LASER & PHOTONICS REVIEWS
www.lpr-journal.org

Adaptive Structured-Light 3D Surface Imaging with Cross-Domain Learning

Xinsheng Li, Shijie Feng, Wenwu Chen, Ziheng Jin, Qian Chen,* and Chao Zuo**

The rapid development of artificial intelligence (AI) technology is leading a paradigm shift in optical metrology, from physics- and knowledge-based modeling to data-driven learning. In particular, the integration of structured-light techniques with deep learning has garnered widespread attention and achieved significant success due to its capability to enable single-frame, high-speed, and high-accuracy 3D surface imaging. However, most algorithms based on deep neural networks (DNNs) face a critical challenge: *they assume the training and test data are independent and identically distributed, leading to performance degradation when applied across different image domains, especially when test images are acquired from unseen systems and environments*. A cross-domain learning framework for adaptive structured-light 3D imaging is proposed to address this challenge. This framework's adaptability is enhanced by a novel mixture-of-experts (MoE) architecture, capable of dynamically synthesizing a network by integrating contributions from multiple expert DNNs. Experimental results demonstrate the method exhibits superior generalization performance across diverse systems and environments over both "specialist" DNNs developed for fixed domains and "generalist" DNNs trained by brute-force approaches. This work offers fresh insights into substantially enhancing the generalization of deep-learning-based structured-light 3D imaging and advances the development of versatile, robust AI-driven optical metrology techniques.

1. Introduction

"If you can't measure it, you can't make it" expresses the great significance of metrology in our world.[1] As a representative optical metrology technique, structured-light 3D surface imaging is a powerful and versatile tool for acquiring 3D shapes, distinguished by its non-contact interaction with measured objects, as well as its high-speed and high-resolution data acquisition capabilities.[2–7] In structured-light 3D imaging, sinusoidal patterns are typically used to encode illumination for high-precision 3D measurements.[8–14] By demodulating the phase information from the captured images and relating it to 3D coordinates, high-accuracy surface information is extracted.

In the past few years, deep learning technology has brought groundbreaking progress in many fields,[15–19] such as natural language processing, computer vision, and biomedical research. In structured-light 3D surface imaging, we have observed many successful applications leveraging artificial intelligence (AI).[20] For fringe analysis, deep neural networks (DNNs) have been trained to demodulate high-precision phase information from a single fringe image[21] and also been applied to extracting phase from non-sinusoidal fringes.[22] For fringe denoising, researchers trained DNNs to remove speckle noise and improved the signal-to-noise ratio (SNR) of captured fringe.[23] For phase unwrapping, deep learning techniques have been proposed to improve the reliability of phase unwrapping, enabling novel strategies for both spatial and temporal phase unwrapping.[24,25] For high-speed 3D reconstruction, DNNs have been carefully designed to demodulate 3D shapes from a single fringe image generated by techniques such as spatial frequency multi-plexing[26], epipolar geometry,[27] and color channel multiplexing.[28]

Although deep learning methods have shown great potential for the above applications, they are not trouble-free. As opposed to traditional algorithms based on mathematical and physical models, deep learning techniques are essentially driven by data and therefore their performance largely relies on the quality of captured data. So, the generalization capability of DNN, that is whether DNN models trained on a certain amount of data still perform well on other unseen data, has always been a crucial issue. To enhance DNN's generalization capability, researchers usually collect as many types of samples as possible during the training process for reducing the possibility of DNN facing out-of-distribution (OOD) test data. To enhance the generalization capability of DNN in fringe analysis, ensemble deep learning has been introduced, where DNNs of different architectures are used to capture different features of the same test object and the phase is predicted by combining the contributions of all these

X. Li, S. Feng, W. Chen, Z. Jin, Q. Chen, C. Zuo
Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, Jiangsu Province 210094, China
E-mail: shijiefeng@njust.edu.cn; chenqian@njust.edu.cn; zuochao@njust.edu.cn
X. Li, S. Feng, W. Chen, Z. Jin, Q. Chen, C. Zuo
Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense
Nanjing, Jiangsu Province 210094, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/lpor.202401609

DOI: 10.1002/lpor.202401609

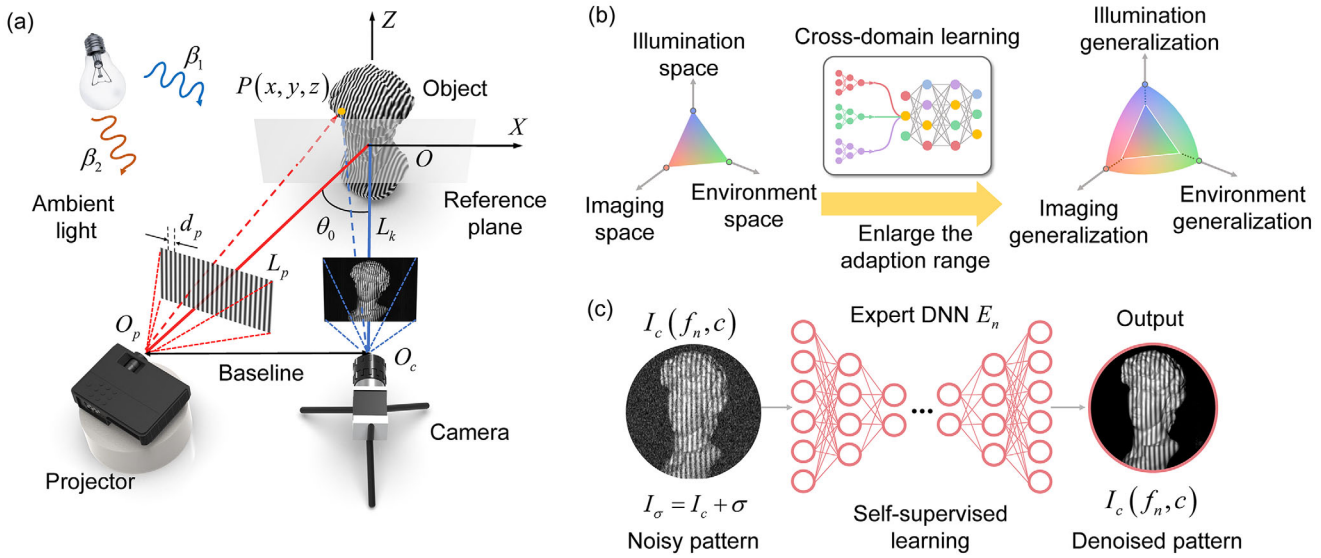

Figure 1. Structured-light 3D imaging using cross-domain learning with enhanced generalization. a) Schematic of structured-light 3D imaging using fringe projection. b) Our method expands the adaption range of illumination, imaging, and environmental attributes, thereby improving the generalization capability of deep learning model for unseen systems and environments. c) Expert DNN learns to extract features of fringe images through self-supervised learning by using a denoising autoencoder framework.

DNNs.[29] To collect a large amount of data, a digital twin of a real-world structured-light system can be built by 3D computer graphics software (e.g., Blender[30]), which can generate various realistic simulation data efficiently and improves the DNN's ability to handle different types of objects.[31,32] In addition, in lensless imaging[33] and computational ghost imaging,[34] techniques such as data augmentation and numerical simulations have also been proposed to increase the diversity of data. Furthermore, Bayesian neural networks (BNNs) are employed to quantify prediction uncertainty, enabling the deep learning model to identify OOD data and provide reliable predictions.[35,36]

Although these methods improve the generalization of DNN to some extent, they have not yet solved the generalizability problem fundamentally. All of these methods rely on a fundamental assumption that the training data and the testing data are independent and identically distributed (i.i.d), which indicates the test data and training data are supposed to be collected under the same system and the environment. When the system and the environment is different, the distribution of collected data can change significantly even if measured objects remain unchanged. Therefore, most DNNs compromise when they are cross-utilized between different systems and imaging environments. Moreover, due to the constraints imposed by this assumption it's often required to collect a large amount of new data and retrain DNNs from scratch every time the system or the environment changes, resulting in a huge cost of time and labor. Therefore, *how to enable DNNs to be cross-utilized in different domains remains a big challenge.*

Inspired by recent successful applications of dynamic neural networks,[37] for the first time, we propose an adaptive cross-domain learning (CDL) framework for generalized structured-light 3D surface imaging. The adaptability is achieved by a new mixture-of-experts (MoE) architecture that consists of a set of expert DNNs that performs feature extraction for fringe images from different domains and a gating neural network that controls synthesis of these expert DNNs dynamically. Our approach can adaptively and dynamically fuse different experts DNNs, establishing a customized way for analyzing various structured-light fringe images. Experimental results show that our method exhibits excellent generalization for both in-distribution (ID) and OOD scenarios. For ID scenes where the system and the imaging conditions are unchanged, the 3D measurement accuracy of unseen objects exceeds traditional DNNs thanks to more information acquired by the collaborations of multiple experts DNNs. For OOD scenarios, where structured-light systems and environments are changed, such as the spatial frequency of projected fringes, the system's working distance, the ambient light and so on, robust and high-accuracy 3D measurements can be achieved with our CDL framework. The presented approach provides a novel and flexible solution for adaptive structured-light 3D imaging.

2. Theory

2.1. Principles of Structured-Light 3D Imaging

A typical structured-light 3D imaging system using fringe projection is shown in **Figure 1**a, where a projector projects a sinusoidal structured-light image onto the measured object and a camera captures the pattern from a different angle. Due to the modulation by the object's shape, a deformed fringe pattern can be captured. The projected sinusoidal fringe pattern can be written as

$$I_p(x,y) = A^p(x,y) + B^p(x,y)\cos(2\pi fx) \quad (1)$$

where A^p represents the background intensity, B^p the amplitude, and f the spatial frequency of the sinusoidal pattern. As shown

in Figure 1a, we assume that the baseline connecting the optical center of the camera and that of the projector is parallel to the reference plane. On the reference plane, we set up a world coordinate system whose origin is the intersection of the optical axis of the camera and that of the projector. The measured point is $P(x, y, z)$ and its captured intensity can be expressed as [38]:

$$I_c = r\left\{A^p + B^p \cos\left[\frac{2\pi\cos\theta_0}{d_p}\left(x - \frac{2x^2\sin\theta_0}{L_p}\right) - \frac{2\pi z\sin\theta_0}{d_p}\left(1 - \frac{2x\sin\theta_0}{L_p}\right)\right] + \beta_1\right\} + \beta_2 \tag{2}$$

For brevity, we omit the pixel coordinate (x, y). In Equation (2), θ_0 is the angle between the optical axis of the camera and that of the projector, L_p the distance between the projector center and the origin O, r the reflectivity of the object, d_p the fringe pitch, β_1 the ambient light reflected by the object, and β_2 the ambient light directly entering the camera. Details of derivation of Equation (2) can be found in the Section S2 (Supporting Information). Equation (2) can be simplified as

$$I_c = A' + B'\cos\left[2\pi\nu(x)(x - z\tan\theta_0)\right] = A' + B'\cos\phi \tag{3}$$

where $A' = rA^p + r\beta_1 + \beta_2$ represents the background intensity of the captured pattern, $B' = rB^p$ the fringe amplitude, and $\nu(x)$ the local frequency of the fringe pattern at the measured surface. ϕ is the phase to be measured and it can be calculated as

$$\phi = \arctan\frac{\alpha B'\sin\phi}{\alpha B'\cos\phi} = \arctan\frac{M}{D} \tag{4}$$

where α is a constant dependent on the phase demodulation algorithm (e.g., $\alpha = 0.5$ in FTP and $\alpha = N/2$ in N-step PS algorithms). Using deep learning techniques, a DNN can be trained to predict M and D, which are then fed into an arctangent function to calculate the phase.[21] By converting the phase into height,[39] we can reconstruct the 3D surface of the measured object.

2.2. Structured-Light 3D Imaging with Various Systems and Environments

According to Equation (3), we can see that the phase ϕ depends not only on the height z but also on the local frequency $\nu(x)$ and the angle θ_0. Therefore, the phase will change for different systems with varying $\nu(x)$ and θ_0 even if the object is unchanged. Next, we further analyze the local frequency $\nu(x)$. Based on Equations (2) and (3), and $d_p = m_p/f$ (where m_p is the magnification factor of the projection system), $\nu(x)$ can be represented as

$$\nu(x) = f\frac{\cos\theta_0}{m_p}\left(1 - \frac{2x}{L_p}\sin\theta_0\right) = f\frac{\cos\theta_0}{m_p}\left(1 - \frac{x\sin 2\theta_0}{L_k}\right) \tag{5}$$

According to Equation (5), we find that: (1) there is a linear relationship between the frequency f of the projected pattern and the local frequency $\nu(x)$; (2) the local frequency $\nu(x)$ exhibits a nonlinear relationship with the angle θ_0; (3) assume that the relative position between the camera and the projector remains constant, the variation of L_k can cause varying θ_0 and m_p. As a result, the local frequency $\nu(x)$ will be changed. Thus, we can see that for different measurement systems, changes in frequency f, angle θ_0, and working distance L_k can cause variations in the grating's local frequency $\nu(x)$.

In addition to the measurement system, the phase calculation is often influenced by various environments. The effects of the environment on 3D imaging mainly stem from vibrations and ambient light. Assuming that the phase is calculated with a single fringe pattern, we ignore the effect of vibrations in the environment. So, we focus on the effect of ambient light on phase measurement. According to Equation (3), the ambient light can cause variations in the intensity of the captured grating. To analyze its effect, we define the contrast of the captured grating as

$$c' = \frac{B'}{A'} = \frac{rB^p}{rA^p + r\beta_1 + \beta_2} \tag{6}$$

where c' characterize the SNR of the captured grating. For different measurement environments, the ambient light are usually different. According to Equation (6), when the ambient light (β_1 and β_2) changes, the contrast c' of the captured fringe will change accordingly, resulting in changes of SNR of the captured grating. Assume the contrast of the projected grating is $c = B^p/A^p$ and substitute it into Equation (6), we have

$$c' = \frac{rA^p}{rA^p + r\beta_1 + \beta_2}c \tag{7}$$

According to Equation (7), we find that c' can also be modulated by the contrast c of the projected grating.

Based on the above analysis, the frequency of the projected pattern, the angle between the optical axes of the camera and projector, the working distance, and the ambient light conditions all vary across different measurement systems and environments. These variations result in diverse local frequency and intensity distributions in the captured images, which can affect phase computation and 3D reconstruction. Consequently, when a DNN is trained only on images from a specific system and environment, it struggles to accurately compute phase and 3D coordinates for images obtained from other unseen systems and environments.

2.3. Development of the Cross-Domain Learning Framework

To adapt a deep learning model to different systems and environments, we develop an adaptive CDL 3D imaging framework. As shown in Figure 1b, our framework can expand the adaption range of illumination, imaging, and environmental attributes, thus improving the generalization capability for unseen systems and environments. For cross-domain learning, a deep learning model should be exposed to diverse systems and environments during training. Instead of varying parameters such as f, θ_0, and L_k for each system, which can be time- and labor-intensive, we propose an efficient training data acquisition strategy that enables us to quickly obtain corresponding fringe images across different systems and environments. Specifically, we observe that changes in the projected pattern's frequency f in a fixed system

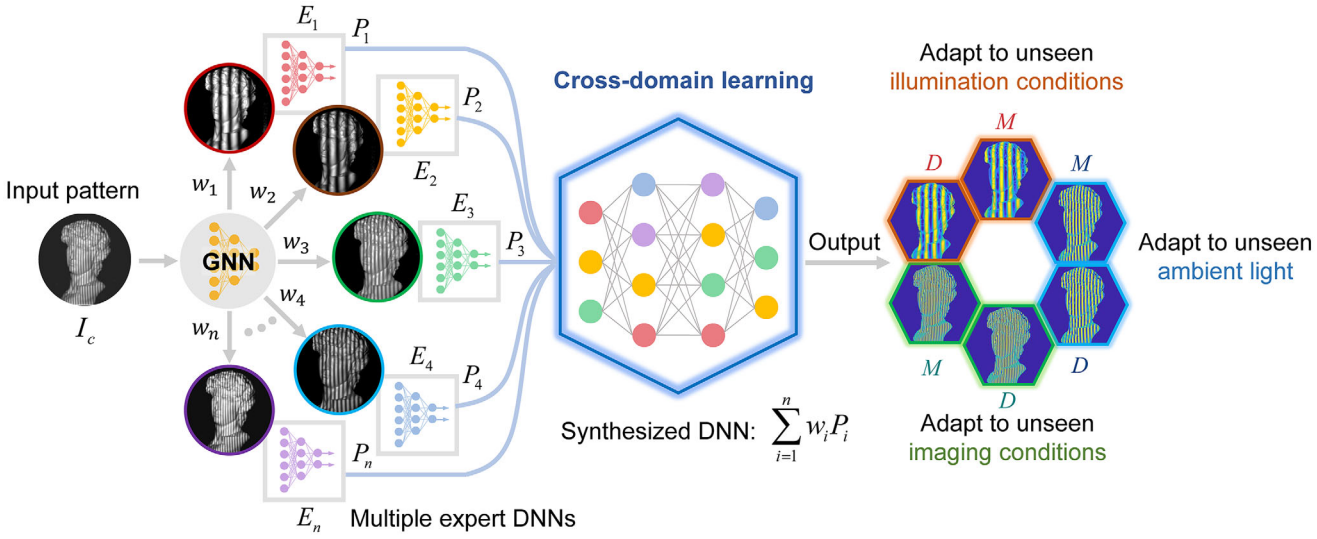

Figure 2. Schematic of our adaptive cross-domain learning framework. It is a hierarchical deep learning framework that generates adaptive behavior by combining multiple DNNs together to produce versatile predictions. The gating neural network (GNN) generates variable weights to fuse the parameters of all expert DNNs such that newly captured fringe patterns are analyzed by dynamically synthesized DNN. The input is an unseen fringe image and the output are *M* and *D* that can be fed into the arctangent function to calculate the phase of input pattern. The framework can adapt to unseen illumination, imaging, and ambient light conditions owing to the advantage of blending useful learned features collectively from the consortium of experts.

can effectively approximate the effects of modifying θ_0 and L_k, as described by Equation (5). This allows for the simulation of system variations by adjusting f. Regarding the measurement environment, the impact of varying ambient light is captured by adjusting the contrast c of the projected pattern, according to Equation (7), providing a more efficient alternative to directly altering the ambient lighting conditions. Our framework consists of multiple expert DNNs, which can be used to dynamically synthesize a new DNN through a gating neural network (GNN) to perform the fringe analysis task.

As shown in Figure 1c, the expert DNN is designed carefully to handle grating images with frequency f and fringe contrast c. Assuming the frequency candidates of projected fringes are $f = \{f_1, f_2, \ldots, f_n\}$ and the contrast candidates of projected fringes are $c = \{c_1, c_2, \ldots, c_n\}$, we have expert DNN E_n to learn to process fringe patterns $I(f_n, c)$ with a specific frequency f_n and a sets of contrast c, where $n = 1, 2, 3, \ldots$. By modifying the structure of Res-UNet,[40] we develop the architecture of expert DNN. It consists of a fully convolutional neural network with an encoder-decoder structure, where each encoder and decoder is composed of four convolutional blocks. To avoid challenges of acquiring a large training dataset and the significant human labor and time costs associated with labeling, inspired by the idea of denoising autoencoders as in ref. [41], we propose a self-supervised pre-training method for expert DNN based on denoising autoencoders. As shown in Figure 1c, the input of the expert DNN is fringe images with additional noise added, and its output is the original fringe image without any noise. By training the network to remove the noise from input fringe images, the network can learn how to extract features of the fringe patterns. The training process does not require any training labels, thereby avoiding labor and time costs. More details on the structure of the expert DNN and its training strategy are provided in Section S3 (Supporting Information).

The schematic of our adaptive CDL framework is shown in **Figure 2**. This framework employs a hierarchical deep learning approach that combines multiple DNNs to generate adaptive behaviors and produce versatile predictions. A GNN dynamically assigns variable weights to integrate the parameters of all expert DNNs, enabling the analysis of newly captured fringe patterns by a synthesized DNN. The model takes an unseen fringe image as input, with the outputs being M and D, which are fed into the arctangent function to calculate the phase of the input pattern (Equation 4). The framework is capable of adapting to varying illumination, imaging, and ambient light conditions by effectively merging the useful learned features from the consortium of expert networks. The GNN is developed by modifying the architecture of VGG.[42] It comprises convolutional layers and pooling layers organized into a total of five convolution blocks. The feature maps from the final convolution block are flattened and passed through fully connected layers to generate weights for allocating to the expert DNNs. More details on the structure of GNN are provided in Section S4 (Supporting Information). The GNN analyzes and extracts features from the input fringe pattern to compute weights for each expert DNN. By adjusting the weights, GNN controls the collaboration among the expert DNNs. The fusion process of the parameters of expert DNNs can be expressed as

$$P_{CDL} = \sum_{i=1}^{n} w_i P_i \tag{8}$$

where P_i represents the parameters of expert DNN E_i and P_{CDL} the parameters of the dynamically synthesized network. The sum of all weights is 1, i.e., $\sum_{i=1}^{n} w_i = 1$. It is noteworthy that since the task of the expert DNN during pre-training is the self-supervised denoising, we retain the encoder parameters of each expert DNN and discard their decoder parameters during the training process of the whole CDL framework.

To further enhance the generalization capability, we incorporate the process of Fourier transform into the loss function and propose a loss function that takes advantage of consistencies in both spatial domain and frequency domain. So, the loss function of this work is expressed as

$$L_{loss} = \lambda_1 L_{Spat} + \lambda_2 L_{Freq} \tag{9}$$

where λ_1 and λ_2 are the weights of L_{Spat} and L_{Freq} respectively. L_{Spat} is the loss function that imposes consistency in spatial domain, which ensures that M and D predicted by our model remain consistent with the label in terms of spatial intensity distribution. The calculation of L_{Spat} is given by

$$L_{Spat} = \frac{\sum_{i=1}^{k} (y_{pred} - y_{true})^2}{k} \tag{10}$$

where y_{pred} represents a pair of M and D predicted by the CDL framework, while y_{true} denotes their ground truth and k is the total number of pixels in the image. Then, L_{Freq} is the Fourier-domain loss (FL) function, which constrains the predicted results to be consistent with the label in the frequency domain. The expression of L_{Freq} is

$$L_{Freq} = \frac{\sum_{i=1}^{k} |\mathcal{F}(y_{pred}) - \mathcal{F}(y_{true})|}{k} \tag{11}$$

where $\mathcal{F}$ represents the discrete Fourier transform. An ablation study of L_{Freq} is presented in Supporting Information, demonstrating the contribution of FL. This hybrid loss function of this work leverages the generalization capability of the physical model, thereby further enhancing the adaptability of our CDL framework for fringe images collected from unseen systems and environments.

3. Experiments

To validate the effectiveness of our method, we constructed a fringe projection system. To collect the training data, fringe images of various objects with different system parameters and environment were captured. The captured fringe patterns are 8-bit gray-scale images. In the data preprocessing stage, the input fringe pattern was divided by 255 for normalization before being fed into the DNNs. All models were implemented by the Tensorflow and computed on a graphic card (GTX 4090, NVIDIA). Section S1 (Supporting Information) provides further details on our system.

3.1. Training of the Proposed CDL Framework

In our experiments, a CDL framework with three expert DNNs was developed for the validation of our method. To train the expert DNNs, candidate frequencies are $f = \{24, 64, 128\}$ and candidate contrast $c = \{0.5, 1\}$. For the purpose of controlling variables, we altered the contrast c by changing the amplitude B^p of the projection grating, while keeping the background intensity A^p constant. Expert DNN E_1 was trained to handle fringe images of $f_1 = 24$ with $c_1 = 0.5$ and $c_2 = 1$, i.e., fringe images with the same frequency but two types of contrast. Similarly, experts DNNs E_2 and E_3 were trained to handle fringe images of frequencies of $f_2 = 64$ and $f_3 = 128$ with these contrast, respectively. These fringe images were captured when the angle between the axes of the camera and the projector is $\theta_0 = 15°$, the system working distance $L_k = 1$m, and no ambient light presents.

First, we pre-trained the three expert DNNs. For each expert, the training dataset contained 450 actual fringe images (involving 450 different scenes), and the test dataset contained 150 fringe images (involving 150 different scenes). Subsequently, we employed the self-supervised learning method described in Section 2 to pre-train each expert DNN. After pre-training, we established the CDL framework. Since the expert DNNs had acquired certain prior knowledge after pre-training while the GNN hasn't been trained, we carried out a warm-up training process for GNN to prevent overfitting. In the first 50 epochs, we only updated the weights of the GNN and frozen those of expert DNNs. After 50 epochs, parameters of both GNN and expert DNNs were updated simultaneously. The warm-up training process allows for a deep integration between GNN and expert DNNs to achieve better collaborative performance. The weights λ_1 and λ_2 in the loss function L_{loss} were set to 1 and 0.1, respectively, during the training process. We trained the CDL framework with the learning rate at 1e-4 by 150 epochs. In addition, when we trained the CDL framework, we utilized data augmentation to expand the training dataset to three times its original size. Details about data augmentation and its ablation study are provided in Section S5 (Supporting Information).

For comparison, we also developed a baseline DNN and a generalist DNN. The baseline DNN had the same structure as the expert DNN. During training, the baseline DNN only saw fringe images collected from a specific system or environment. As for the generalist DNN, it also shared the same structure as the expert DNN. However, during training it was exposed to all of the fringe images captured by various systems and environments.

3.2. 3D Imaging with Seen Systems and Environments

First, we tested our method by using three seen domains, namely $I_c(f_1 = 24, c_1 = 0.5)$, $I_c(f_2 = 64, c_2 = 1)$ and $I_c(f_3 = 128, c_2 = 1)$. We used the 12-step phase-shifting method to calculate the ground-truth (GT) phase, and assessed the results by computing the average phase error. For the baseline DNN, the fringe images used for training and testing were from the same domain, i.e., with the same system and environmental parameters. For the generalist DNN, it was exposed to images from these three domains simultaneously during both training and testing processes. The experimental results are presented in **Figure 3**. According to the mean absolute error (MAE) of the retrieved phase as shown in Figure 3a, it can be observed that our framework, the generalist DNN, and the baseline DNN all achieved high accuracy in phase measurement for these fringe images with different frequencies and contrast. By comparison, our CDL framework exhibited higher precision in phase measurement for reconstructing objects' edges and fine details. Figure 3b illustrates the 3D

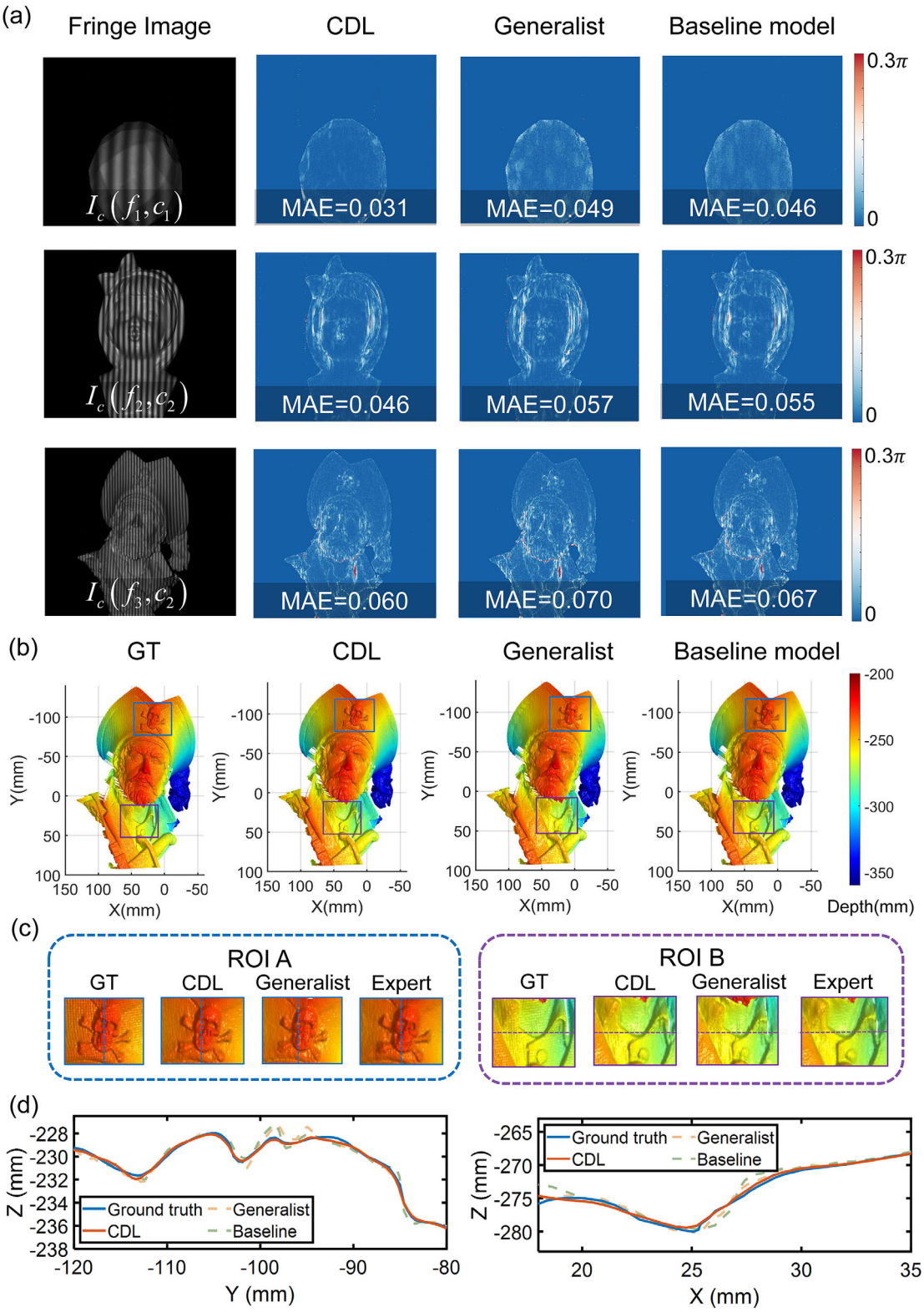

Figure 3. Experimental results on 3D imaging with fringe images obtained from seen systems and environments. Panel (a) displays the distribution of phase errors, where the first row to the third row correspond to the cases of I_c ($f_1 = 24, c_1 = 0.5$), I_c ($f_2 = 64, c_2 = 1$) and I_c ($f_3 = 128, c_2 = 1$). It is noted that these objects are not seen during training. Panel (b) illustrates the 3D reconstructions obtained by the 12-step phase-shifting method (ground truth), our CDL framework, the baseline DNN, and the generalist DNN for the fringe image I_c ($f_3 = 128, c_2 = 1$). Panel (c) presents the comparison of 3D reconstructions of two ROIs. Panel (d) compares the cross-sectional 3D profile calculated by these methods for the two ROIs.

Table 1. Phase MAE for the CDL framework, the baseline model, and the generalist model for several seen systems and environments.

Domains	$f_1 = 24$		$f_2 = 64$		$f_3 = 128$	
	$c_1 = 0.5$	$c_2 = 1$	$c_1 = 0.5$	$c_2 = 1$	$c_1 = 0.5$	$c_2 = 1$
Baseline model	0.086	0.048	0.067	0.044	0.069	0.047
Generalist	0.075	0.040	0.059	0.035	0.068	0.041
Our CDL	**0.065**	**0.037**	**0.054**	**0.034**	**0.060**	**0.040**

reconstructions of the object in the third row of Figure 3a. For comparison, we selected two regions of interest (ROIs) with rich details, and their 3D reconstruction results and cross-sectional views are shown in Figures 3c,d, respectively. We can see that for smooth areas these three 3D reconstructions are very close to the GT 3D model. However, for areas with depth changes, the proposed framework demonstrates higher accuracy compared to the baseline DNN and the generalist DNN.

Next, we expanded the testing scenario to include 150 different scenes and enlarged the range of seen domains to six kinds of fringe patterns. The results are shown in **Table 1**. For the baseline DNN, it performed well with these different types of fringe images as the system and environment remained unchanged. For the generalist DNN, we observed that its performance was superior to the baseline DNN, indicating that the generalist DNN has learned features of fringe images in different domains. Compared to the baseline DNN and the generalist DNN, our CDL framework demonstrated smaller phase measurement errors for all these fringe images with different frequencies and contrast. This suggests that our framework can dynamically generate DNNs that are more suitable for handling different fringes through combining contributions from different expert DNNs, leading to further reduced measurement errors. This experiment demonstrates that our framework can not only enhance phase accuracy for seen systems and environments but also this strategy is superior to the brute-force method of feeding all data to the neural network all at once.

3.3. 3D Imaging with Unseen Systems and Environments

In this experiment, we tested the proposed method by using different systems and environments. Specifically, we changed the spatial frequency, the fringe contrast, the angle between the projector axis and camera axis, the working distance, and ambient light intensity. For each case, 150 scenes were tested and the MAE was calculated over all these scenes.

Firstly, we tested our method using fringe patterns with unseen frequencies $f = \{30, 48, 72, 96, 120\}$. The results are shown in **Figure 4a**. The baseline DNN, trained with $f = 72$, had the lowest error at this frequency (0.066 rad) but showed increased errors as the frequency deviated from 72. In contrast, the generalist DNN performed better across unseen frequencies, with reduced errors compared to the baseline DNN. Notably, the CDL framework achieved the minimum error at $f = 72$ (0.052 rad) and outperformed other methods at low frequencies (e.g., $f = 30$ with error 0.064 rad), as shown in Figure 4a. Moreover, we further tested fringe frequencies outside the range seen by expert DNNs. Figure 4b,c shows the performance of our framework for $f < 24$ and $f > 128$, respectively. The baseline DNN exhibited large errors and thus is not shown. We can see that the performance trends of the CDL framework and the generalist DNN were similar, with both showing increased errors as the frequencies deviated from the seen range. Specifically, errors increased sharply for frequencies below 12 and above 170, as shown in Figure 4b,c. However, the CDL framework consistently maintained lower errors than the generalist DNN, demonstrating better generalization for different fringe patterns of unseen frequencies.

For different systems, the fringe contrast, the angle between the camera optical axis and the projector optical axis, and the working distance are usually not the same. Therefore, we also tested the proposed framework under these varying conditions. In the first case, for different fringe contrasts, the trained contrasts were $c_1 = 0.5$ and $c_2 = 1$, while the test contrasts were $c = \{0.6, 0.7, 0.75, 0.8, 0.9\}$. The baseline DNN was trained with $c = 0.75$. In the second case, for different angles between the optical axis, the training angle was $\theta_0 = 15°$ for the CDL

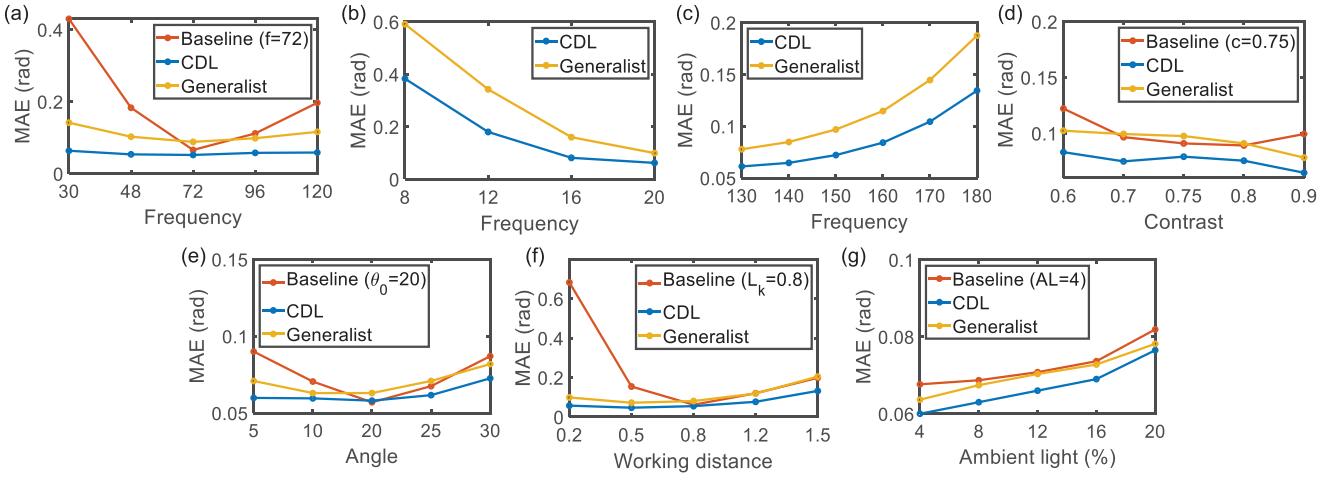

Figure 4. The phase error distribution of the CDL framework for unseen systems and environments. a) Unseen spatial frequency (interpolation). b) Unseen lower frequency (extrapolation). c) Unseen higher frequency (extrapolation). d) Unseen fringe contrast. e) Unseen angles between the optical axes of the projector and the camera. f) Unseen working distances. g) Unseen ambient light intensity.

framework and the generalist DNN, and $\theta_0 = 20°$ for the baseline DNN. Testing angles were $\theta_0 = \{5°, 10°, 20°, 25°, 30°\}$. In the third case, for varying working distances, training distances were $L_k = 1$m for the CDL framework and the generalist DNN, and $L_k = 0.8$m for the baseline DNN. Testing distances were $L_k = \{0.2, 0.5, 0.8, 1.2, 1.5\}$. The results for these three cases are shown in Figure 4d–f. The baseline DNN performs well within the trained conditions, such as at the contrast of $c = 0.75$, the angle of $\theta_0 = 15°$ and the working distance of 0.8 m, but exhibits a rapid increase in errors as the conditions diverge from the training set (Figure 4d–f). The generalist DNN shows better generalization than the baseline DNN, but still struggles as the contrast, the angle, and the distance deviate from those seen during training. In contrast, the CDL framework consistently outperforms both the baseline DNN and the generalist DNN, with smaller errors under these conditions.

Finally, we tested performance of our method under varying ambient light (AL) conditions. We placed an adjustable light source next to the measurement system to control ambient lighting (JINBEI EFl-200). Training data were collected without ambient light for the CDL framework and the generalist DNN, and at AL = 4% for the baseline DNN. Figure 4g shows that the baseline DNN had the smallest error at AL = 4%, but errors increased rapidly with higher AL. The generalist DNN performed similarly to the baseline DNN at AL = 12% and slightly better otherwise. The CDL framework demonstrates the smallest errors, with a maximum of 0.077 rad at AL = 20%.

For a more intuitive comparison, we compared the phase error maps and 3D reconstructions in **Figure 5**. In Figure 5a, the first to the fourth rows show the fringe images captured with unseen frequency f, contrast c, angle θ_0, and ambient light, respectively. For the unseen frequency $f = 96$, the baseline DNN shows a large phase error of 0.27 rad. Although the generalist DNN partially suppressed the error, it is still relatively large (0.11 rad). In comparison, our CDL framework achieved the smallest error which is only 0.061 rad. Regarding the unseen contrast $c = 0.8$, the baseline DNN demonstrates some degree of generalization but still has an phase error of 0.11 rad. In contrast, our CDL framework achieved the smallest error of 0.08 rad. Next, for the unseen angle $\theta_0 = 5°$, the baseline DNN shows a large phase error of 0.1 rad. The generalist DNN reduced the error to 0.094 rad. In contrast, the CDL framework demonstrates the smallest error of 0.079 rad. For the unseen ambient light AL = 8%, all these networks show adaptability to the captured fringe image, where the CDL framework demonstrates the best generalization capability and the smallest measurement error of 0.068 rad.

We then converted the phase of the first scene shown in Figure 5a into 3D reconstructions and the results are shown in Figure 5b. We find that the baseline DNN and the generalist DNN show significant ripple errors in the facial and neck regions of the reconstructed statue. For our method, however, these areas have been recovered smoothly, indicating that our framework is not sensitive to fringe patterns of unseen frequency. To further analyze the 3D reconstructions, we selected two ROIs, and the 3D reconstructions and cross-sectional views are shown in Figure 5c,d respectively. We can observe severe reconstruction errors in the eyes and the neck from the 3D results of the baseline DNN and the generalist DNN. In contrast, our CDL framework overcame the errors caused by unseen frequencies and accurately reconstructed the 3D shape of the statue's face. The 3D reconstruction of our CDL framework is comparable to the ground-truth 3D model that was measured by 12-step PS algorithm.

Moreover, to further investigate the performance of our network in complex imaging environments, we introduced two light sources to illuminate a measured industrial workpiece from both sides. As shown in **Figure 6**d, both light sources alternated flashing synchronously with a 200 ms interval, and the camera captured the workpiece at different moments. Therefore, in this case, the camera captured the measurement object under three different ambient lighting conditions: left-side ambient light, right-side ambient light, and ambient light from both sides simultaneously. The fringe frequency used for testing was $f = 128$, and the ambient light intensity was set to AL = 8%. Figure 6a–c display the fringe images, phase errors, and 3D reconstruction results under these scenarios. It can be seen that our CDL framework maintains stable phase measurement accuracy and 3D measurement results in these complex imaging environments. Figure 6e presents cross-sectional curves of ROIs of reconstructed 3D models. The generalist DNN shows significant error fluctuations when facing complex ambient lighting conditions, while the CDL framework's results are consistently close to the ground truth. These results highlight the adaptability and stability of our CDL framework in handling complex imaging environments.

3.4. Dynamic 3D Reconstructions Using CDL Framework

3.4.1. Rotating Fan Blades

We further tested the performance of the proposed method for dynamic 3D measurements. The test object is a fan and its speed is adjustable. The frequencies used for training our CDL framework were still $f = \{24, 64, 128\}$, while the baseline DNN was trained on fringe patterns with a frequency of $f = 128$. During the testing process, three unseen frequencies of projected patterns are $f = \{110, 115, 120\}$, which correspond to different speeds $V = \{425, 400, 375\}$ rotations per minute (RPM) of the fan. **Figure 7**a shows the fringe images of the fan of different frequencies collected at $T = \{10, 20, 30\}$ ms. Figure 7b demonstrates the 3D shape obtained by our method and the baseline DNN for these moments, where the left column shows the results of the baseline DNN (in the red background) and the right column those of our framework (in the blue background). Visualization 1 demonstrates the overall 3D reconstruction of the rotating fan. It can be observed that the 3D reconstruction error of the baseline DNN becomes more obvious as the frequency decreases. Figure 7c illustrates the cross-sectional profiles of the blades of the rotating fan shown in Figure 7b, where the first and the second row shows the results of the baseline DNN and our CDL framework, respectively. From the 3D reconstructions of the baseline DNN, it is evident that when the grating frequency decreases to 110, the ripple errors are found on the reconstruction blade surface. In contrast, our CDL framework shows more reliable performance in handling these dynamic fringes. Despite it did not see these dynamic fringe images of different frequencies before, it was still capable of accurately reconstructing the profile of the rotating blades.

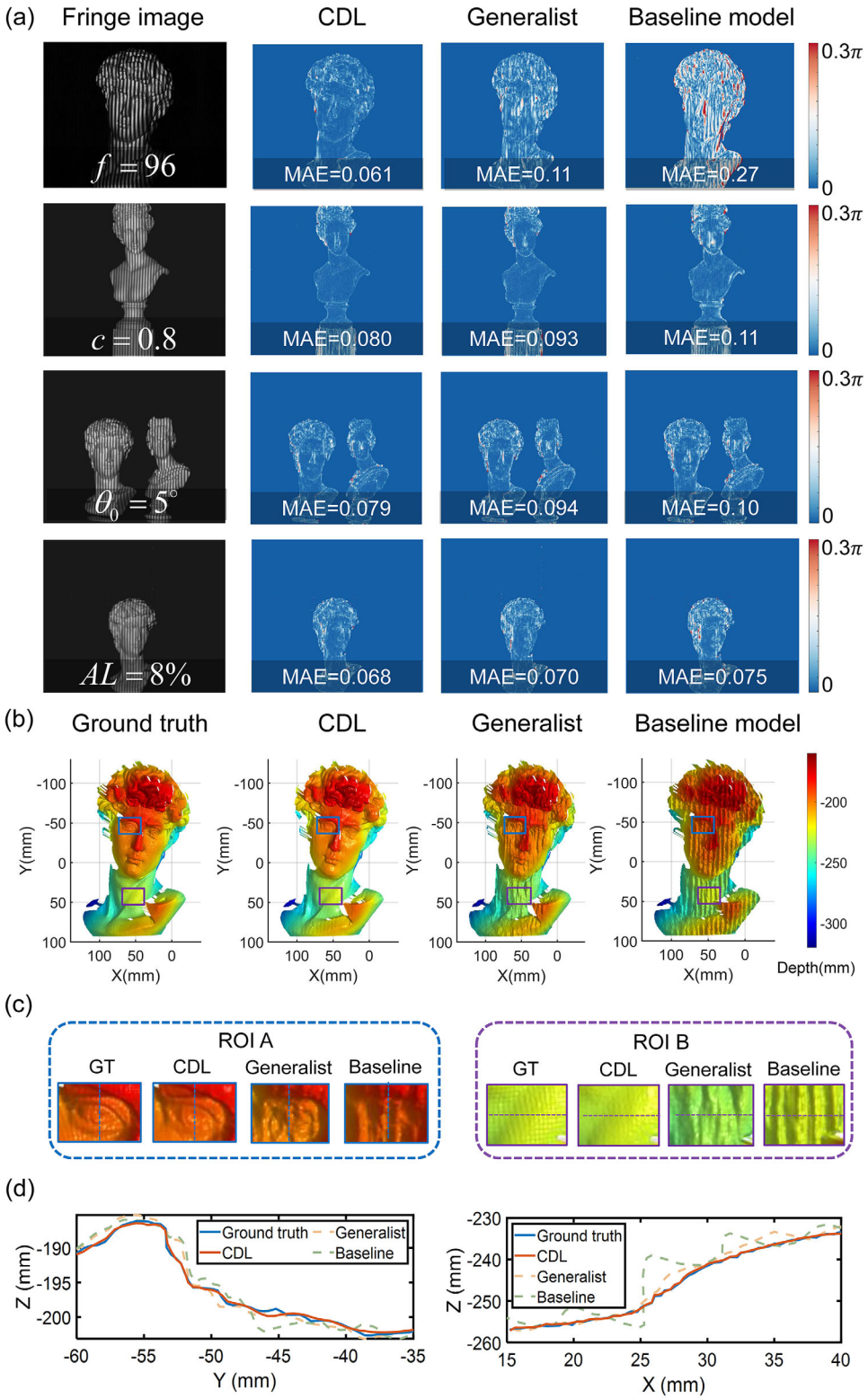

Figure 5. Experimental results on phase measurements and 3D imaging with fringe images obtained from unseen systems and environments. a) Fringe images obtained by unseen systems and environments, and the phase error of our method, the generalist DNN and the baseline model. b) 3D reconstructions obtained by the 12-step phase-shifting method (ground truth), our CDL framework, the generalist DNN, and the baseline DNN for the unseen fringe image of $f = 96$ as shown in (a). c) Comparison of 3D reconstructions for two ROIs. d) Comparison of cross-sectional 3D profile calculated by these methods for the two ROIs.

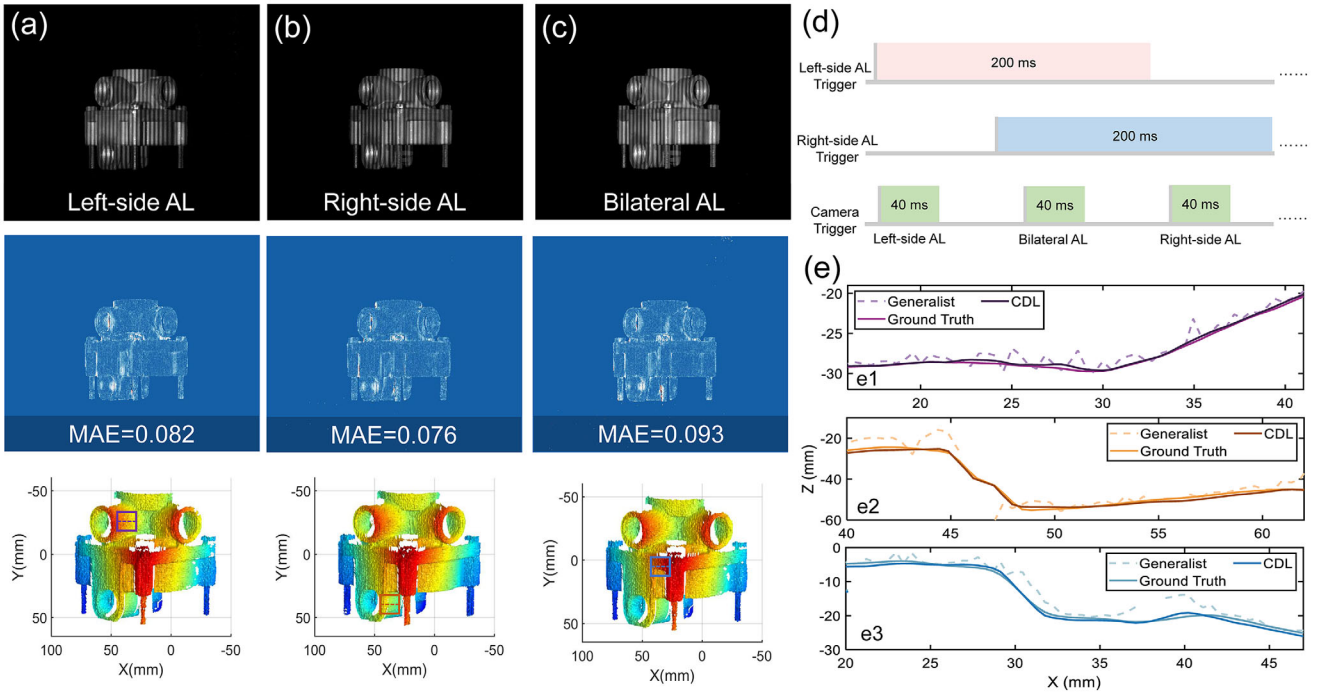

Figure 6. 3D reconstruction under complex lighting conditions. a) Input fringe images captured under left-side ambient light, b) right-side ambient light, and c) bilateral ambient light, along with phase errors and 3D results obtained using our CDL framework. d) Triggers of two ambient light sources and the camera. e) The cross-sections of ROIs in 3D reconstructions, where (e1–3) correspond to left-side ambient light, right-side ambient light, and bilateral ambient light, respectively.

Furthermore, we randomly selected three points on the blades (points A, B, and C in Figure 7d) to demonstrate the cyclic displacements of the blades. Displacement in the z (out-of-plane) direction at the chosen point locations are plotted as a function of time, as shown in Figure 7d. Regarding the fringe of an unseen frequency of 110, the plot shows that the fan has a rotation period of approximately 140 ms, corresponding to a rotation speed of 425 RPM. For the fringe image with frequencies of 115 and 120, the plots show that the fan has a rotation period of approximately 150 and 160 ms, corresponding to the rotation speed of 400 RPM and 375 RPM, respectively. These plots also indicate a good repeatability of the proposed approach. Furthermore, we tested the performance of our method for dynamic 3D measurements in different unseen ambient light. The experimental results show that the well-trained CDL framework can adapt to fringe images captured under different ambient lighting conditions and accurately retrieve the 3D shape of dynamic objects. More details on this experiment are provided in the Section S6 (Supporting Information).

3.4.2. Dynamic Quantitative Evaluation

We conducted quantitative evaluations of our framework in dynamic scenarios by measuring standard ceramic spheres (**Figure 8a**). Consistent with the dynamic measurement experiment of the rotating fan in Section 3.4.1, the fringe frequency used in our system was $f = 120$, while the baseline network was trained with a fringe frequency of $f = 128$. Figure 8b displays the captured input fringe image at $T = 0$ ms. The ceramic spheres moved along the Y-axis, and Figure 8c,d illustrate the 3D reconstruction results at different time points and the corresponding error maps produced by our framework. At $T = 1000$ ms, the root mean square errors (RMSE) for the two spheres were 67.50 μm and 89.35 μm, respectively, with an RMSE of 56.50 μm for the center-to-center distance. And at $T = 6000$ ms, the RMSE for the two spheres were 50.50 μm and 65.80 μm, respectively, while the RMSE for the center-to-center distance was 62.98 μm.

For detailed analysis, Figure 8e,f shows the error variation over time for arbitrary points on spheres A and B during motion (points C and D in Figure 8c,d), measured using our CDL framework, the baseline network, and the generalist network. Overall, the baseline network exhibits the highest error compared to the other methods, with significant fluctuations, reflecting its limited generalization ability in the face of unknown conditions. The generalist network shows lower error but still has a relatively high level of error. In contrast, the CDL framework demonstrates the smallest error for the select points during the whole moving process. Figure 8g–i provides a detailed visualization of the 3D reconstruction results and error distributions for the three methods at $T = 3000$ ms. As can be seen, due to the discrepancy between the training fringe frequency and the test fringe frequency, the baseline network's prediction results exhibited significant ripple-like errors, causing the measured RMSE to exceed 500 μm. The generalist network demonstrated slightly better generalization and accuracy, but still had an RMSE greater than 200 μm. In contrast, our CDL framework, with its improved generalization and robustness, achieved the best performance, with the RMSE for

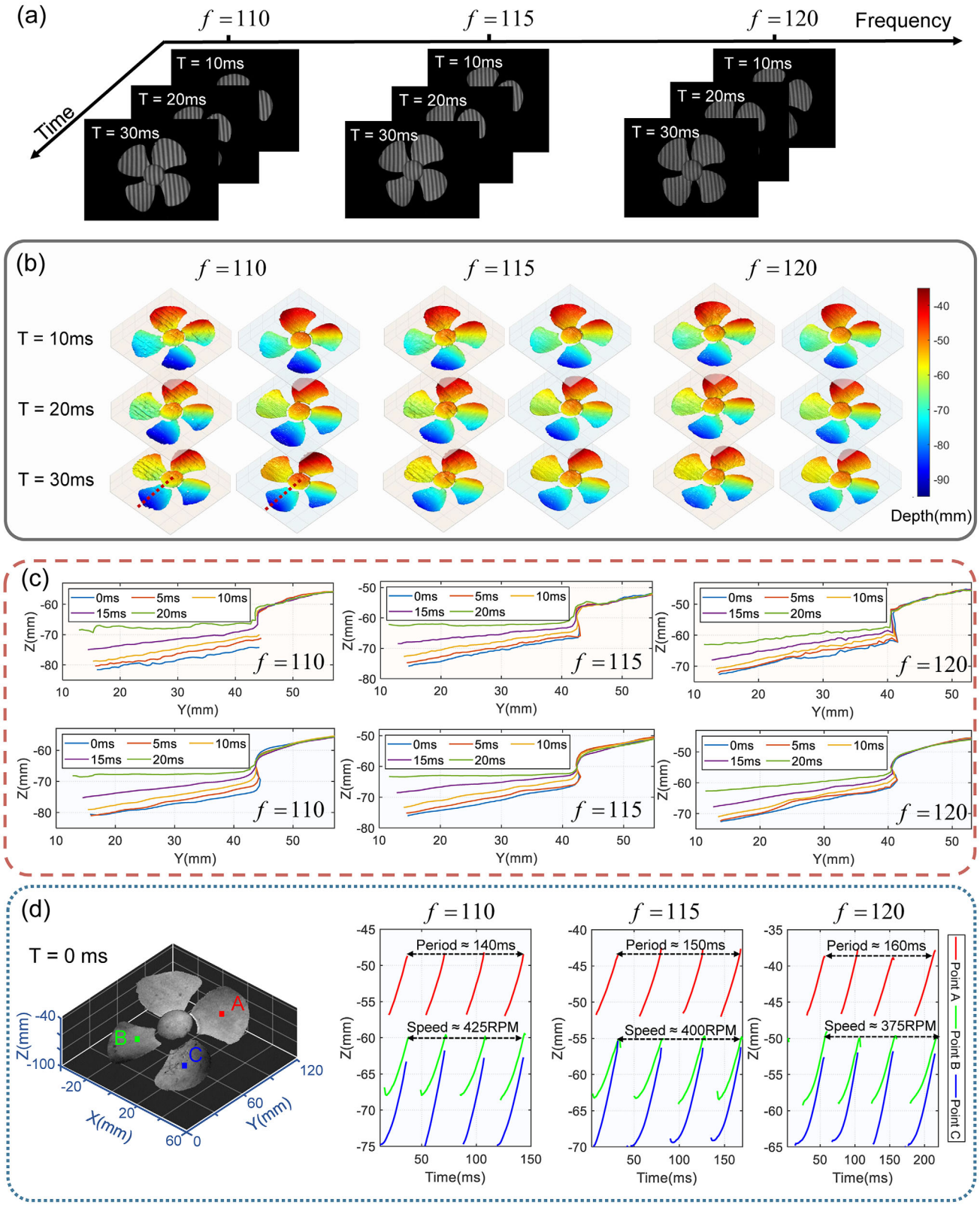

Figure 7. Dynamic 3D reconstructions of a rotating fan using our CDL framework. a) Fringe images of unseen frequencies captured at different moments. b) 3D reconstructions obtained by the baseline DNN (left column) and our CDL framework (right column). c) Cross-sectional profiles of the rotating blades reconstructed by the baseline DNN (the first row) and our method (the second row). d) Three points (A, B, and C) were selected and the height variations at these points were calculated, from which we computed the rotating period and speed of the fan.

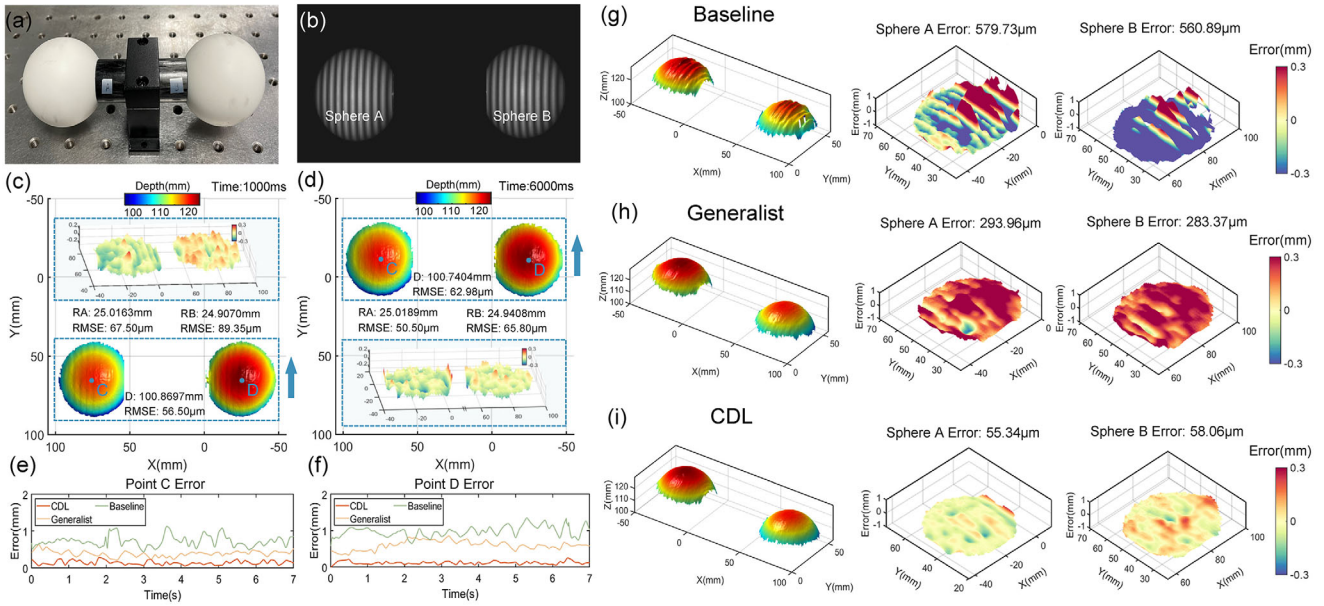

Figure 8. 3D reconstruction accuracy analysis of dynamic standard ceramic spheres. a) The standard ceramic spheres under test, each with a radius of 25 mm and a center-to-center distance of 100 mm. b) Input fringe pattern captured by the camera at $T = 0$ ms. c,d) 3D reconstruction results and error distributions obtained using our CDL framework at $T = 1000$ ms and $T = 6000$ ms, respectively. e,f) 3D measurement errors over time for two arbitrary points (points C and D) on spheres A and B, as measured by the baseline network, the generalist network, and our CDL framework. g–i) Comparison of the baseline network, the generalist network, and our CDL framework in terms of 3D reconstruction results and error distributions at $T = 3000$ ms.

the radius measurements of the two spheres remaining below 100 μm. Visualization 3 demonstrates the overall 3D reconstruction of the moving standard spheres.

3.5. Computational Complexity Analysis

In **Table** 2, we analyzed the baseline DNN, the generalist DNN, and our CDL framework according to the total number of parameters, the training time, and the accuracy in terms of average RMSE of the radii measurements of the two test spheres at different time points. We can see that the total number of parameters of the baseline DNN and the generalist DNN are the same which is 3.46 million. The reason is that they had the same structure in our experiments. The training time of them are 2 h and 10 h respectively. As the amount of data used to train the generalist DNN increased, the training time correspondingly became longer. The RMSE of the baseline DNN and the generalist DNN are 780.35 and 212.43 μm. By comparison, the total number of parameters and the training time of our CDL framework are 12.33 million and 14 h, which indicates that more computational resources and time costs have to be taken for training our CDL framework, as several expert DNNs are involved. Nevertheless, our framework demonstrates the RMSE of 73.88 μm, which is the smallest error among these DNNs.

Table 2. Comparison of the computational resource for the baseline network, the generalist network, and the CDL framework. Metrics include the total number of parameters (in millions), training time, and dynamic measurement errors for standard ceramic spheres.

Methods	Parameters [M]	Training time [h]	RMSE [μm]
Baseline model	3.46	2	780.35
Generalist	3.46	10	212.43
Our CDL	12.33	14	73.88

To further optimize the CDL framework, several strategies may be adopted to reduce computational resources and time costs without significantly sacrificing its accuracy. These strategies include the pruning technique (which eliminates less essential model parameters), the knowledge distillation technique (which transfers knowledge from a larger model to a more compact one), and so on ref. [43, 44]. By implementing these approaches, the computational burden of the CDL framework could be alleviated, while still maintaining its accuracy.

3.6. Experimental Analysis of GNN

Our CDL framework optimally fuses expert DNNs by leveraging a GNN that adaptively calculates the appropriate weights for each expert. Here, we analyzed the performance of the GNN to gain deeper insights into the behavior of our CDL framework. Specifically, we tested the CDL framework on 15 sets of fringe images with different spatial frequency and contrast. There are 150 test fringe images of unseen scenarios in each set. To obtain the distribution of GNN's weights, we used the UMAP technique that is able to demonstrate the distribution by reducing the dimensionality of the predicted weights.[45] As shown in **Figure** 9a, each point represents a prediction of GNN. Different colors were used to label different sets of test fringe images. We can see that points of the same color tend to cluster together, indicating that

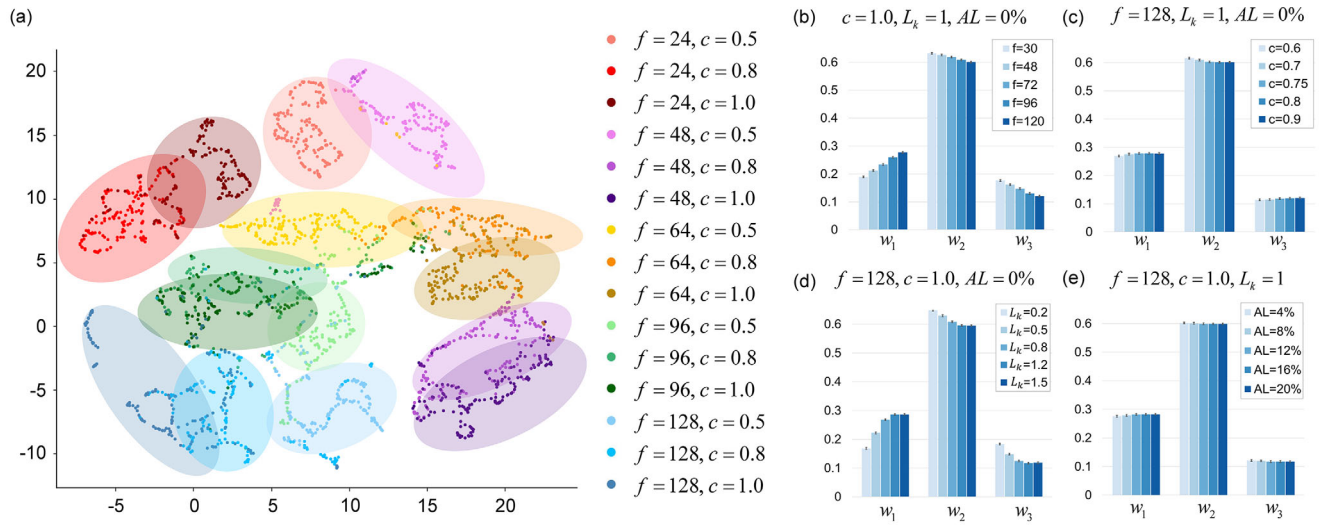

Figure 9. Visualization and analysis of the prediction of GNN for a variety of systems and environments. a) Visualization of the weights predicted by GNN using UMAP. b–e) The numerical distribution of GNN's weights respectively for these cases when the spatial frequency, the contrast, the working distance and the ambient light are changed.

for fringe images collected under the same system and environment, the GNN provides similar fusion strategies. Additionally, we observe that points of different colors are more widely separated, suggesting that when the system or environment changes, the GNN generates distinct fusion strategies.

Furthermore, we studied the mean and variance of the weights of the gating network. Figure 9b–e respectively show the weights predicted by GNN when dealing with fringe images captured with unseen frequencies, contrast, working distances, and ambient light. It can be observed that the variance of prediction is small when GNN handles fringe images obtained under the same condition. This indicates that GNN can output consistent results when facing scenes captured using the same system and environment. Based on the average weights presented in Figure 9b–e, it is evident that the GNN assigns a higher weight to the second expert DNN, while the first and the third expert DNNs receive lower weights. This indicates that the second expert DNN has a predominant influence in the fringe demodulation process.

Figure 9b shows the weights' distribution of GNN when it handled fringe images with unseen frequencies. It can be seen that as the spatial frequency increases, the weight of expert E_1 continues to rise while the weights of expert E_2 and expert E_3 decrease, indicating an increasing importance of expert E_1. As shown in Figure 9c, increasing the contrast of the fringes results in a decrease in the weight assigned to expert E_2, while the weights for experts E_1 and E_3 increase. This suggests that experts E_1 and E_3 become more influential in processing high-contrast fringe images. Figure 9d demonstrates that as the working distance increases, the weight for expert E_1 rises whereas the weights of experts E_2 and E_3 decrease, indicates a growing significance of expert E_1 with greater working distance. Finally, Figure 9e shows that changes in ambient light have minimal impact on the weight distribution from GNN, with expert E_2 maintaining the primary role, followed by expert E_1, and expert E_3 contributing the least. In short, the experimental results indicate that for fringe images captured in different systems and environments, our GNN can assign varying weights to different experts, suggesting that the GNN has recognized the differences among these fringe images. Conversely, for fringe images captured in the same system and environment, the weights output by our GNN are more consistent, indicating that it has detected the similarities in these fringes and established a stable expert fusion strategy. Moreover, we observe that fringe patterns possess intrinsic features that reveal underlying frequency and contrast information. Our CDL framework can extract these features to adaptively predict the weights for orchestrating the collaboration among these expert DNNs.

4. Conclusion

In this work, we have proposed an adaptive cross-domain learning framework designed for generalized structured-light 3D imaging across various systems and environments. Our framework dynamically synthesizes DNNs by integrating a MoE architecture with a GNN. The MoE architecture efficiently extracts features from fringe images across different domains, while the GNN adaptively orchestrates the synthesis of these expert DNNs. Our experimental results demonstrate superior generalization performance across diverse systems and environments. It's worth noting that our method's reliance on multiple experts demands more computational resources compared to approaches using a single DNN. Therefore, for resource-constrained scenarios, such as mobile applications, it may be necessary to streamline the proposed framework. Additionally, the effectiveness of cross-domain learning requires a richly diverse dataset for training. In future, we will explore the use of digital twin technology to rapidly generate diverse and realistic training data virtually. Our approach not only enhances the generalization capabilities of deep-learning-based structured-light 3D imaging but also lays the foundation for the development of robust and versatile AI-driven optical metrology techniques. Furthermore, our framework can be extended to other applications, such as

digital holography, interferometry, and fluorescence microscopy. We believe that our cross-domain learning framework represents a significant advancement in the development of generic and highly adaptive AI-based optical metrology techniques.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

X.L. and S.F. contributed equally to this work. This work was supported by National Key Research and Development Program of China (2022YFB2804603), National Natural Science Foundation of China (62075096, 62005121, U21B2033), Leading Technology of Jiangsu Basic Research Plan (BK20192003), "333 Engineering" Research Project of Jiangsu Province (BRA2016407), Fundamental Research Funds for the Central Universities (30920032101, 2023102001, 2024202002).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

3D imaging, deep learning, fringe analysis, optical metrology, phase measurement

Received: September 28, 2024
Revised: December 9, 2024
Published online:

[1] W. Osten, in *Speckle 2018: VII International Conference on Speckle Metrology*, vol. *10834*, SPIE, **2018**, pp. 1083402.
[2] J. Geng, *Adv. Opt. Photonics* **2011**, *3*, 128.
[3] K. J. Gåsvik, *Optical metrology*, John Wiley & Sons, **2003**.
[4] K. Harding, *Handbook of optical dimensional metrology*, CRC Press, Boca Raton, Florida, USA **2013**.
[5] P. J. De Groot, *Rep. Prog. Phys.* **2019**, *82*, 056101.
[6] M. Takeda, K. Mutoh, *Appl. Opt.* **1983**, *22*, 3977.
[7] Q. Kemao, *Opt. Lasers Eng.* **2007**, *45*, 304.
[8] S. Zhang, *Opt. Lasers Eng.* **2018**, *106*, 119.
[9] S. Van der Jeught, J. J. Dirckx, *Opt. Lasers Eng.* **2016**, *87*, 18.
[10] J. Salvi, J. Pages, J. Batlle, *Pattern Recognit.* **2004**, *37*, 827.
[11] C. Zuo, S. Feng, L. Huang, T. Tao, W. Yin, Q. Chen, *Opt. Lasers Eng.* **2018**, *109*, 23.
[12] L. Lu, V. Suresh, Y. Zheng, Y. Wang, J. Xi, B. Li, *Opt. Lasers Eng.* **2021**, *141*, 106573.
[13] W. Guo, Z. Wu, Q. Zhang, Y. Wang, *Opt. Express* **2021**, *29*, 23822.
[14] Z. Liu, P. C. Zibley, S. Zhang, *Opt. Express* **2018**, *26*, 12632.
[15] Y. LeCun, Y. Bengio, G. Hinton, *nature* **2015**, *521*, 436.
[16] G. Barbastathis, A. Ozcan, G. Situ, *Optica* **2019**, *6*, 921.
[17] D. W. Otter, J. R. Medina, J. K. Kalita, *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 604.
[18] S. Azizi, L. Culp, J. Freyberg, B. Mustafa, S. Baur, S. Kornblith, T. Chen, N. Tomasev, J. Mitrović, P. Strachan, et al., *Nat. Biomed. Eng.* **2023**, *7*, 756.
[19] C. Yang, K. Yuan, Q. Zhu, W. Yu, Z. Li, *Sci. Rob.* **2020**, *5*, eabb2174.
[20] C. Zuo, J. Qian, S. Feng, W. Yin, Y. Li, P. Fan, J. Han, K. Qian, Q. Chen, *Light: Sci. Appl.* **2022**, *11*, 1.
[21] S. Feng, Q. Chen, G. Gu, T. Tao, L. Zhang, Y. Hu, W. Yin, C. Zuo, *Adv. Photonics* **2019**, *1*, 025001.
[22] S. Feng, C. Zuo, L. Zhang, W. Yin, Q. Chen, *Photonics Res.* **2021**, *9*, 1084.
[23] K. Yan, Y. Yu, C. Huang, L. Sui, K. Qian, A. Asundi, *Opt. Commun.* **2019**, *437*, 148.
[24] K. Wang, Q. Kemao, J. Di, J. Zhao, *Adv. Photonics Nexus* **2022**, *1*, 014001.
[25] X. Guo, Y. Li, J. Qian, Y. Che, C. Zuo, Q. Chen, E. Y. Lam, H. Wang, S. Feng, *Opt. Express* **2023**, *31*, 16659.
[26] Y. Li, J. Qian, S. Feng, Q. Chen, C. Zuo, *Opto-Electronic Advances* **2022**, *5*, 210021.
[27] J. Qian, S. Feng, T. Tao, Y. Hu, Y. Li, Q. Chen, C. Zuo, *Apl Photonics* **2020**, *5*, 4.
[28] J. Qian, S. Feng, Y. Li, T. Tao, J. Han, Q. Chen, C. Zuo, *Optics Letters* **2020**, *45*, 1842.
[29] S. Feng, Y. Xiao, W. Yin, Y. Hu, Y. Li, C. Zuo, Q. Chen, *Adv. Photonics Nexus* **2023**, *2*, 036010.
[30] J. M. Blain, *The complete guide to Blender graphics: computer modeling & animation*, AK Peters/CRC Press, Boca Raton, Florida, USA **2019**.
[31] Y. Zheng, S. Wang, Q. Li, B. Li, *Opt. Express* **2020**, *28*, 36568.
[32] F. Wang, C. Wang, Q. Guan, *Opt. Express* **2021**, *29*, 8024.
[33] Y. Zhang, M. Ouyang, A. Ray, T. Liu, J. Kong, B. Bai, D. Kim, A. Guziak, Y. Luo, A. Feizi, et al., *Light: Sci. Appl.* **2019**, *8*, 91.
[34] F. Wang, H. Wang, H. Wang, G. Li, G. Situ, *Opt. Express* **2019**, *27*, 25560.
[35] Y. Xue, S. Cheng, Y. Li, L. Tian, *Optica* **2019**, *6*, 618.
[36] S. Feng, C. Zuo, Y. Hu, Y. Li, Q. Chen, *Optica* **2021**, *8*, 1507.
[37] W. Tahir, H. Wang, L. Tian, *Light: Sci. Appl.* **2022**, *11*, 42.
[38] G. S. Spagnolo, G. Guattari, C. Sapia, D. Ambrosini, D. Paoletti, G. Accardo, *Opt. Lasers Eng.* **2000**, *33*, 141.
[39] Z. Zhang, S. Huang, S. Meng, F. Gao, X. Jiang, *Opt. Express* **2013**, *21*, 12218.
[40] X. Xiao, S. Lian, Z. Luo, S. Li, in *2018 9th international conference on information technology in medicine and education (ITME)*. IEEE, **2018**, pp. 327–331.
[41] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, in *Proceedings of the 25th international conference on Machine learning*, **2008**, pp. 1096–1103.
[42] K. Simonyan, A. Zisserman, *arXiv preprint arXiv:1409.1556* **2014**.
[43] S. Han, H. Mao, W. J. Dally, *arXiv preprint arXiv:1510.00149* **2015**.
[44] G. Hinton, O. Vinyals, J. Dean, *arXiv preprint arXiv:1503.02531* **2015**.
[45] L. McInnes, J. Healy, J. Melville, *arXiv preprint arXiv:1802.03426* **2018**.