One such approach is iterative reconstruction (IR) [24–26], which involves the progressive adjustment of output polarization parameters using optimization algorithms to better align with the theoretical mathematical model. These methods often include regularization terms to control the adjustment process. Although IR has successfully overcome the limitations of FR, it is computationally intensive and suffers from parameter couplings. Thus, some researchers have attempted to overcome these

* Corresponding author at: State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science & Technology, Wuhan 430074, China.

E-mail address: xiuguo@hust.edu.cn (X. Chen).

challenges by incorporating compressed sensing (CS) into IR [27–30]. In particular, they transformed the demodulation problem into a convex optimization problem by parameterizing the polarization parameters. This approach has effectively mitigated parameter couplings, resulting in higher measurement accuracy. However, it requires precise hardware and calibration because it relies heavily on the accuracy of the mathematical model and the calibrated system parameters.

In recent years, machine learning (ML) has evolved rapidly and achieved great success in various optical fields, such as nano-photonic design [32–34] and optical measurement [35–38]. With available data, ML can establish mappings to predict the required parameters for solving specific problems. In the demodulation problem of CSE, many data-driven ML methods [39–43] have greatly enhanced the measurement accuracy. These methods leverage large datasets to train neural networks for mapping the measured spectra (or optical path difference domain spectra) to polarization parameters [39–42]. And there are some intriguing approaches that use neural networks to predict filters for demodulation problems [43]. However, their predictions may not always be physically plausible. It is also difficult to create comprehensive datasets covering all possible conditions in CSE. Furthermore, the generalization of these ML models, which are primarily determined by the training data and the computational costs required for training, can pose challenges in their practical applications. Some studies have suggested that untrained neural networks (UNN) and physics-informed networks are suited for incorporating physical priors during parameter updating [44–51]. Furthermore, the burden of creating datasets and training models is eliminated because UNNs do not require training. Recently, there have been several attempts to solve the demodulation problem in channeled polarimetry, such as deep image prior plus sparsity prior (DIP-SP) and physics-guided neural network (PGNN), which employs a UNN [50,51], thus demonstrating excellent performance. However, due to the similarities in principles, applying DIP-SP in CSE still faces the same limitations as CS, demanding precise hardware and calibration.

In this paper, we introduce CSE enabled by the physics-informed tandem untrained neural networks (PITUNN), which employs three UNNs in tandem to incorporate prior information during updating, achieving higher measurement accuracy in the presence of system errors while maintaining short computational time. The PITUNN partially alleviates the strict requirements for hardware and calibration accuracy in

passes through the polarization state analyzer (PSA), which consists of the multi-order waveplate R2, the polarizer A, and finally the spectrum is obtained by the spectrometer.

As shown in Fig. 1, the system model in terms of the Stokes–Mueller formalism can be expressed as follows

$$\mathbf{S}_{out} = [\mathbf{M}_A \mathbf{T}(\alpha_2)] [\mathbf{T}(-\alpha_4) \mathbf{M}_R(\delta_2) \mathbf{T}(\alpha_4)] \mathbf{M}_S [\mathbf{T}(-\alpha_3) \mathbf{M}_R(\delta_1) \mathbf{T}(\alpha_3)] [\mathbf{T}(-\alpha_1) \mathbf{M}_P] \mathbf{S}_{in} \quad (1)$$

where \mathbf{S}_{in} and \mathbf{S}_{out} respectively represent the Stokes vectors of the incident and outgoing light; \mathbf{M}_S represents the sample Mueller matrix; \mathbf{M}_P and \mathbf{M}_A represent the Mueller matrices of P and A, respectively; $\mathbf{M}_R(\delta_1)$ and $\mathbf{M}_R(\delta_2)$ represent the corresponding Mueller matrices of R1 and R2 with the respective retardances of δ_1 and δ_2 , which are calibrated before measurements [21,30]; $\mathbf{T}(\cdot)$ represents the Mueller rotation transformation matrix; α_1 and α_2 are the transmission-axis orientations of P and A, respectively; and α_3 and α_4 are the corresponding fast-axis orientations of R1 and R2. At every single wavelength, \mathbf{S}_{in} and \mathbf{S}_{out} are 4×1 vectors; \mathbf{M}_P , \mathbf{M}_A , $\mathbf{M}_R(\cdot)$ and $\mathbf{T}(\cdot)$ are 4×4 matrices.

By setting $\alpha_1 = \alpha_4 = \pi/4$, $\alpha_2 = \alpha_3 = \pi/2$ and performing matrix multiplication, we can obtain the output modulated intensity spectrum I_{out} at every single wavenumber for isotropic samples as follows

$$I_{out} = \frac{1}{8} I_{in} \cdot \{ 2 - 2 \cdot N(\sigma) \cdot \cos[\delta_2(\sigma)] + C(\sigma) \cdot \cos[\delta_1(\sigma) - \delta_2(\sigma)] + S(\sigma) \cdot \sin[\delta_1(\sigma) - \delta_2(\sigma)] - C(\sigma) \cdot \cos[\delta_1(\sigma) + \delta_2(\sigma)] - S(\sigma) \cdot \sin[\delta_1(\sigma) + \delta_2(\sigma)] \} \quad (2)$$

where σ represents the wavenumber, I_{in} is the incident intensity, $N(\sigma) = -\cos 2\Psi(\sigma)$, $C(\sigma) = \sin 2\Psi(\sigma) \cdot \cos \Delta(\sigma)$, and $S(\sigma) = \sin 2\Psi(\sigma) \cdot \sin \Delta(\sigma)$, in which $\Psi(\sigma)$ and $\Delta(\sigma)$ are the ellipsometric parameters. In this way, CSE modulates the sample Mueller matrix elements into the spectrum, and the measurement becomes a demodulation problem.

In practice, systematic errors such as the alignment error of each device are inevitable, and these can considerably change the mathematical model of the spectrum [52]. Here, we denote the alignment errors of P, A, R1, and R2 as ε_1 , ε_2 , ε_3 , and ε_4 , respectively. Considering that the alignment errors are typically small values, we can carry out approximate simplification as $\sin \varepsilon_i = \varepsilon_i$, $\cos \varepsilon_i = 1$, $\varepsilon_i \cdot \varepsilon_j = 0$ ($i, j = 1, 2, 3, 4$; $i \neq j$). In this case, the mathematical model can be rewritten as follows

$$I_{out} = \frac{1}{8} I_{in} \cdot \{ 2 - 4 \cdot N(\sigma) \cdot (\varepsilon_1 - \varepsilon_2) - [2 \cdot N(\sigma) - 4 \cdot (\varepsilon_1 - \varepsilon_2)] \cdot \cos[\delta_2(\sigma)] + [C(\sigma) \cdot (1 - 2 \cdot \varepsilon_3) + 2 \cdot \varepsilon_2] \cdot \cos[\delta_1(\sigma) - \delta_2(\sigma)] + S(\sigma) \cdot (1 - 2 \cdot \varepsilon_3) \cdot \sin[\delta_1(\sigma) - \delta_2(\sigma)] - [C(\sigma) \cdot (1 + 2 \cdot \varepsilon_3) - 2 \cdot \varepsilon_2] \cdot \cos[\delta_1(\sigma) + \delta_2(\sigma)] - S(\sigma) \cdot (1 + 2 \cdot \varepsilon_3) \cdot \sin[\delta_1(\sigma) + \delta_2(\sigma)] + [4 \cdot C(\sigma) \cdot (\varepsilon_3 - \varepsilon_4) - 4 \cdot N(\sigma) \cdot \varepsilon_2] \cdot \cos[\delta_1(\sigma)] + 4 \cdot S(\sigma) \cdot (\varepsilon_3 - \varepsilon_4) \cdot \sin[\delta_1(\sigma)] \} \quad (3)$$

CSE and further expands the theoretical boundaries of UNNs, thus presenting a possible solution paradigm for similar problems. In addition, its high computational efficiency aligns well with the snapshot principle of CSE, thus offering great potential for real-time measurements.

2. Method

2.1. Principle of CSE

To illustrate the measurement process, Fig. 1 exemplifies a schematic diagram of CSE. As can be seen, the light from a broad-spectrum light source passes through the polarization state generator (PSG) consisting of the polarizer P and the multi-order waveplate R1 and is then reflected at the sample surface at an incidence angle of θ . The reflected light

There are many other systematic errors aside from alignment errors, such as retardance errors and incident intensity errors, and the PITUNN can be easily extended to these cases. All that required is to re-derive the mathematical model of the intensity spectrum under specific errors, while everything else remains unchanged.

2.2. Physics-informed tandem untrained neural networks

Inspired by CS and physics-informed machine learning, we transform the demodulation problem of CSE into an optimization problem via sparsity prior and physics-informed regularization:

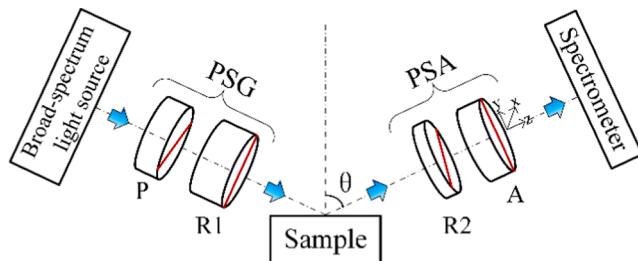


Fig. 1. Schematic diagram of channeled spectroscopic ellipsometry (CSE). The red line represents the transmission axes of the polarizers and the fast axes of the multi-order waveplates.

$$\begin{aligned} & \underset{n.c.s}{\text{minimize}} \|I_{exp} - I_{model}\|_2^2 + \beta \text{Spa}(\hat{n}, \hat{c}, \hat{s}) + \gamma \text{Phy}(\hat{n}, \hat{c}, \hat{s}) \\ & \text{subject to } \hat{n}_i, \hat{c}_i, \hat{s}_i = 0, i \geq \tau(f_i) \\ & \text{where } \text{Spa}(\hat{n}, \hat{c}, \hat{s}) = \|\hat{n} + \hat{c} + \hat{s}\|_1 \quad \text{Phy}(\hat{n}, \hat{c}, \hat{s}) = \|N^2 + C^2 + S^2 - 1\|_2^2 \end{aligned} \quad (4)$$

where I_{exp} represents the spectrum obtained by the spectrometer; I_{model} represents the spectrum calculated based on the mathematical model with the demodulated N , C , and S ; \hat{n} , \hat{c} , and \hat{s} represent the basis coefficients that can be calculated to be N , C , and S via parameterization. In addition, $\text{Spa}(\cdot)$ and $\text{Phy}(\cdot)$ represent sparsity prior and physics-informed regularization, respectively. The details of parameterization,

sparsity prior, and physics-informed regularization are presented in [Appendix A](#). Furthermore, β and γ are hyperparameters to adjust the weights of different priors; \hat{n}_i , \hat{c}_i , and \hat{s}_i denote the i th coefficients; and $\tau(f_i)$ represents the index associated with frequency f_i , which is also a hyperparameter wherein the constraint means the basis coefficients corresponding to the frequencies higher than f_i are set to zero.

Note that the alignment errors can significantly change the mathematical model of the spectrum, thus posing a challenge to the measurements. The sample parameters and alignment errors can be heavily coupled to each other, leading to decreased measurement accuracy. This problem exists whether Eq. (4) is solved by convex optimization methods [27,28] or the UNNs [50,51]. Thus, it is important to develop a method that avoids such coupling while maintaining high accuracy, low computational cost, and low bandwidth limitation.

2.2.1. Workflow of the PITUNN

The paradigm for building the PITUNN is shown in [Fig. 2](#). The measured spectrum I_{exp} (as shown in [Fig. 2\(a\)](#)) is fed into the PITUNN, which consists of three UNNs (with random initial weights, as shown in [Fig. 2\(b\)](#) and (c)) connected in tandem. Within the PITUNN, I_{exp} will sequentially pass through the three UNNs. The first UNN conducts initial estimation of N , C , and S , the second UNN estimates alignment errors based on the initial estimation of N , C , and S , and with the update of network weights, adaptive compensation for alignment errors is performed. The third UNN adjusts N , C , and S again based on the compensation for alignment errors and outputs the final N , C , and S . The

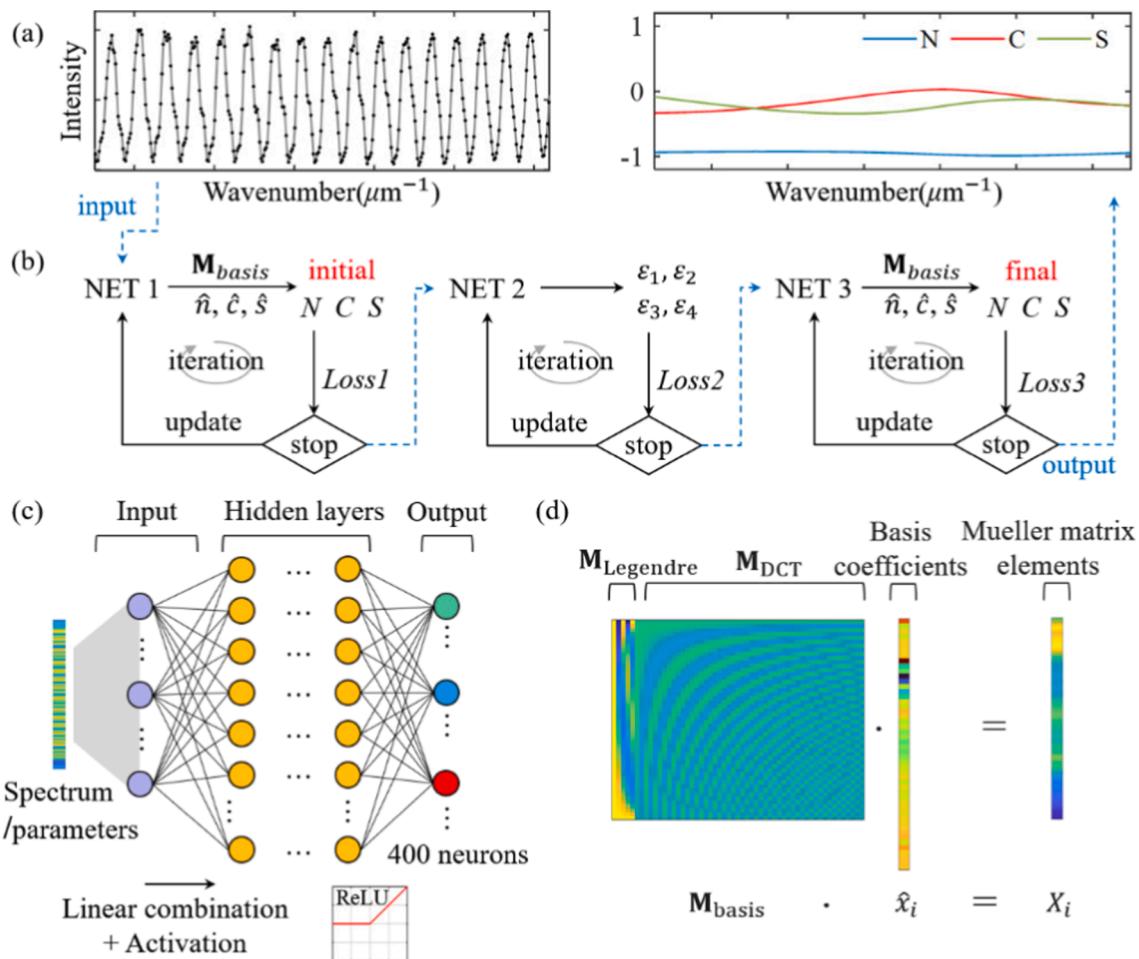


Fig. 2. Overview of the PITUNN. (a) Input of a measured spectrum, which contains information about the sample, systematic errors, and random noise. (b) Overall workflow of the PITUNN, inputting the spectrum and outputting N , C , and S of the sample. (c) Schematic illustration of the networks used in the PITUNN. (d) Parameterization process in accordance with the basis matrix consisting of DCT and Legendre bases.

operational details are as follows:

After receiving the spectrum I_{exp} , NET 1 outputs the basis coefficients \hat{n} , \hat{c} , and \hat{s} , and a set of N , C , and S is calculated by parameterization based on the basis matrix (as shown in Fig. 2(d), details shown in Appendix A). We construct the loss function in NET 1 as follows

$$Loss1 = \|I_{exp} - I_{model}\|_2^2 + \beta_1 \cdot Spa(\hat{n}, \hat{c}, \hat{s}) + \gamma_1 \cdot Phy(\hat{n}, \hat{c}, \hat{s}) \quad (5)$$

where I_{model} is calculated based on the mathematical model in Eq. (2) with N , C , and S from NET 1. Then, we can calculate the value of $Loss1$ of this iteration. If this iteration does not meet the stopping criterion, the weights of NET 1 will be updated based on the backpropagation algorithm, and it will proceed to the next iteration until the stopping criterion is met. Finally, the obtained initial N , C , and S will be outputted to NET 2.

After receiving the initial N , C , and S , NET 2 outputs four values corresponding to the alignment errors. Here, we construct the loss function in NET 2 as follows

$$Loss2 = \|I_{exp} - I_{model}\|_2^2 \quad (6)$$

where I_{model} is calculated based on the mathematical model in Eq. (3) with initial N , C , and S from NET 1 and the alignment errors from NET 2. Next, we calculate the $Loss2$ value of this iteration. When $Loss2$ meets the stopping criterion of NET 2, the obtained alignment errors will be outputted to NET 3, which operates similarly to NET 1. After receiving the input alignment errors, NET 3 outputs the basis coefficients \hat{n} , \hat{c} , and \hat{s} , and a set of N , C , and S is calculated by parameterization. Here, we construct the loss function in NET 3 as follows

$$Loss3 = \|I_{exp} - I_{model}\|_2^2 + \beta_2 \cdot Spa(\hat{n}, \hat{c}, \hat{s}) + \gamma_2 \cdot Phy(\hat{n}, \hat{c}, \hat{s}) \quad (7)$$

where I_{model} is calculated based on the mathematical model in Eq. (3) with the alignment errors from NET 2 and N , C , and S from NET 3. Then, the weights of NET 3 are iteratively updated until the stopping criterion is met. Finally, the obtained N , C , and S will be outputted.

Thus, we can develop the PITUNN. Drawing on the parameterization of CS, we utilize tandem UNNs to automatically compensate for possible alignment errors during network weight updates, ultimately outputting N , C , and S that satisfy physical priors, while other methods neglect system errors. This approach enhances the robustness against system errors and reduces the hardware precision requirements of CSE.

In consideration of the workflow, the PITUNN requires predetermined hyperparameters, which encompass six values, namely, $\tau_1(f_r)$, β_1 , γ_1 , $\tau_2(f_r)$, β_2 , and γ_2 . The discussion on the selection of optimal hyperparameters is presented in Section 5.1.

2.2.2. Structure of the networks and stopping criterion

To optimize computational efficiency, we choose shallow fully connected neural networks (SFCN), which is a classic handcrafted network structure with all neurons connected as illustrated in Fig. 2(c) [53,54], to serve as the core solver of the three UNNs in the PITUNN. The discussion about the network structure is presented in Section 5.2. Each hidden layer consists of 400 neurons, and we use a rectified linear unit (ReLU) as the activation function. Here, NET 1 consists of 4 hidden layers, and the output can be partitioned into the initial N , C , and S . NET 2 consists of 2 hidden layers, and the output can be partitioned into four values corresponding to the alignment errors. NET 3 consists of 2 hidden layers, and the output can be partitioned into the final N , C , and S . The learning rates of all networks are 0.01, 0.01, and 0.001, respectively.

In the PITUNN, we set the stopping criterion as follows

$$\left| \frac{Loss(n) - \frac{1}{5} \sum_{i=1}^5 Loss(n-i)}{\frac{1}{5} \sum_{i=1}^5 Loss(n-i)} \right| \leq \epsilon \quad (8)$$

where $Loss(n)$ represents the value of the loss function at the latest (the n th, $n > 20$) iteration, and ϵ represents the stopping threshold. We stop the iteration and enter the next step when the loss function of the n th iteration does not satisfy Eq. (8). In NET 1, we set the threshold ϵ to 0.5 % to obtain a suitable margin for the alignment errors. As for NET 2 and NET 3, there is no consideration of margin, thus we set the threshold ϵ to 0.02 %. The detailed analysis is presented in Section 5.3.

3. Simulation

3.1. Simulation setting and evaluation criteria

In the simulation, we set the measurement range of the spectrometer to 400–800 nm with a spectral resolution of 1 nm (the number of sample points $K = 400$). Retarders R1 and R2 were quartz waveplates with thicknesses of 4.5 and 1.5 mm, respectively. Three SiO₂ thin films on silicon substrates with different thicknesses (1.9, 112 and 1037 nm) were chosen to generate the simulated spectrum. In addition, a set of alignment errors were randomly generated ($\varepsilon_1 = 0.0804$, $\varepsilon_2 = 0.0886$, $\varepsilon_3 = 0.0502$, and $\varepsilon_4 = 0.0536$), all the alignment errors were in radian unit. With the above settings, we were able to generate the simulated spectrum according to Eq. (3). Here, we also provided a suggested set of hyperparameter values: $\tau_1(f_r) = 0.06 \cdot K$, $\beta_1 = 0.02$, $\gamma_1 = 0.03$, $\tau_2(f_r) = 0.09 \cdot K$, $\beta_2 = 0.02$, $\gamma_2 = 0.3$ for simulation. Note that these values were not considered the optimal hyperparameters for the PITUNN. Further discussion about selecting these hyperparameter values is presented in Section 5.1. In addition, random noise was not considered in the simulation to explore the performance of different methods in the presence of alignment errors. The actual experimental results in Section 4.2 demonstrate the robustness of the PITUNN against random noise.

As for the evaluation criteria of measurement accuracy, we calculate the root mean squared error (RMSE) [36] between the demodulated N , C , and S and the ground truth using the following equation:

$$RMSE = \frac{1}{3} \sum_{i=1}^3 \sqrt{\sum_{j=1}^K [X_i(j) - X_i^{GT}(j)]^2} / K \quad (9)$$

where K is the number of sample points included in the spectrum; X_i ($i = 1, 2, 3$) represents the demodulated N , C , and S ; $X_i(j)$ represents the value of X_i at the j th point; and X_i^{GT} ($i = 1, 2, 3$) represents the ground truth of

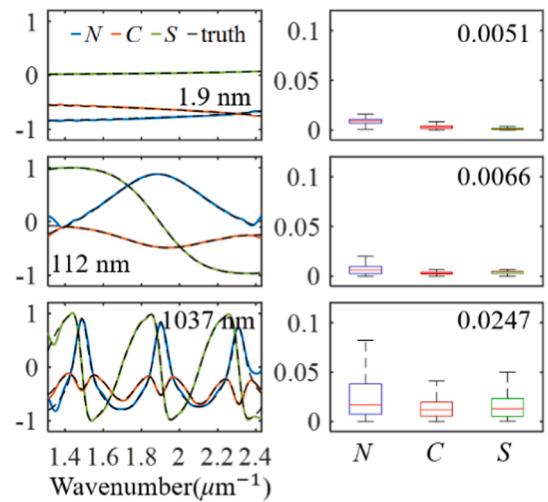


Fig. 3. Simulated measurement results in the presence of alignment errors using the PITUNN. Each row represents the simulation results of SiO₂ thin films of different thicknesses. The first column shows the demodulated N , C , and S , while the second column visualizes the absolute error using box plots. The RMSE values between the reconstructed spectra and ground truth are appended in the corresponding box plots.

N , C , and S . Moreover, we calculated the respective absolute errors of the demodulated N , C , and S and visualized them using box plots. Smaller RMSE and absolute error values indicate higher measurement accuracy.

3.2. Results

Fig. 3 presents the simulated measurement results in the presence of alignment errors using CSE enabled by the PITUNN. The results reveal that as the sample thickness increases, the frequencies of N , C , and S also increase. The measurement accuracy exhibits fluctuations, but overall, it aligns well with the ground truth with respective RMSE values of 0.0051, 0.0066, and 0.0247. The demodulated N , C , and S using the PITUNN also exhibit a small number of anomalies at the two ends due to significant errors in the calibrated system parameters at the two ends of the spectral range. These anomalies indicate relatively large measurement errors, thus highlighting the challenge posed by calibration accuracy to the CSE.

We compared the results of the PITUNN with the state-of-the-art demodulation methods, namely, FR [23], CS [27], and DIP-SP [50], as introduced in Section 1, to further evaluate the performance of the PITUNN. The simulation conditions were identical for all the demodulation methods for fairness of comparison. The corresponding measurement results are presented in Fig. B1 of Appendix B. Furthermore, the RMSE results are listed in Table 1. For clarity of comparison, we use red font to indicate the best performance and blue font to indicate the second-best performance.

As shown in Fig. B1 (a), FR exhibits relatively low overall measurement accuracy, with respective RMSE values of 0.0521, 0.0519, and 0.2382. When the sample thickness is 1037 nm, corresponding to the highest frequency of N , C , and S , FR nearly becomes ineffective due to channel cross-talk, as analyzed in Section 1. At the same time, Fig. B1(b) illustrates the simulated measurement results using CS, showing a moderate improvement in measurement accuracy compared to FR, with respective RMSE values of 0.0747, 0.0154, and 0.1133. Fig. B1(c) illustrates the simulated measurement results using DIP-SP, which achieves higher measurement accuracy compared with FR and CS, with respective RMSE values of 0.0158, 0.0160, and 0.0616. However, these three methods lack robust tolerance to errors in calibrated system parameters, resulting in more anomalies at the two ends compared with the PITUNN as shown in Fig. B1(d).

Thus, through comparison, it is apparent that CSE enabled by the PITUNN achieves the highest measurement accuracy and does not demand strict hardware precision or accurate calibration of system parameters, thus alleviating the drawbacks of existing demodulation methods.

4. Experiments

4.1. Experimental setup

As shown in Fig. 4, a prototype was developed to further verify CSE enabled by the PITUNN. The broad-spectrum light source used was EQ-99-3.1, the converging lenses were both Thorlabs AC254-030-A, the polarizers were both Union optics PGT5012, the multi-order waveplates were customized quartz waveplates of 1.5 and 4.5 mm thicknesses, the

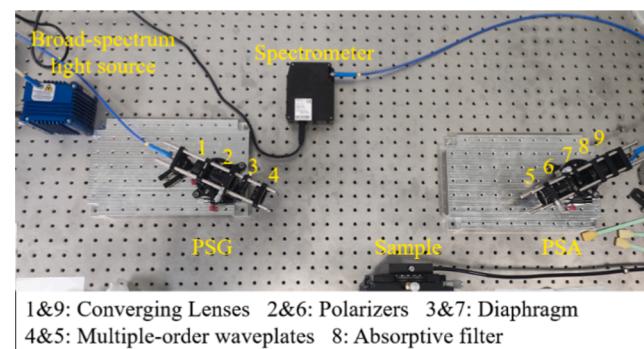


Fig. 4. Experimental setup. The angle of incidence was set to 65°. The sample was mounted vertically by a vacuum chuck.

Table 2

Samples used in the experiments.

Sample	Film structure (surface/substrate)	Thickness (nm)
#1	Au/Glass	19.60
#2	SiO ₂ /Si	1037
#3	Al/Al ₂ O ₃	20.37

absorptive filter was Thorlabs NE10A, and the spectrometer was SE2090-010-VNIR (the resolution and the spectral range were set as in the simulations).

The samples used in the experiments are listed in Table 2, encompassing various substrates and surface materials. Their data (the ground truth) were measured by a commercial ellipsometer (RC2, J. A. Woollam Co., USA). In all the experiments, we employed the same hyperparameters as those used in the simulations: $\tau_1(f_\tau) = 0.06 \cdot K$, $\beta_1 = 0.02$,

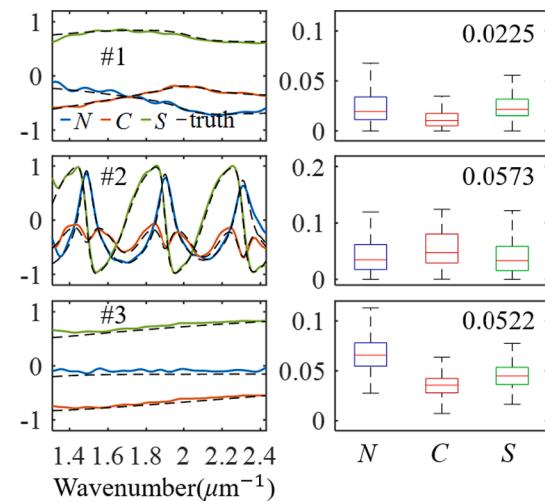


Fig. 5. Experimental measurement results using the PITUNN on different samples.

Table 1

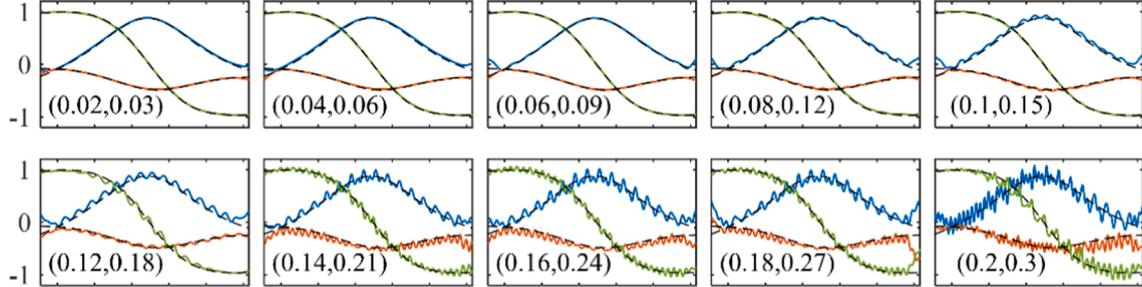
RMSE of different methods in the simulation.

Sample	FR	CS	DIP-SP	PITUNN
1.9 nm	0.0521	0.0747	0.0158	0.0051
112 nm	0.0519	0.0154	0.0160	0.0066
1037 nm	0.2382	0.1133	0.0616	0.0247

Table 3

RMSE of different methods in the experiments.

Sample	FR	CS	DIP-SP	PITUNN
#1	0.0574	0.0257	0.0245	0.0225
#2	0.1772	0.1008	0.0827	0.0573
#3	0.1118	0.0546	0.0565	0.0522

**Fig. 6.** Simulated measurement results using the PITUNN when $\tau_1(f_r)$ and $\tau_2(f_r)$ are taken as different groups of values. The specific values of $\tau_1(f_r)/K$ and $\tau_2(f_r)/K$ are labeled in the subplots, and the other hyperparameter values are as those used in the simulations ($\beta_1 = 0.02$, $\gamma_1 = 0.03$, $\beta_2 = 0.02$, $\gamma_2 = 0.3$).

$$\gamma_1 = 0.03, \tau_2(f_r) = 0.09 \cdot K, \beta_2 = 0.02, \gamma_2 = 0.3.$$

4.2. Results

The experimental measurement results using the PITUNN are presented in Fig. 5. Overall, the demodulated values of N , C , and S align well with the ground truth, with respective RMSE values of 0.0225, 0.0573, and 0.0522.

The experimental measurement results using other competitive methods are presented in Fig. B2 of Appendix B, and the RMSE results are listed in Table 3. We use red font to indicate the best performance and blue font to indicate the second-best performance. Upon comparing the experimental measurement results among these competitive methods, it is observed that the PITUNN maintains the highest measurement accuracy across varying samples in the same measurement condition. This robust performance demonstrates the ability of the PITUNN to handle random noise and alignment errors effectively.

Next, we conducted multiple repetitions of the experiments to compare the computational time of these methods. The average processing time for CS was about 0.5 s; for DIP-SP, it was about 7 s; and for the PITUNN, it was about 1 s. Evidently, due to its simple network structure and appropriate stopping criterion, the PITUNN exhibited high computational efficiency, making it a promising method for real-time measurements.

5. Discussion

In this section, we present a detailed theoretical discussion of the PITUNN based on the simulated data of 112 nm SiO₂ thin films in the presence of alignment errors. First, we discuss some empirical rules for hyperparameter selection and determine the optimal hyperparameter values for the simulated measurements of the 112 nm SiO₂ sample. Then, we compare the measurement results using different network architectures as the core solver of the three UNNs in the PITUNN and discuss the early stopping of NET 1. Finally, we explore the contribution of physics-informed regularization in the PITUNN.

5.1. Selection of hyperparameter values

Although the simulated and experimental results presented above have demonstrated the effectiveness of using the same set of recommended hyperparameters, it is still possible to fine-tune the hyperparameters for each specific measurement considering the physical significance of the hyperparameters and the prior information of the samples. However, determining suitable hyperparameter values remains a challenging task. Therefore, in this section, we offer some empirical adjustment strategies. Notably, complex coupling and trade-offs exist among the hyperparameters, and empirical strategies may only partially guide us to the optimal hyperparameter values. Subsequently, we employ a genetic algorithm to identify a more optimal set of hyperparameter values based on simulated data of the 112 nm SiO₂ sample, thus forming the basis for further discussions.

The PITUNN involves several hyperparameters: $\tau_1(f_r)$, β_1 , γ_1 , $\tau_2(f_r)$, β_2 , and γ_2 . Among these, $\tau_1(f_r)$ and $\tau_2(f_r)$ regulate the highest frequency in the basis matrix. Consequently, if prior information suggests a lower frequency for sample parameters, smaller values of $\tau_1(f_r)$ and $\tau_2(f_r)$ can be chosen; conversely, larger values are suitable for higher frequencies. As shown in Fig. 6, the frequency components encompassed in the demodulated N , C , and S increase progressively by gradually increasing $\tau_1(f_r)$ and $\tau_2(f_r)$.

The hyperparameters β_1 and β_2 act as weighting coefficients that control the sparsity prior and are mainly linked to random noise during measurements. When facing significant random noise, such as dust on the sample surface or low reflectivity of the sample, larger β_1 and β_2 values are appropriate, as discussed in Ref. [27]. Additionally, larger β_1 and β_2 tend to output N , C , and S with lower frequency, thus necessitating a trade-off. As shown in Fig. 7, the frequency components encompassed in the demodulated N , C , and S decrease progressively by gradually increasing β_1 and β_2 .

The hyperparameters γ_1 and γ_2 are the weighting coefficients that control the physics-informed regularization and introduce physical priors during the network parameter update process. Generally, these coefficients require little adjustment. As shown in Fig. 8(a), we

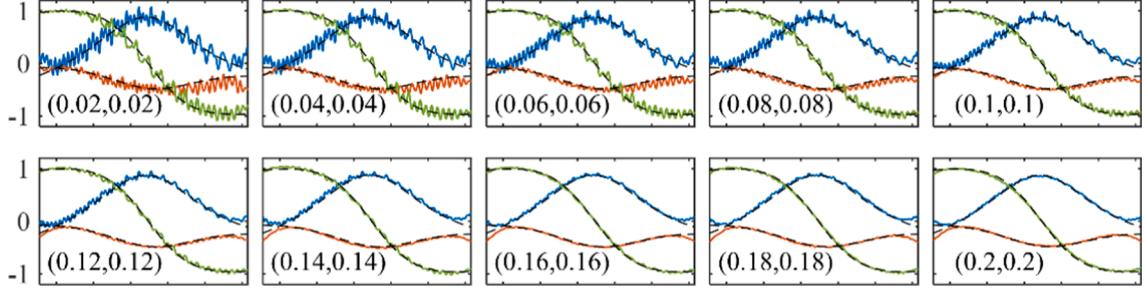


Fig. 7. Simulated measurement results using the PITUNN when β_1 and β_2 are taken as different groups of values. The specific values of β_1 and β_2 are labeled in the subplots, and the other hyperparameter values are $\tau_1(f_r) = 0.2 \cdot K$, $\gamma_1 = 0.03$, $\tau_2(f_r) = 0.3 \cdot K$, $\gamma_2 = 0.3$.

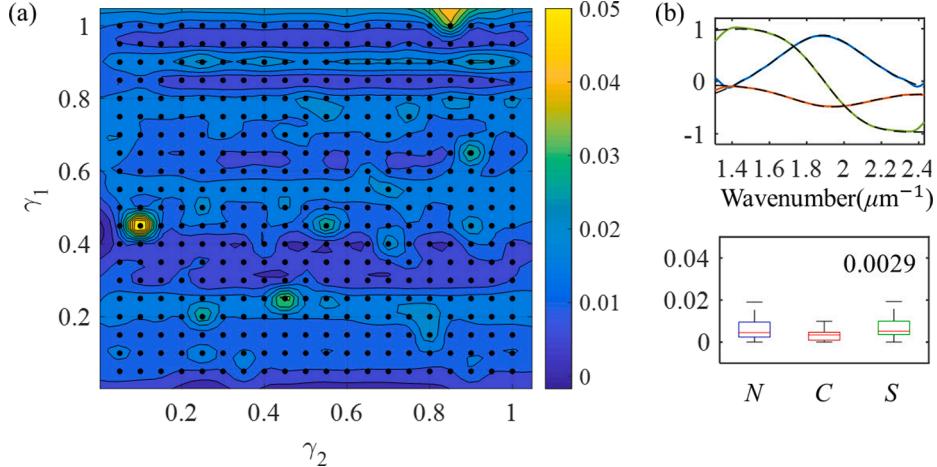


Fig. 8. (a) RMSE values using the PITUNN when γ_1 and γ_2 are taken as different groups of values and presented in the form of a contour plot; (b) Simulated measurement results using the PITUNN with optimized hyperparameters.

considered 19 discrete values for γ_1 and γ_2 in the range of 0.05–1, with an interval of 0.05, leading to a total of 361 simulations. As can be observed, the contour lines of RMSE are primarily horizontally distributed, indicating that γ_1 has a more significant impact on the measurement accuracy of the PITUNN. Furthermore, most regions exhibit relatively low RMSE values, suggesting that the measurement accuracy is not highly sensitive to γ_1 and γ_2 . Therefore, extensive adjustments are not necessary.

Subsequently, for the simulated data of the 112 nm SiO₂ sample, we employed a genetic algorithm to optimize the hyperparameters, resulting in the following optimized set of values: $\tau_1(f_r) = 0.0807 \cdot K$, $\beta_1 = 0.3357$, $\gamma_1 = 0.1023$, $\tau_2(f_r) = 0.0804 \cdot K$, $\beta_2 = 0.0112$, and $\gamma_2 = 0.8770$. With these optimized hyperparameters, the simulated measurement results of the PITUNN are displayed in Fig. 8(b), yielding an RMSE value of 0.0029.

As can be seen, the adjusted hyperparameters enhance the performance of the PITUNN. Furthermore, the theoretical upper limit of the PITUNN's accuracy exceeded what was obtained by using the fixed

hyperparameters. Thus, it can be concluded that hyperparameter tuning can integrate more prior information about measurement conditions and the sample into the measurement process. Subsequent discussions are based on the optimized hyperparameter values: ($\tau_1(f_r) = 0.0807 \cdot K$, $\beta_1 = 0.3357$, $\gamma_1 = 0.1023$, $\tau_2(f_r) = 0.0804 \cdot K$, $\beta_2 = 0.0112$, and $\gamma_2 = 0.8770$).

5.2. Structure of the networks

As can be observed from Eq. (3), the negative impact of the alignment errors on the measurement can be roughly summarized as a “global offset” [52] (although it should be noted that this oversimplification overlooks the complex coupling phenomena). This implies that the alignment errors have a similar mathematical significance as N , C , and S in the model. Thus, unlike previous works on random noise in images [55,56], it is impossible to isolate the alignment errors from N , C , and S using an underparameterized network like the deep decoder or a theory of structural bias. Therefore, we can say that the

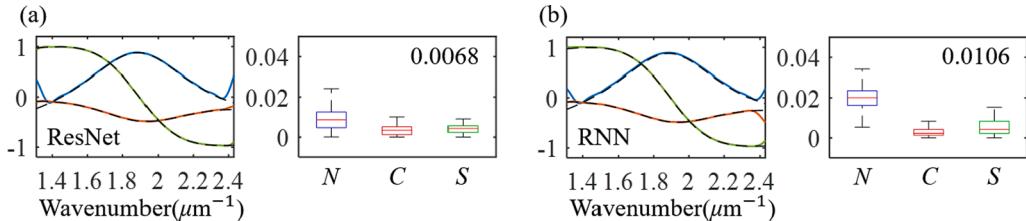


Fig. 9. Simulated measurement results taking (a) ResNet and (b) RNN as the core solver of the three UNNs of the PITUNN.

Table 4

Computational times of the PITUNN based on different neural networks.

	NET 1	NET 2	NET 3	Total (s)
SFCN	0.70	0.06	0.35	1.11
ResNet	2.30	1.27	6.75	10.32
RNN	1.65	0.50	2.26	4.41

network structure is unlikely to affect the measurement accuracy significantly. Moreover, employing relatively simple neural networks is feasible, considering that the mathematical model of CSE is relatively simple.

To validate this hypothesis, we select three representative neural networks as the core solvers of the three UNNs in the PITUNN, including SFCN (the representative underparameterized networks, as shown in Fig. 2(c)), residual network (ResNet, the representative over-parameterized networks) [57], and recurrent neural networks (RNN, the representative serial data processing networks) [58]. We compare the measurement accuracy achieved by these different networks.

The measurement results using ResNet and RNN are shown in Fig. 9, and the results using SFCN are shown in Fig. 3. The results reveal that the various neural network structures employed do not substantially influence the final measurement accuracy. This observation is in accordance with the analysis presented above.

Then, we recorded the computational time of the PITUNN using

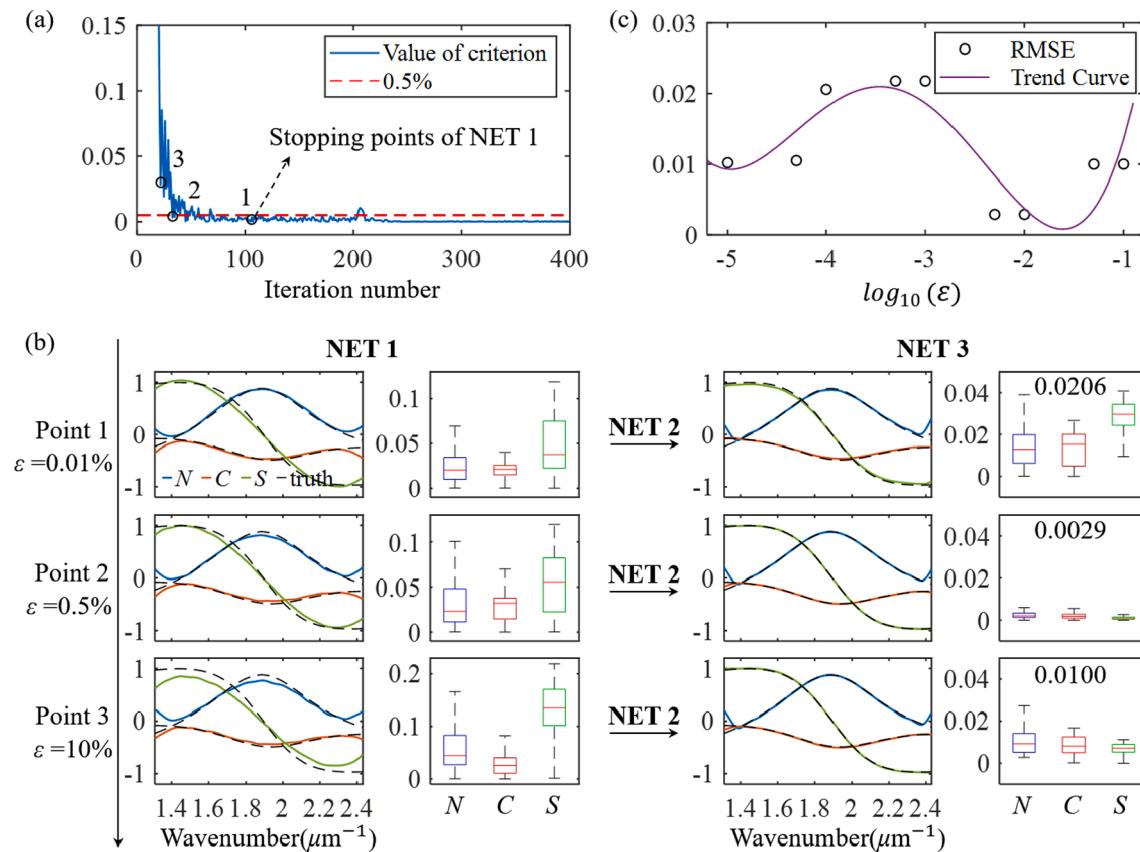


Fig. 10. Discussion on the early stopping of NET 1. (a) Variation of the constructed stopping criterion in the simulation. (b) Initial N , C , and S output by NET 1 and the final N , C , and S output by NET 3 when taking different values of ϵ . (c) RMSE values of the PITUNN when taking different values of ϵ and the trend curve.

different neural networks by performing several simulations under the same condition. The average computational time is shown in Table 4. From this, we can see that, due to the simple structure of SFCN, the computational cost required is the smallest when it is used as the core solver of the three UNNs in the PITUNN. Therefore, we choose SFCN as the core solver of the three UNNs in the PITUNN.

5.3. Early stopping in NET 1

As discussed in Section 5.2, it is impossible to isolate N , C , S and the alignment errors by designing the network structure. Thus, we propose to develop an early stopping mechanism for NET 1 to handle the alignment errors. On the one hand, the output N , C , and S cannot serve as a reliable reference for the following networks if the iterative updating of NET 1 is stopped too early. On the other hand, if the iterative updating of NET 1 is stopped too late, it may result in an insufficient margin for the following networks to compensate for the alignment errors. Therefore, we must set a suitable stopping criterion to obtain a high measurement accuracy.

For this reason, we used Eq. (8) and conducted simulations based on a range of ϵ values to establish a suitable stopping criterion. Throughout the simulations, we monitored the value of the stopping criterion as the iterative updating of NET 1 progresses, as shown in Fig. 10(a). Specifically, we selected three different values of ϵ , 10 %, 0.5 %, and 0.01 % and visualized the calculation procedure of the PITUNN to provide a

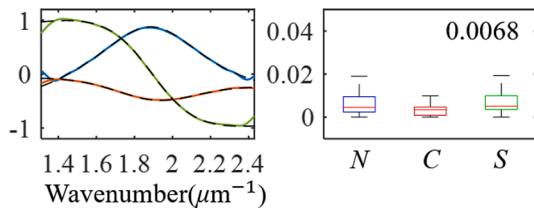


Fig. 11. Measurement results of the PITUNN when no physics-informed regularization is incorporated.

more precise illustration of the iterative process. The output of NET 1 (initial N , C , and S) and the output of NET 3 (final N , C , and S) are shown in Fig. 10(b). As can be seen, when $\epsilon = 0.5\%$, the final N , C , and S have the highest measurement accuracy. When $\epsilon = 10\%$, the reference N , C , and S output from NET 1 has relatively large errors (as in the case of “stopping too early” in the above analysis), and the final N , C , and S outputs from NET 3 have relatively low accuracies. Similarly, when $\epsilon = 0.01\%$, the reference N , C , and S outputs from NET 1 have relatively small errors (as in the case of “stopping too late” in the above analysis), and the final N , C , and S output from NET 3 also have relatively low accuracies. The results of the simulation above prove that NET 1 does need an appropriate stopping criterion.

Thus, we conduct further simulations with more values of ϵ . The relationship between the RMSE values and ϵ is illustrated in Fig. 10(c). The results reveal that when the stopping criterion is set to approximately $\epsilon = 0.5\%$, a high measurement accuracy can be achieved. Moreover, the results in Sections 3.2 and 4.2 consistently demonstrate that this particular stopping criterion leads to stable and high measurement accuracy across various scenarios.

5.4. Contribution of physics-informed regularization

For the demodulation problem in CSE, a calibration must be performed before the measurement to determine the specific values of the system parameters. The intensity spectrum I_{exp} was obtained by the spectrometer so that the demodulation problem is underdetermined. As more information is incorporated into the underdetermined problem, the possibility of obtaining a more accurate solution theoretically increases. Thus, we integrated the physics-informed regularization term into the optimization problem, as shown in Eq. (4). Notably, a physics-informed regularization does not exist in all cases, such as when measuring samples with depolarization where $N^2 + C^2 + S^2 \neq 1$. Thus, in this section, we discuss the contribution of the physics-informed regularization to the PITUNN.

As discussed in Section 5.1, the sensitivity of measurement accuracy to γ_1 and γ_2 is not pronounced, thereby highlighting the flexibility and convenience of physics-informed regularization. Subsequently, we set both γ_1 and γ_2 to 0, thus excluding physics-informed regularization. The simulated measurement results shown in Fig. 11 indicate a slight decline in measurement accuracy compared with the results in Section 3.2. These results not only verify the contribution of physics-informed regularization to the performance of the PITUNN but also demonstrate that even in the absence of physics-informed regularization, the PITUNN can achieve relatively high measurement accuracy guided by other physical information, which can still be utilized for measurements of samples with different characteristics. Therefore, we can state that the PITUNN exhibits strong generalization capabilities.

6. Conclusion

In this work, we have introduced CSE enabled by the physics-informed tandem untrained networks. The proposed PITUNN integrates sparsity prior and physics-informed regularization, as well as utilizes a tandem network consisting of three UNNs to address system errors (e.g., alignment errors) and random noise encountered during measurements.

The analysis based on simulated and experimental results demonstrates the following contributions of this work:

Pioneering paradigm for CSE measurement. Methods based on data-driven machine learning face challenges in generalization and training costs, while their outputs are not always physically plausible. In the PITUNN, we integrate the capabilities of physics-informed machine learning with tandem UNNs for the first time, developing a novel paradigm for the demodulation problem in CSE. Such a paradigm ensures that the outputs are physically plausible and eliminates the challenges related to generalization and training costs, holding potential implications for other domains, such as channeled spectroscopic polarimetry.

High accuracy and robustness. Traditional CSE are constrained by their stringent requirements for hardware precision and calibration, leading to relatively low measurement accuracy. Through comprehensive simulations and experiments, we have validated the superior measurement accuracy of the PITUNN, demonstrating its exceptional robustness against system errors and random noise.

Rapid performance with high efficiency. Due to its simplistic network structure and judicious early stopping mechanism during network updates, the PITUNN boasts an average single-run time of approximately 1–1.5 s in both simulations and experiments. While maintaining high measurement accuracy and robustness, it also retains rapid processing speeds. The efficiency of the PITUNN aligns perfectly with the snapshot nature of CSE, marking a significant advancement for the online applications of CSE.

Furthermore, the PITUNN offers new insights for addressing systematic errors in optical measurement processes, while also expanding the application of UNNs in measurements.

Funding

This work was partially funded by the National Natural Science Foundation of China (Grant No. 62175075, 52022034, 52130504).

CRediT authorship contribution statement

Shilong Yang: Conceptualization, Methodology, Software, Writing – original draft, Investigation, Writing – review & editing. **Xiuguo Chen:** Funding acquisition, Supervision, Writing – review & editing. **Wenlong Chen:** Formal analysis, Investigation. **Jing Hu:** Investigation. **Yifu Wang:** Investigation. **Shuo Liu:** Investigation, Software. **Shiyuan Liu:** Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

can be seen at <https://github.com/xiuguochen/PITUNN.git>.

We have uploaded all of the source data and code to Github, which

Appendix

Appendix A: Sparsity prior and physics-informed regularization

Prior 1: Sparsity prior

Following the theory of compressed sensing, we can conclude the first prior: Sparsity prior. We can use a family of orthogonal functions to represent any function in their function space parametrically. It is empirically known that most of the isotropic sample Mueller matrix elements are fluctuating quasi-periodic functions. Thus, we can use Fourier basis functions, Legendre polynomials, wavelet basis functions, and so on to parameterize the Mueller matrix elements.

For example, some researchers used the discrete cosine transform (DCT) basis and Legendre polynomials to parameterize the Stokes parameters and Mueller matrix elements in their works [27–30]. The parameterization process can be expressed as follows

$$\mathbf{X}_i = \mathbf{M}_{\text{basis}} \cdot \hat{\mathbf{x}}_i \quad (\text{A1})$$

where \mathbf{X}_i ($i = 1, 2, 3$) represents N, C , and S in vector form; $\mathbf{M}_{\text{basis}}$ represents the basis matrix; and $\hat{\mathbf{x}}_i$ represents the corresponding basis coefficients in vector form. Here, the demodulation problem for CSE is converted to solving for a set of basis coefficients.

The basis matrix $\mathbf{M}_{\text{basis}}$ is constructed by the DCT basis \mathbf{M}_{DCT} and Legendre basis $\mathbf{M}_{\text{Legendre}}$

$$\mathbf{M}_{\text{DCT}}^{K \times K}(i,j) = \begin{cases} \sqrt{\frac{2}{K}} \cos\left(\frac{\pi}{2K}(2j-1)(i-1)\right), & \text{for } i = 2, \dots, K \\ \sqrt{\frac{1}{K}}, & \text{for } i = 1 \end{cases}, j = 1, 2, \dots, K \quad (\text{A2})$$

$$\mathbf{M}_{\text{Legendre}}^{K \times L} = \begin{bmatrix} P_1(t_1) & \dots & P_L(t_1) \\ \vdots & \ddots & \vdots \\ P_1(t_K) & \dots & P_L(t_K) \end{bmatrix} \quad (\text{A3})$$

where K is the number of sample points included in the spectrum, and t_1, \dots, t_K uniformly sample the interval $[-1, 1]$. In addition, the Legendre polynomial $P_i(t)$, ($i = 1, 2, \dots, L$) is given by

$$P_i(t) = 2^i \sum_{j=0}^i t^j \binom{i}{j} \binom{i+j-1}{2} \quad (\text{A4})$$

$$\mathbf{M}_{\text{basis}}^{K \times (K+L)} = \begin{bmatrix} \mathbf{M}_{\text{Legendre}}^{K \times L} & \mathbf{M}_{\text{DCT}}^{K \times K} \end{bmatrix} \quad (\text{A5})$$

Here, we set $L = 5$.

After a suitable parameterization, we tend to use fewer non-zero basis coefficients to characterize the Mueller matrix elements [27]. This can be summarized in terms of sparsity, and the sparsity of the basis coefficients is usually expressed as

$$\text{Spa}(\hat{n}, \hat{c}, \hat{s}) = \|\hat{n} + \hat{c} + \hat{s}\|_1 \quad (\text{A6})$$

Prior 2: Physics-informed regularization

In the specific situation of samples without depolarization, there exists a physical prior expressed as

$$N^2 + C^2 + S^2 = 1 \quad (\text{A7})$$

Thus, we can construct the constraint in the loss function:

$$\text{Phy}(\hat{n}, \hat{c}, \hat{s}) = \|N^2 + C^2 + S^2 - 1\|_2^2 \quad (\text{A8})$$

Appendix B: Experimental measurement results using other methods

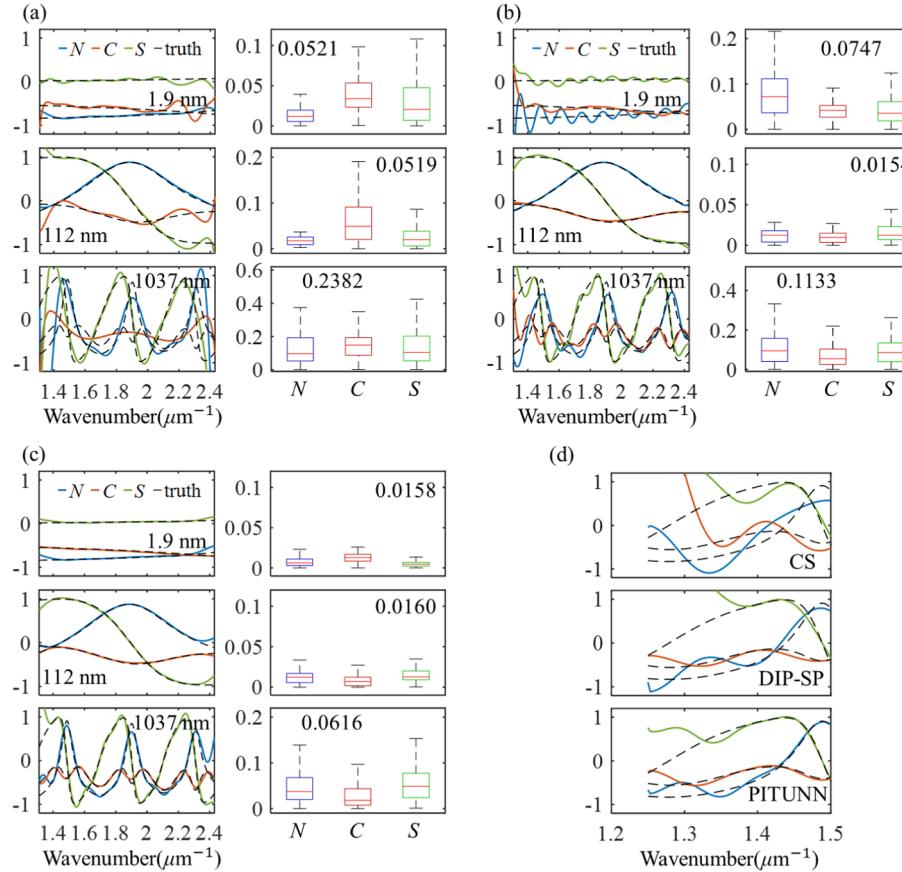


Figure B1. Simulated measurement results using (a) FR, (b) CS, (c) DIP-SP on different samples. (d) Localized enlargements of simulated measurement results on 1037 nm sample at the left end

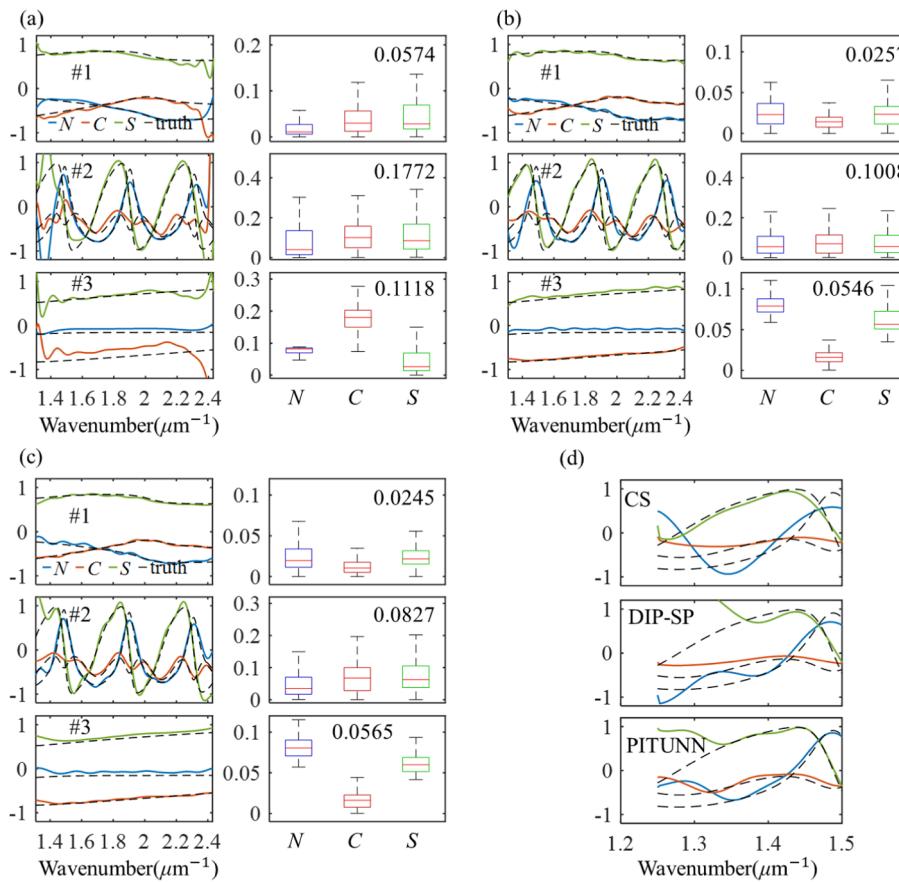


Figure B2. Experimental measurement results using (a) FR, (b) CS, (c) DIP-SP on different samples. (d) Localized enlargements of experimental measurement results on sample #2 at the left end

References

- [1] E. Garcia-Caurel, A. De Martino, J.P. Gaston, L. Yan, Application of spectroscopic ellipsometry and Mueller ellipsometry to optical characterization, *Appl. Spectrosc.* 67 (2013) 1–21.
- [2] D.E. Aspnes, Spectroscopic ellipsometry—past, present, and future, *Thin Solid Films* 571 (2014) 334–344.
- [3] X. Chen, J. Hu, W. Chen, S. Yang, Y. Wang, Multi-spectral snapshot diffraction-based overlay metrology, *Opt. Lett.* 48 (13) (2023) 3383–3386.
- [4] H. Fujiwara, *Spectroscopic Ellipsometry: Principles and Applications*, John Wiley & Sons, 2007.
- [5] X. Chen, E.P.J. Parrott, Z. Huang, H.P. Chan, E. Pickwell-MacPherson, Robust and accurate terahertz time-domain spectroscopic ellipsometry, *Photon. Res.* 6 (2018) 768–775.
- [6] Y. Yuan, C. Yao, W. Shen, X. Hu, C. Hu, Polarization Measurement Method Based on Liquid Crystal Variable Retarder (LCVR) for Atomic Thin-Film Thickness, *Nanomanuf. Metrol.* 5 (2) (2022) 159–166.
- [7] J.N. Hilfiker, N. Hong, S. Schoeche, Mueller matrix spectroscopic ellipsometry, *Adv. Opt. Technol.* 11 (3–4) (2022) 59–91.
- [8] X. Chen, H. Gu, J. Liu, C. Chen, S. Liu, Advanced Mueller matrix ellipsometry: Instrumentation and emerging applications, *Sci. China Technol. Sci.* 65 (2022) 2007–2030.
- [9] H. Okabe, M. Hayakawa, J. Matoba, H. Naito, K. Oka, Error-reduced channeled spectroscopic ellipsometer with palm-size sensing head, *Rev. Sci. Instrum.* 80 (2009) 083104.
- [10] S.W. Lee, S.Y. Lee, G. Choi, H.J. Park, Co-axial spectroscopic snap-shot ellipsometry for real-time thickness measurements with a small spot size, *Opt. Express* 28 (2020) 25879–25893.
- [11] N.G. Orji, M. Badaroglu, B.M. Barnes, C. Beitia, B.D. Bunday, U. Celano, R.J. Kline, M. Neisser, Y. Obeng, A.E. Vladar, Metrology for the next generation of semiconductor devices, *Nat. Electron.* 1 (2018) 532–547.
- [12] G. Choi, S.W. Lee, S.Y. Lee, H.J. Pahk, Single-shot multispectral angle-resolved ellipsometry, *Appl. Opt.* 59 (21) (2020) 6296–6303.
- [13] G.G. Politano, C. Vena, G. Desiderio, C. Versace, Variable angle spectroscopic ellipsometry characterization of turbostratic CVD-grown bilayer and trilayer graphene, *Opt. Mater.* 107 (2020) 110165.
- [14] A. Tuniyazi, T. Mu, X. Jiang, F. Han, H. Li, Q. Li, H. Gong, W. Wang, B. Qin, Snapshot polarized light scattering spectroscopy using spectrally-modulated polarimetry for early gastric cancer detection, *J. Biophoton.* 14 (2021) e202100140.
- [15] C. He, H. He, J. Chang, B. Chen, H. Ma, M.J. Booth, Polarisation optics for biomedical and clinical applications: a review, *Light Sci. Appl.* 10 (2021) 194.
- [16] J.S. Tyo, D.L. Goldstein, D.B. Chenault, J.A. Shaw, Review of passive imaging polarimetry for remote sensing applications, *Appl. Opt.* 45 (2006) 5453–5469.
- [17] Q. Li, L. Dong, Y. Hu, Q. Hao, W. Wang, J. Cao, Y. Cheng, Polarimetry for bionic geolocation and navigation applications: a review, *Remote Sens.* 15 (2023) 3518.
- [18] K. Oka, T. Kato, Spectroscopic polarimetry with a channeled spectrum, *Opt. Lett.* 24 (1999) 1475–1477.
- [19] H. Okabe, M. Hayakawa, H. Naito, A. Taniguchi, K. Oka, Spectroscopic polarimetry using channeled spectroscopic polarization state generator (CSPSG), *Opt. Express* 15 (2007) 3093–3109.
- [20] A.S. Alenin, J.S. Tyo, Generalized channeled polarimetry, *J. Opt. Soc. Am. A* 31 (2014) 1013–1022.
- [21] N. Hagen, Calibration and demonstration of a snapshot Mueller matrix spectropolarimeter, *Proc. SPIE* 11833 (2021) 125–134.
- [22] N. Hagen, Design of channeled spectropolarimeters, *Appl. Opt.* 61 (12) (2022) 3381–3389.
- [23] M.W. Kudnenov, N.A. Hagen, E.L. Dereniak, G.R. Gerhart, Fourier transform channeled spectropolarimetry in the MWIR, *Opt. Express* 15 (20) (2007) 12792–12805.
- [24] D.J. Lee, C.F. LaCasse, J.M. Craven, Channeled spectropolarimetry using iterative reconstruction, *Proc. SPIE* 9853 (2016) 240–253.
- [25] F. Han, T. Mu, D. Bao, A. Tuniyazi, Q. Li, H. Gong, Z. Chen, C. Zhang, Iterative reconstruction for general linear imaging polarimetry without polarimetric calibration, *Opt. Lett.* 45 (1) (2020) 57–60.
- [26] F. Han, T. Mu, A. Tuniyazi, D. Bao, H. Gong, Q. Li, Z. Chen, C. Zhang, Iterative reconstruction for snapshot intensity-modulated linear imaging spectropolarimetry without Fourier transform and phase calibration, *Opt. Lasers Eng.* 134 (2020) 106286.
- [27] D.J. Lee, C.F. LaCasse, J.M. Craven, Compressed channeled spectropolarimetry, *Opt. Express* 25 (25) (2017) 32041–32063.
- [28] W. Ren, C. Fu, D. Wu, Y. Xie, G.R. Arce, Channeled compressive imaging spectropolarimeter, *Opt. Express* 27 (3) (2019) 2197–2211.
- [29] G. Zhou, Y. Li, K. Liu, Adaptive linear reconstruction with regularizer for channeled spectropolarimeter, *Opt. Eng.* 60 (6) (2021) 064111.

- [30] G. Zhou, Y. Li, K. Liu, Reconstruction and calibration methods for a Mueller channeled spectropolarimeter, *Opt. Express* 30 (2) (2022) 2018–2032.
- [31] H. Bendada, B. Bakhouche, L.O. González-Siu, N.C. Bruce, O. Arteaga, Spectrally modulated polarimetry with wavelength domain analysis, *Appl. Opt.* 61 (19) (2022) 5608–5613.
- [32] P.R. Wiecha, A. Arbouet, C. Girard, O.L. Muskens, Deep learning in nanophotonics: inverse design and beyond, *Photon. Res.* 9 (5) (2021) B182–B200.
- [33] X. Xu, C. Sun, Y. Li, J. Zhao, J. Han, W. Huang, An improved tandem neural network for the inverse design of nanophotonics devices, *Opt. Commun.* 481 (2021) 126513.
- [34] C. Yeung, J.M. Tsai, B. King, B. Pham, J. Liang, D. Ho, M.W. Knight, A.P. Raman, Designing multiplexed supercell metasurfaces with tandem neural networks, *Nanophotonics* 10 (2021) 1133–1143.
- [35] S. Liu, X. Chen, T. Yang, C. Guo, J. Zhang, J. Ma, C. Chen, C. Wang, C. Zhang, S. Liu, Machine learning aided solution to the inverse problem in optical scatterometry, *Measurement* 191 (2022) 110811.
- [36] J. Liu, D. Zhang, D. Yu, M. Ren, J. Xu, Machine learning powered ellipsometry, *Light Sci. Appl.* 10 (1) (2021) 55.
- [37] S. Zhu, E. Guo, J. Gu, L. Bai, J. Han, Imaging through unknown scattering media based on physics-informed learning, *Photon. Res.* 9 (5) (2021) B210–B219.
- [38] H. Kwak, J. Kim, Semiconductor Multilayer Nanometrology with Machine Learning, *Nanomanuf. Metrol.* 6 (1) (2023) 15.
- [39] X. Lv, Z. Yang, Y. Wang, K. Zhou, J. Lin, P. Jin, Channeled imaging spectropolarimeter reconstruction by neural networks, *Opt. Express* 29 (22) (2021) 35556–35569.
- [40] Z. Wang, Y.C. Lin, K. Zhang, W. Wu, and S. Huang, “EllipsoNet: Deep-learning-enabled optical ellipsometry for complex thin films,” arXiv:2210.05630 (2022).
- [41] Q. Li, J. Song, A.S. Alenin, J.S. Tyo, Spectral-temporal channeled spectropolarimetry using deep-learning-based adaptive filtering, *Opt. Lett.* 46 (17) (2021) 4394–4397.
- [42] C. Huang, S. Wu, Y. Chang, Y. Fang, Z. Zou, H. Qiu, Convolutional neural network-based spectrum reconstruction solver for channeled spectropolarimeter, *Opt. Express* 30 (7) (2022) 10367–10386.
- [43] J. Song, R. Abay, J.S. Tyo, A.S. Alenin, Transcending conventional snapshot polarimeter performance via neuromorphically adaptive filters, *Opt. Express* 29 (12) (2021) 17758–17774.
- [44] V. Lempitsky, A. Vedaldi, D. Ulyanov, Deep image prior, *IEEE Conference on Computer Vision and Pattern Recognition* 2018 (2018) 9446–9454.
- [45] A. Qayyum, I. Ilahi, F. Shamshad, M. Bennamoun, and J. Qadir, “Untrained neural network priors for inverse imaging problems: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.* (2022).
- [46] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, *Nat. Rev. Phys.* 3 (6) (2021) 422–440.
- [47] S.A. Faroughi, N. Pawar, C. Fernandes, M. Raissi, S. Das, N.K. Kalantari, and S.K. Mahjour, “Physics-guided, physics-informed, and physics-encoded neural networks in scientific computing,” arXiv:2211.07377, 2022.
- [48] P. Shi, Z. Zeng, T. Liang, Physics-informed convnet: Learning physical field from a shallow neural network, *Commun. Nonlinear Sci. Numer. Simul.* 107911 (2024).
- [49] Z. Tian, L. Pei, J. Wang, K. Hu, J. Zheng, J. Li, T. Ning, Physics-driven mode decomposition based on a training-free shallow neural network, *Opt. Fiber Technol.* 76 (2023) 103239.
- [50] F. Han, T. Mu, H. Li, A. Tuniyazi, Deep image prior plus sparsity prior: toward single-shot full-Stokes spectropolarimetric imaging with a multiple-order retarder, *AP Nexus* 2 (3) (2023) 036009.
- [51] C. Huang, H. Liu, S. Wu, X. Jiang, L. Zhou, J. Hu, Physics-guided neural network for channeled spectropolarimeter spectral reconstruction, *Opt. Express* 31 (15) (2023) 24387–24403.
- [52] M. Dubreuil, S. Rivet, B.L. Jeune, J. Cariou, Systematic errors specific to a snapshot Mueller matrix polarimeter, *Appl. Opt.* 48 (6) (2009) 1135–1142.
- [53] J.P. Heo, C.G. Im, K.H. Ryu, S.W. Sung, C. Yoo, D.R. Yang, Shallow fully connected neural network training by forcing linearization into valid region and balancing training rates, *Processes* 10 (6) (2022) 1157.
- [54] J. Wen, L. Hao, C. Gao, H. Wang, K. Mo, W. Yuan, X. Chen, Y. Wang, Y. Zhang, Y. Shao, C. Yang, W. Shen, Deep learning-based miniaturized all-dielectric ultracompact film spectrometer, *Acs Photon.* 10 (1) (2023) 225–233.
- [55] R. Heckel and P. Hand, “Deep decoder: Concise image representations from untrained non-convolutional networks,” arXiv:1810.03982 (2018).
- [56] R. Heckel and M. Soltanolkotabi, “Denoising and regularization via exploiting the structural bias of convolutional generators,” arXiv:1910.14634 (2019).
- [57] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Proc. IEEE Comput. Soc. Conf. Comput. vis. Pattern Recognit.* (2016) 770–778.
- [58] L.R. Medsker, L.C. Jain, Recurrent neural networks: design and applications, CRC Press, 1999.