RESEARCH ARTICLE | APRIL 07 2025

Multimodal adaptive temporal phase unwrapping using deep learning and physical priors

Yiheng Liu ; Xinsheng Li ; Ziheng Jin ; Wenwu Chen ; Edmund Y. Lam ; Shijie Feng ; Qian Chen ; Chao Zuo

APL Photonics 10, 046104 (2025)
https://doi.org/10.1063/5.0252363

Multimodal adaptive temporal phase unwrapping using deep learning and physical priors

Cite as: APL Photon. 10, 046104 (2025); doi: 10.1063/5.0252363
Submitted: 9 December 2024 • Accepted: 20 March 2025 •
Published Online: 7 April 2025

Yiheng Liu,[1,2] Xinsheng Li,[1,2] Ziheng Jin,[1,2] Wenwu Chen,[1,2] Edmund Y. Lam,[3] Shijie Feng,[1,2,a)] Qian Chen,[1,2,b)] and Chao Zuo[1,2,c)]

AFFILIATIONS

[1] Smart Computational Imaging Laboratory (SCILab), Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
[2] Jiangsu Key Laboratory of Visual Sensing & Intelligent Perception, Nanjing, Jiangsu Province 210094, China
[3] Department of Electrical and Electronic Engineering, University of Hong Kong, Pokfulam Road, Hong Kong, China

[a)] shijiefeng@njust.edu.cn
[b)] chenqian@njust.edu.cn
[c)] **Author to whom correspondence should be addressed:** zuochao@njust.edu.cn

ABSTRACT

In optical 3D measurement, temporal phase unwrapping (TPU) is widely employed in fringe projection and interferometry, as it is crucial for resolving wrapped phase ambiguities and obtaining absolute phase distributions. Recently, deep learning has significantly enhanced TPU performance, particularly in noise robustness. However, existing deep learning-based TPU methods often struggle with generalization, as they typically assume that training and testing data share the same distribution, such as maintaining a constant spatial frequency for fringes in both training and testing processes. When fringe patterns become sparser or denser, the phase unwrapping accuracy declines sharply. Moreover, conventional learning-based methods develop deep neural networks (DNNs) that operate in a single modality, meaning that once the training is complete, the DNN can only perform a specific TPU algorithm. If other TPU methods need to be characterized, the DNN must be retrained, which is a time-consuming process. To address these challenges, we propose for the first time a deep learning-based multimodal adaptive TPU method that integrates prior information obtained by mathematical models of TPU. This approach allows a trained DNN to effectively perform multi-frequency TPU, multi-wavelength TPU, and number-theoretic TPU at the same time while adaptively processing unseen fringes from diverse systems. Experimental results demonstrate that while a U-Net-based TPU method nearly fails with varying test fringes, our method maintains a high accuracy of ~96%. This work offers a novel perspective for developing robust, generalizable AI-driven optical metrology techniques.

I. INTRODUCTION

3D measurement technology plays a critical role in various fields, including reverse engineering, quality inspection, and medical imaging. Traditional contact-based methods, such as coordinate measuring machines, provide high accuracy but are hindered by slow measurement speeds, challenges in capturing complex shapes, and the risk of damaging object surfaces. In contrast, optical 3D measurement has gained widespread application due to its non-contact, fast, and flexible characteristics.[1–5] In various optical 3D measurement technologies, fringe projection profilometry (FPP) and interferometry are two widely used techniques for precise 3D reconstruction.[6–10] Both techniques are based on phase measurement principles, which convert phase information into depth data, enabling the reconstruction of the object's three-dimensional structure.

In phase-based 3D reconstruction methods, the accuracy of phase measurement determines the accuracy of the 3D reconstruction.[11–13] In these methods, the arctangent function is typically used to calculate the object's phase information. However, due to the periodic nature of the arctangent, the measured phase contains 2π discontinuities, which must be removed for

obtaining the absolute phase without ambiguity. Current phase unwrapping methods can be divided into two types: spatial phase unwrapping[14] (SPU) and temporal phase unwrapping[15] (TPU). Representative SPU methods include Goldstein's method,[16] the quality-guided method,[17] Flynn's method,[18] and the minimum Lp-norm method.[19] While SPU assumes phase continuity between neighboring pixels, making it unsuitable for discontinuous surfaces, TPU methods, such as multi-frequency approaches,[20] binary Gray code,[21,22] spatial binary coding,[23–25] and phase encoding,[26] offer an alternative by uniquely labeling phase values through auxiliary patterns. In this work, we focus on multi-frequency (MF, hierarchical) approaches[27–31] and their variants, such as multi-wavelength (MW, heterodyne) approaches[32–37] and number-theoretic (NT) approaches,[38–42] as they can leverage a small number of auxiliary phase maps with global information to resolve the ambiguities pixel by pixel with high accuracy and adaptability, even in the presence of large phase jumps or noise. Zuo *et al.*[15] compared these multi-frequency methods, showing that their stability decreases as the spatial frequency increases due to the system noise.

In recent years, deep learning has been widely applied in 3D measurement. In the field of fringe analysis, Feng *et al.*[43–45] used deep neural networks (DNNs) to predict the wrapped phase from a single fringe pattern, achieving accuracy comparable to that of the 12-step phase-shifting method. Moreover, deep learning has also been applied to SPU. Zhang *et al.*[46] proposed a deep convolutional neural network (DCNN) method. This method utilizes the DeepLabV3+ architecture, which offers noise suppression and feature representation capabilities, achieving better results compared to traditional phase unwrapping algorithms. Wang *et al.*[47] proposed a one-step deep learning approach to solve the phase unwrapping problem. By training a DNN, the authors unwrapped the phase of live mouse osteoblasts and dynamic candle flames, demonstrating the method's performance in noise resistance and anti-aliasing. Spoorthi *et al.*[48] transformed the phase unwrapping problem into a classification task. They utilized a fully convolutional network to predict the fringe order of each pixel in the wrapped phase map, achieving high noise robustness. However, deep learning-based SPU methods still encounter difficulties when it comes to absolute phase unwrapping on discontinuous surfaces. Qian *et al.*[49] proposed a deep learning-based stereo phase unwrapping algorithm. By utilizing DNNs and geometric constraints, the method can recover high-quality wrapped phase from a pair of fringe images, enabling absolute phase unwrapping for discontinuous surfaces. For deep learning-based TPU, Yin *et al.*[50] introduced deep learning into the MF TPU by training a multi-path DNN to establish a mapping between wrapped phase maps and the corresponding unwrapped phase. Recently, Guo *et al.*[51] have proposed a FOA-Net that is based on multi-scale residual modules to achieve a unifying TPU framework. This model can be applied to MF TPU, MW TPU, and NT TPU. It suppresses system noise and improves the phase unwrapping accuracy for dense gratings.

However, existing deep learning-based TPU methods are not without problems. First, they face generalization issues. All of these methods are based on the fundamental assumption that the training and testing data are independent and identically distributed (i.i.d). This means that both datasets are expected to be collected by using the fringe images of the same spatial frequency. However, when there are differences in the spatial frequency, the distribution of the collected data can change significantly even if the measured objects remain the same. Second, existing deep learning methods are typically designed for specific TPU methods, and a trained model can only implement one type of TPU at a time. If other TPU methods are characterized, the model should be retrained, leading to high training costs and low efficiency. To address these challenges, this paper introduces a multimodal adaptive TPU method based on deep learning. The proposed model incorporates physical priors, enabling it to adapt to different imaging systems and fringe frequencies. Furthermore, this model requires only a single training session to simultaneously implement multi-frequency, multi-wavelength, and number-theoretic modalities, eliminating the need for retraining for different modes. Experimental results demonstrate the effectiveness of the proposed method on images collected with various spatial frequencies and imaging systems. Compared to traditional deep learning approaches, our method exhibits enhanced generalization ability, maintaining an unwrapping accuracy of ~96% in scenarios where conventional U-Net fails. We believe that our work paves the way for developing versatile and highly adaptable AI-driven optical metrology techniques.

II. METHOD

A. Phase calculation

As shown in Fig. 1, a typical FPP system consists of a projector and a camera. The projector projects pre-designed sinusoidal fringes onto the object, and as the surface height of the object varies, the fringes undergo deformation. Using phase-shifting algorithms, the phase information of the sinusoidal fringes can be extracted and converted into depth information. The N-step phase-shifting algorithm offers high precision, noise insensitivity, and pixel-wise reconstruction, making it widely used for phase recovery.[11]

The light intensity $I_n(x,y)$ of the N-step phase-shifted fringe pattern can be expressed as

$$I_n(x,y) = A(x,y) + B(x,y)\cos\left[\phi(x,y) - \frac{2n\pi}{N}\right], \tag{1}$$

where $A(x,y)$ is the ambient light intensity, $B(x,y)$ denotes the modulation, n $(n = 1,2,\ldots,N)$ refers to the index of phase-shifting, and $\phi(x,y)$ is the phase of the object being measured. Using the least-squares method, the wrapped phase $\phi(x,y)$ can be calculated as

$$\phi(x,y) = \tan^{-1}\frac{\sum_{n=1}^{N} I_n(x,y)\sin\left(\frac{2n\pi}{N}\right)}{\sum_{n=1}^{N} I_n(x,y)\cos\left(\frac{2n\pi}{N}\right)}. \tag{2}$$

B. Temporal phase unwrapping

From Eq. (2), the range of the arctangent function is $(-\pi,\pi]$, which means that each pixel of the wrapped phase has an ambiguity of $2k\pi$. Therefore, phase unwrapping algorithms are required to eliminate the ambiguity in the wrapped phase. The principle of phase unwrapping can be represented as

$$\Phi(x,y) = \phi(x,y) + 2\pi k(x,y), \tag{3}$$

where $\phi(x,y)$ represents the absolute phase and $k(x,y)$ is the fringe order $(k \in \mathbb{Z})$. According to Eq. (3), $\Phi(x,y)$ can be obtained by

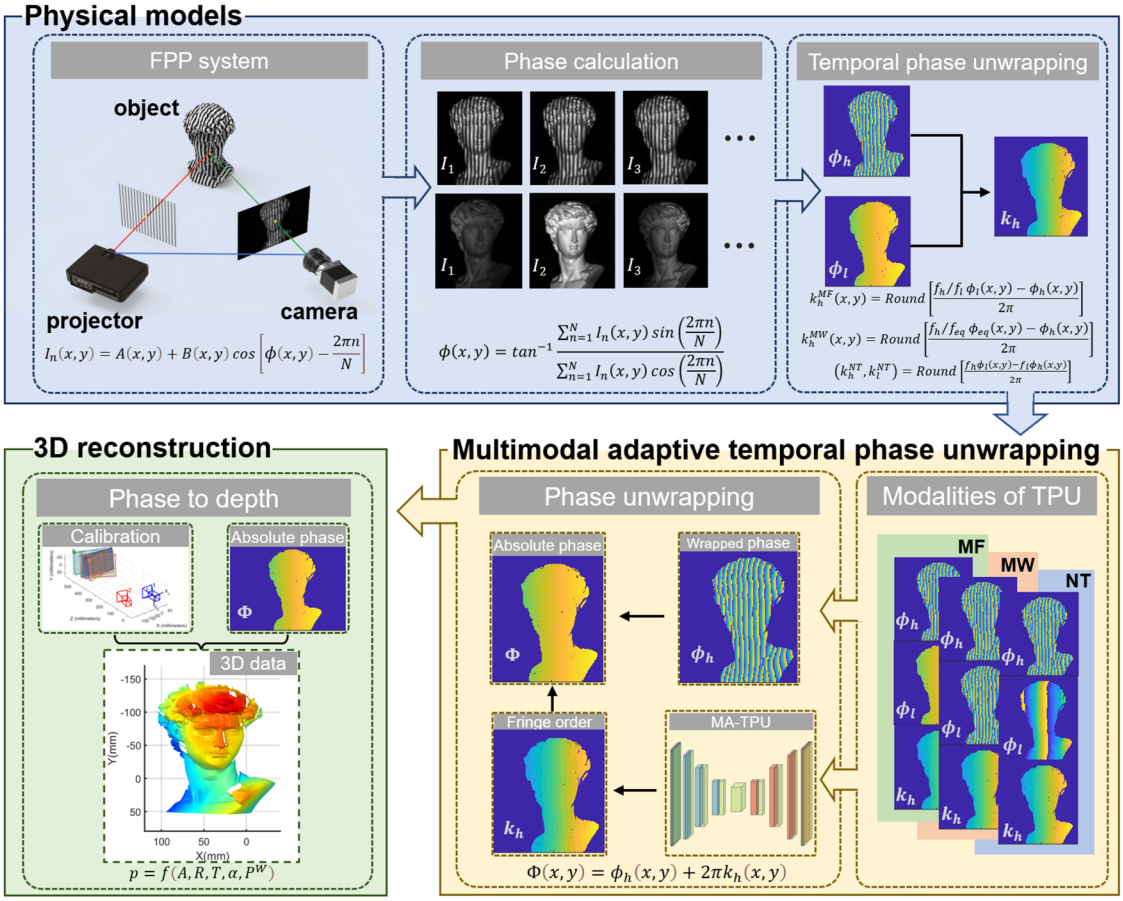

FIG. 1. Diagram of fringe projection profilometry based on the proposed MA-TPU.

$\phi(x,y)$ and $k(x,y)$. Therefore, the crucial aspect of phase unwrapping is to precisely determine the $k(x,y)$ value for each pixel in the wrapped phase image. The commonly used TPU methods include the MF approach, MW approach, and NT approach. These three methods all rely on one or more sets of auxiliary fringe patterns with lower frequencies to assist in determining the fringe order.[15] Here, we take the dual-frequency phase unwrapping as an example to introduce the principles of these three methods. We assume that there are two sets of grating images with different frequencies, f_h and f_l, where h represents high-frequency and l represents low-frequency gratings. The wrapped phases corresponding to these two sets of gratings are $\phi_h(x,y)$ and $\phi_l(x,y)$. TPU aims to use the additional information provided by $\phi_l(x,y)$ to obtain the unambiguous absolute phase $\Phi_h(x,y)$. For these three phase unwrapping methods, they all share the following relationship:

$$\Phi_h(x,y) = \frac{f_h}{f_l}\Phi_l(x,y). \tag{4}$$

This formula describes the proportional relationship between the phase generated by the high-frequency grating and the low-frequency grating. It shows that at the same spatial position, the absolute phase value of the high-frequency grating is $\frac{f_h}{f_l}$ times that of the low-frequency grating. Based on this relationship, we further introduce the TPU methods.

1. Multi-frequency temporal phase unwrapping

The MF method employs the wrapped phase from a lower frequency as an auxiliary phase and then determines the fringe order for each pixel of the wrapped phase based on the relationship between the auxiliary phase and the high-frequency wrapped phase.[27] From Eq. (3), for these two wrapped phases, we have

$$\begin{cases} \Phi_h(x,y) = \phi_h(x,y) + 2\pi k_h(x,y), \\ \Phi_l(x,y) = \phi_l(x,y) + 2\pi k_l(x,y), \end{cases} \tag{5}$$

where $k_h(x,y)$ and $k_l(x,y)$ represent the fringe orders of the high and low frequencies. In the dual-frequency MF method, $\Phi_l(x,y)$ typically uses a fringe pattern with only one sinusoidal period. Hence, $k_l(x,y)$ is zero, which means $\Phi_l(x,y) = \phi_l(x,y)$. Therefore,

according to Eqs. (4) and (5), the fringe order can be calculated by

$$k_h(x,y) = Round\left[\frac{\frac{f_h}{f_l}\Phi_l(x,y) - \Phi_h(x,y)}{2\pi}\right], \tag{6}$$

where $Round(*)$ denotes rounding to the nearest integer.

2. Multi-wavelength temporal phase unwrapping

The idea of the MW method is to use the difference between the reference signal and the measurement signal to produce an unambiguous signal for phase information demodulation.[32] The two wrapped phases with different frequencies are subtracted, and the resulting differential signal is used to unwrap the high-frequency wrapped phase. The resulting phase is referred to as the equivalent phase Φ_{eq}, and its corresponding equivalent frequency is denoted as f_{eq}. This can be expressed as

$$\begin{cases} \Phi_{eq}(x,y) = mod(\phi_h - \phi_l, 2\pi), \\ f_{eq} = f_h - f_l. \end{cases} \tag{7}$$

To eliminate the ambiguity of phase unwrapping, we need to choose an appropriate low-frequency grating such that the equivalent frequency f_{eq} satisfies $f_{eq} \le 1$, ensuring that the range of the auxiliary wrapped phase covers the illumination range of the projector, i.e., $\Phi_{eq}(x,y) = \phi_{eq}(x,y)$. With the help of the equivalent phase, the fringe order $k_h(x,y)$ can be computed by the following formula:

$$k_h(x,y) = Round\left[\frac{\frac{f_h}{f_{eq}}\Phi_{eq}(x,y) - \Phi_h(x,y)}{2\pi}\right]. \tag{8}$$

3. Number-theoretic temporal phase unwrapping

The NT method determines the fringe order pair (k_h, k_l) by using two sets of sinusoidal fringe patterns with wavelengths λ_h and λ_l that are coprime.[41] By combining Eqs. (3) and (4) and relating the spatial frequency to the wavelength, we have

$$\begin{cases} \dfrac{LCM(\lambda_h,\lambda_l)}{\lambda_l}\Phi_h(x,y) = \dfrac{LCM(\lambda_h,\lambda_l)}{\lambda_h}\Phi_l(x,y), \\ f = \dfrac{W}{\lambda}, \end{cases} \tag{9}$$

where $LCM(*)$ denotes the least common multiple of the two wavelengths that are coprime and W represents the projector's resolution in the projection direction (assuming vertical gratings are projected). From Eq. (9), the fringe order pair can be calculated as follows:

$$\frac{f_h\phi_h - f_l\phi_l}{2\pi} = f_l k_h - f_h k_l. \tag{10}$$

Once $LCM(\lambda_h,\lambda_l)$ exceeds the resolution in the projection direction, the phase ambiguity across the entire figure can be eliminated.[52] For a projected pattern with projection direction resolution W, by selecting two different suitable wavelengths λ_h and λ_l to meet the following inequality, phase ambiguity can be eliminated,

$$LCM(\lambda_h,\lambda_l) \ge W. \tag{11}$$

The fringe order pair (k_h, k_l) can be determined using a precomputed lookup table (LUT). By calculating the weighted difference between the wrapped phases at a given position and rounding the result to the nearest integer, the fringe order pair (k_h, k_l) can be identified using the LUT as

$$(k_h, k_l) = LUT\left[Round\left(\frac{f_h\phi_l - f_l\phi_h}{2\pi}\right)\right]. \tag{12}$$

C. Multimodal adaptive temporal phase unwrapping using deep learning

To achieve multimodal adaptive temporal phase unwrapping, we construct the MA-TPU neural network. To improve the generalization ability of the network, physical models are introduced into the network. In particular, the inputs to the network include the high-frequency wrapped phase to be unwrapped, the corresponding auxiliary wrapped phase, and a rough fringe order calculated by the TPU physical models. The introduction of rough fringe order allows the network to adapt to phase unwrapping for unseen fringe patterns. To improve the prediction speed of the network, we adopt the lightweight network BiSeNet[53] as the main structure of the network. BiSeNet is an efficient network that provides fast and accurate pixel-level prediction. Here, we modify the BiSeNet for fringe order prediction. By combining global features and local details, BiSeNet can achieve superior performance while maintaining a low computational cost. As shown in Fig. 2, the MA-TPU network takes the high and low frequency wrapped phases corresponding to three different TPU algorithms along with rough fringe orders as inputs. During forward propagation, the input information first passes through a MultiRes block[54] to extract multi-scale information. To alleviate the loss of spatial information and shrinking receptive field during down-sampling, the network employs a context path to down-sample the image and extract information at different scales while using a spatial path to retain rich image details. In the decoder, global context information is introduced to enhance the phase unwrapping performance, using a feature fusion module (FFM) in place of up-sampling and an attention refinement module (AFR) instead of skip connections to integrate features from the encoder into the decoder for better prediction. Finally, the features are up-sampled in the decoder to restore the original size, yielding high-quality fringe orders. The neural network uses Mean Squared Error (MSE) as the loss function, expressed as

$$L_{\mathrm{MSE}} = \frac{1}{HW}\frac{\sum_{n=1}^{N}\left(k_{pred}^{n} - k_{true}^{n}\right)^2}{N}, \tag{13}$$

where k_{pred}^{n} represents the predicted fringe order of the n-th data in the training set, k_{true}^{n} represents the label fringe order of the n-th data in the training set, N represents the size of the training set, and H and W represent the height and width of the image, respectively. After training, the wrapped phase maps and their corresponding rough fringe orders for any of the MF, MW, or NT methods can be fed into the trained DNN, which will predict the high-quality fringe orders specific to the corresponding TPU method.

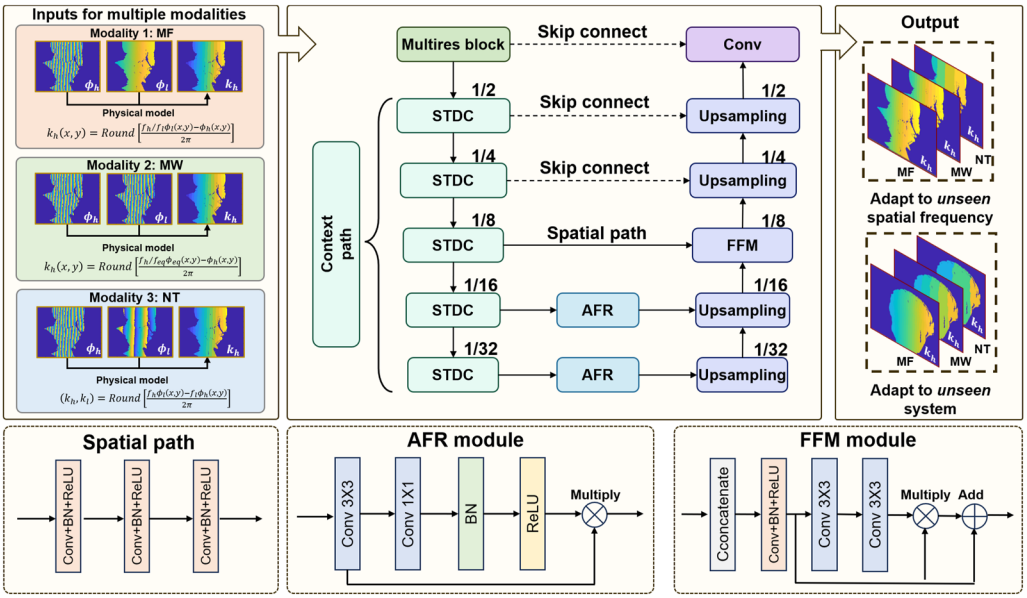

FIG. 2. Network architecture of MA-TPU.

III. EXPERIMENTS

To validate the proposed method, we constructed a structured light illumination system consisting of a projector (DLP 4500, Texas Instruments) and an industrial camera (acA640-750 μm). The focal length of the camera was 12 mm, and the baseline distance between the camera and the projector was ~0.17 m, with the object placed at a distance of around 0.4 m from the system. The angle between the optical axes of the camera and the projector was about 6.21°. To collect training data, the high-frequency fringe images of $f_h = \{16, 48\}$ were used. The corresponding low-frequency fringe images were $f_l = \{1, 1\}$ for MF approach, $f_l = \{15, 47\}$ for MW approach, and $f_l = \{10.9, 10.9\}$ for NT approach. We trained the neural network to unwrap the wrapped phase of these high-frequency fringe patterns. For these three TPU algorithms, we captured 300 sets of dual-frequency three-step phase-shifting fringe patterns from various scenes, and we got a total of 1800 sets of data for training. The test data consisted of 50 sets of data from different scenes. To obtain ground truth data, more auxiliary fringe patterns were used for different TPU algorithms, allowing us to accurately determine the fringe order. Since the errors in the fringe order primarily manifest as jump errors, we assessed performance using the error rate. When calculating the error rate, we only compare the valid points of the ground truth. If the predicted result differs from the ground truth, it is counted as an error point. The error rate is then obtained by taking the ratio of the number of error points to the number of valid points, which can also be written as $E = \frac{P_{error}}{P_{valid}}$, where E represents error rates, P_{error} represents the amount of error points in the predicted result, and P_{valid} represents the amount of valid points in the ground truth.

In the preprocessing of the experimental data, since the range of wrapped phase can be transferred into $(0, 2\pi]$, we normalized the value by dividing it by 2π before feeding it into the neural network, making it easier for the network to learn. Our neural network was implemented in the PyTorch framework, and computations were performed on an RTX 4090 (NVIDIA) GPU. We used the Adam optimizer with a learning rate of 0.001, a batch size of 6, and Mean Squared Error (MSE) as the loss function. The training period was set to 300 epochs. It took ~7 h for U-Net model training, while it took about 3 h for MA-TPU training. To better enable the network to learn the three modalities of TPU, we ensured that each batch contained an equal amount of data from all three modalities. All objects used during testing were not present during the training phase. For comparison, we also trained U-Net models for these TPU algorithms separately, in which the physical priors were not utilized.

A. TPU with fringe images of seen spatial frequencies

To assess the effectiveness of our method, we first evaluated its performance in phase unwrapping for seen frequencies (where only the frequency of the wrapped phase is seen but not the object being measured) under varying noise levels. In our experiment, we set the σ^2 to 2.46. Figure 3 shows the results of traditional physical models, U-Net and MA-TPU. For visual comparison, the unwrapped phases were converted into 3D reconstruction results. The background of the reconstruction is marked in red, with deeper red indicating a higher error rate of phase unwrapping.

As shown, traditional methods performed well at low noise (1σ) and low frequency ($f_h = 16$), but as the frequency increased ($f_h = 48$) and the noise level rose (2σ), the phase unwrapping error rate increased significantly. In contrast, U-Net slightly outperformed traditional methods under 1σ noise, with a relatively lower error rate. However, under higher noise and frequency (especially 2σ and $f_h = 48$), the error rate also increased slightly, although it was still better than traditional methods. Finally, MA-TPU performed the best under these conditions, showing the lowest error rate

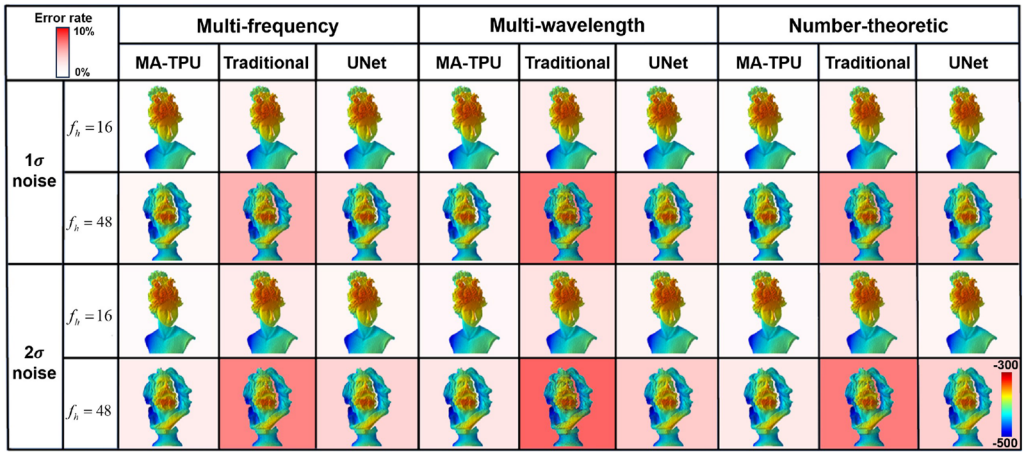

FIG. 3. Matrix of 3D reconstructions that indicate the phase unwrapping error rate of different approaches when fringe patterns of different f_h and noise levels were used. The background color represents the average error rate of phase unwrapping for 50 scenes.

under both low and high noise conditions and exhibiting greater robustness and stability under these conditions.

Table I shows a comparison of the error rates of different methods for the 50 test scenes under seen frequencies (f_h = 16 and f_h = 48) and different noise levels (1σ and 2σ). For traditional methods, the error rates increased significantly with rising noise levels at a given frequency f_h = 16, particularly at 2σ, where the error rate for the MF method increased from 0.46% to 0.76%. Similar trends were observed for the MW and NT methods. The U-Net method demonstrated a significant improvement over traditional approaches, exhibiting much lower error rates at both 1σ and 2σ. In particular, the MF error rate decreased from 0.76% to 0.32%, representing a 57.9% reduction when the noise level was high. In comparison, our method outperformed all others, demonstrating the lowest error rates under both 1σ and 2σ conditions. Notably, for MF, the error rate was only 0.09% at 1σ and 0.19% at 2σ. As the frequency increased to f_h = 48, the error rates for all methods rose significantly compared to those at lower frequencies. The traditional methods showed error rates of 4.75% for MF, 6.06% for MW, and 5.10% for NT at 2σ, indicating substantial errors at high frequencies. Although U-Net exhibited improved performance, the MF error rate still reached 1.68% under high noise conditions. In contrast, our deep learning method demonstrated strong robustness under all conditions, and the maximum error rate does not exceed 0.97% (for the MW method in the presence of high noise).

In addition to comparing phase unwrapping error rates, we also evaluated our method against U-Net in terms of network size, prediction speed, and memory occupancy. Table II demonstrates that, with comparable network parameters, our network's memory occupancy is only 53% of that of U-Net while achieving 3.16 times the

TABLE II. Parameter and speed analysis of the U-Net and MA-TPU on one NVIDIA 4090 card. "M" denotes million and "G" denotes gigabyte.

Method	Params	Memory	FPS
U-Net	17.6M	4.19G	38.71
MA-TPU	17.7M	2.26G	122.35

TABLE I. Error rate of TPU for the traditional method, U-Net, and our method when the spatial frequencies were seen during training. Boldface denotes the best performance among the compared methods.

		MF		MW		NT	
Seen frequency	Method	1σ (%)	2σ (%)	1σ (%)	2σ (%)	1σ (%)	2σ (%)
f_h = 16	Traditional	0.46	0.76	0.68	1.11	0.54	1.04
	U-Net	0.23	0.32	0.28	0.41	0.26	0.38
	MA-TPU	**0.09**	**0.19**	**0.14**	**0.30**	**0.11**	**0.26**
f_h = 48	Traditional	3.01	4.75	5.28	6.06	3.66	5.10
	U-Net	1.23	1.68	1.72	2.01	1.59	1.81
	MA-TPU	**0.35**	**0.68**	**0.62**	**0.97**	**0.40**	**0.81**

prediction speed. Thus, our method can predict fringe order more accurately and at a faster rate.

B. TPU with fringe images of unseen spatial frequencies

In this experiment, we tested the trained neural network with fringe patterns of unseen spatial frequencies $f_h = \{24, 64, 80\}$.

As shown in Fig. 4(a), the differences between the input and output for our proposed MA-TPU and U-Net are illustrated when tested with an unseen frequency $f_h = 80$ and 1σ noise. Our method first computed a rough fringe order using the corresponding TPU algorithm, which is also the result of the traditional method. This rough fringe order and the corresponding wrapped phase maps are then fed into the network, ultimately yielding a high-precision output of fringe order. In contrast, U-Net directly used the wrapped phase maps as input to predict the fringe order, which completely failed when encountering unseen frequencies.

The experimental results are shown in Fig. 4(b). Traditional methods performed well under low-frequency and low-noise conditions ($f_h = 24$ and 1σ); however, as the frequency increased ($f_h = 64$ and $f_h = 80$) and noise levels rose (2σ), error rates significantly increased, particularly for MW and NT. Among the three methods,

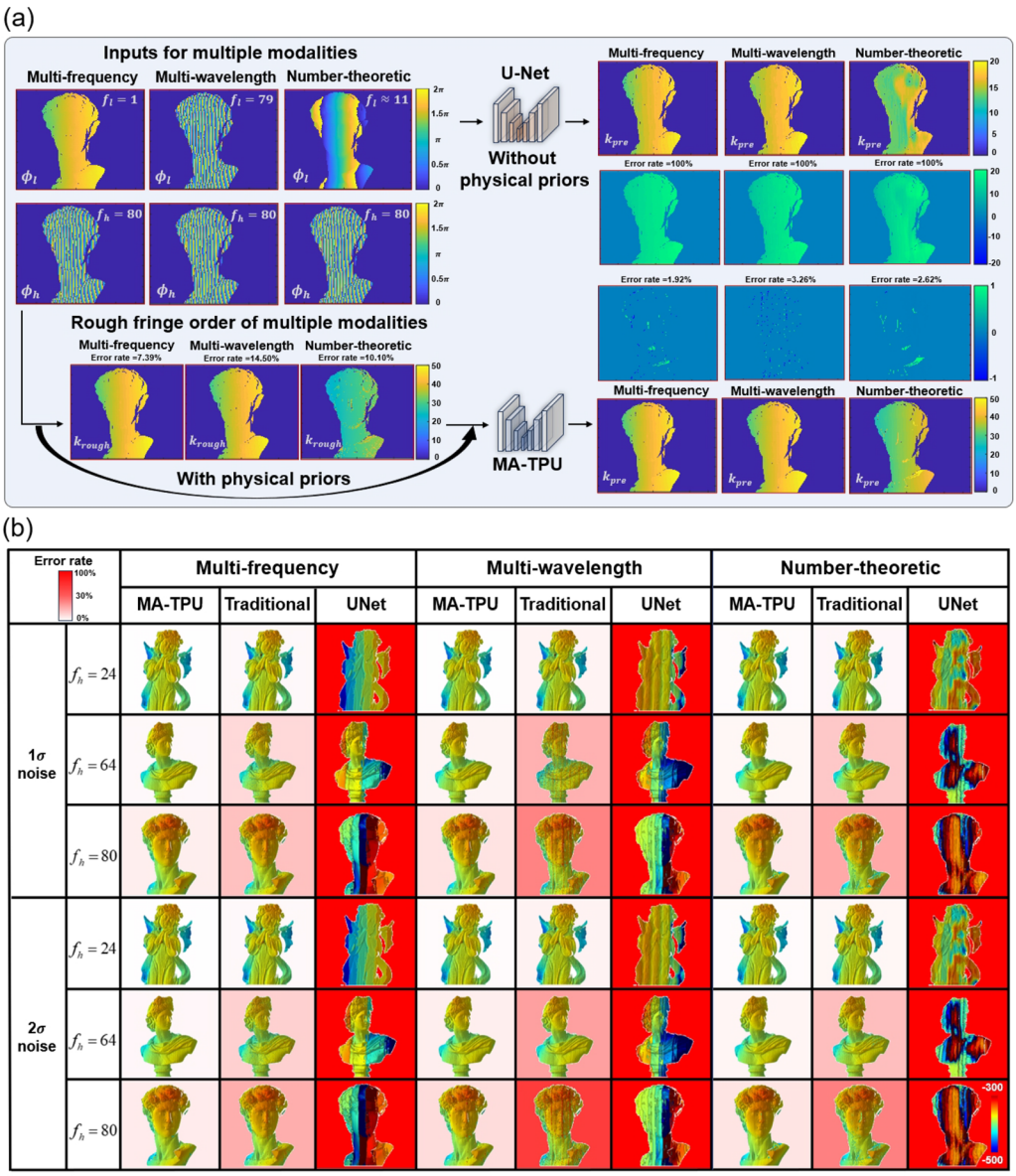

FIG. 4. (a) The diagram shows a comparison of the phase unwrapping performance between MA-TPU and U-Net regarding the input and output of the networks at unseen frequencies. (b) Matrix of 3D reconstructions that indicate the phase unwrapping error rate of different approaches when unseen fringe patterns were used. The background color represents the average error rate of phase unwrapping for 50 scenes.

U-Net exhibited the poorest performance, displaying the highest error rates for MF, MW, and NT, indicating a lack of generalization capability when addressing unseen frequencies. In contrast, our method outperformed both traditional methods and U-Net under all conditions, maintaining low phase unwrapping error rates regardless of noise level or frequency. In particular, for MF and NT, our method exhibited exceptional robustness and stability, demonstrating superior performance in managing complex scenes and unseen frequencies.

Table III shows a comparison of the accuracy of the traditional TPU method and our proposed MA-TPU method across 50 different test scenes. The U-Net model consistently failed with a 100% error rate under all conditions, highlighting its inability to generalize to unseen spatial frequencies. In contrast, our MA-TPU method demonstrated superior generalization across a wide range of fringe frequencies and maintained strong robustness under high-frequency and high-noise conditions, where traditional methods and U-Net struggle. For fringes of $f_h = 24$, the MF method's error was only 0.20% under 1σ noise, increasing to 0.35% under 2σ, representing a relative increase of 75%. As the fringe frequency rose to $f_h = 64$, the noise level significantly impacted the different methods. Under 1σ noise conditions, our method achieved an error rate of 0.78%, which increased to 0.90% under 2σ. In comparison, traditional methods recorded error rates of 4.42% and 5.45% under the same conditions. This discrepancy clearly illustrates that our method excels in high-frequency and high-noise environments, far outperforming traditional methods and demonstrating strong robustness. Moreover, at the highest frequency $f_h = 80$, the error rates of traditional methods reached 22.9% for the MW method and 14.9% for the NT method under 2σ noise conditions, while our method showed relatively low error rates of 4.09% and 3.09% (i.e., accuracy of 95.91% and 96.91%). This difference underscores the generalization capability and stability of our model in handling complex frequency conditions. Even when the U-Net model failed completely, with an error rate of 100%, our method still achieved a low error rate, demonstrating its broad adaptability in feature extraction.

C. TPU with fringe images of unseen fringe projection systems

To further evaluate the adaptability of our method across unseen systems, we constructed another fringe projection system with a camera lens focal length of 8.5 mm and a baseline distance of ~0.6 m between the camera and projector. The test objects were positioned about 1 m away from the system. The angle between the optical axes of the camera and the projector was about 29.65°. During the test, we utilized fringe patterns of unseen frequencies $f_h = \{32, 40, 72\}$, and the experimental results are shown in Fig. 5.

As we can see, traditional methods performed relatively well under low noise (1σ) and low frequency ($f_h = 32$); however, as noise levels increased (2σ) and frequencies rose ($f_h = 40$ and $f_h = 72$), the error rates significantly escalated, particularly for MW and NT. The U-Net exhibited poor performance under all conditions, with consistently high error rates. In contrast, our method demonstrated substantial advantages in all scenarios, showcasing strong robustness and adaptability when confronted with fringe images obtained from a new fringe projection system with unseen spatial frequencies.

We also expanded the test dataset to include 50 sets for evaluation. Table IV shows a comparison of the accuracy of traditional TPU methods with our approach. Notably, the U-Net method failed completely under all frequency and noise conditions, resulting in a 100% error rate. This performance indicates that the U-Net model could not adapt to unseen frequencies or varying system parameters, reflecting its extremely limited generalization capabilities. For traditional methods, the error was relatively low at lower frequencies ($f_h = 32$ and $f_h = 40$). However, as frequency and noise levels increased, the error rose significantly. At $f_h = 72$, particularly for the MW method, the error reached 9.84% under 2σ noise conditions, demonstrating sensitivity to high frequencies and elevated noise levels. Overall, traditional methods performed poorly under high-noise conditions, with errors increasing significantly as noise levels rose. In contrast, our method demonstrated excellent performance under all test conditions. Even with unseen frequencies and different system parameters, the error remained very low. Under single noise conditions, whether for MF, MW, or NT, our method consistently achieved an error rate below 1.5%. Even under 2σ noise conditions, when handling TPU tasks at a frequency $f_h = 72$, our method's error remained within 2.30%. This experiment confirms that our method exhibits strong robustness and generalization ability when dealing with high noise and high frequencies, far outperforming traditional methods and U-Net in robustness when addressing unseen frequencies collected from different systems.

TABLE III. Error rate of TPU for the traditional method and our method when the spatial frequencies were not seen during training. U-Net consistently failed across all tested frequencies, with a consistent 100% error rate. Boldface denotes the best performance among the compared methods.

		MF		MW		NT	
Unseen frequency	Method	1σ (%)	2σ (%)	1σ	2σ (%)	1σ (%)	2σ (%)
$f_h = 24$	Traditional	0.67	0.80	1.31	1.46	0.90	1.03
	MA-TPU	**0.20**	**0.35**	**0.38**	**0.53**	**0.33**	**0.41**
$f_h = 64$	Traditional	4.42	5.45	9.22	11.1	6.39	7.56
	MA-TPU	**0.78**	**0.90**	**1.20**	**1.93**	**1.05**	**1.17**
$f_h = 80$	Traditional	7.39	9.49	14.5	22.9	10.1	14.9
	MA-TPU	**1.92**	**2.29**	**3.26**	**4.09**	**2.62**	**3.09**

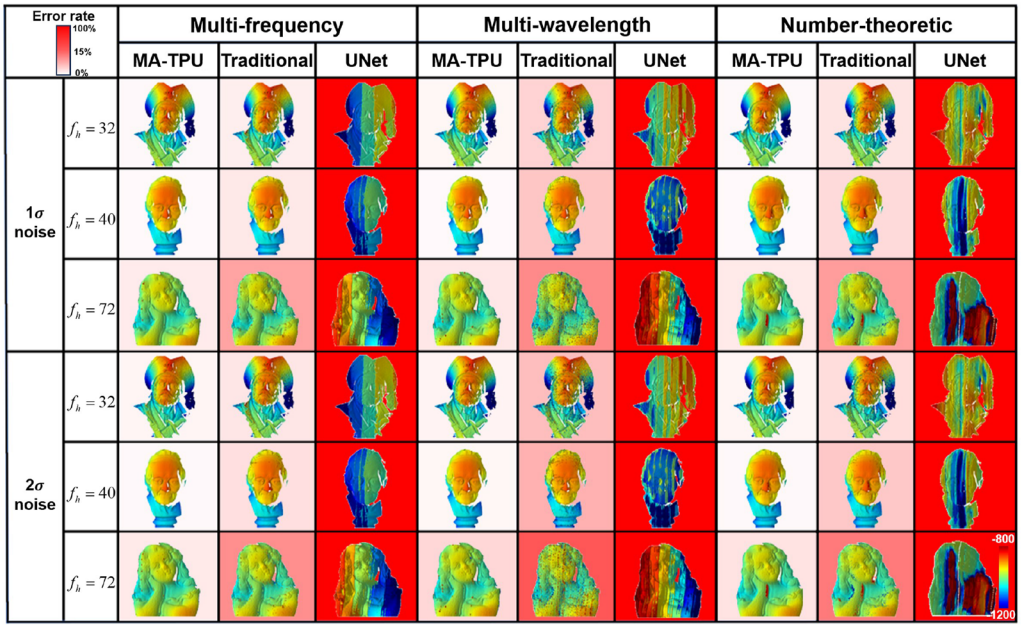

FIG. 5. Matrix of 3D reconstructions that indicate the phase unwrapping error rate of different approaches when fringe patterns were captured with unseen frequencies and from an unseen system. The background color represents the average error rate of phase unwrapping for 50 scenes.

TABLE IV. Error rate of TPU for the traditional method and our method when both the spatial frequencies and the measurement system were not seen during training. U-Net consistently failed across all tested frequencies and unseen systems, with a consistent 100% error rate. Boldface denotes the best performance among the compared methods.

		MF		MW		NT	
Unseen frequency	Method	1σ (%)	2σ (%)	1σ (%)	2σ (%)	1σ (%)	2σ (%)
f_h = 32	Traditional	1.05	1.45	2.53	3.81	1.54	1.99
	MA-TPU	**0.47**	**0.61**	**0.50**	**0.69**	**0.46**	**0.64**
f_h = 40	Traditional	1.31	1.85	3.33	4.53	2.15	3.23
	MA-TPU	**0.30**	**0.45**	**0.39**	**0.49**	**0.39**	**0.46**
f_h = 72	Traditional	5.36	6.62	6.86	9.84	5.77	7.75
	MA-TPU	**1.15**	**1.28**	**1.41**	**2.30**	**1.24%**	**1.68**

D. Dynamic experiment with fringe images of unseen frequencies and systems

To evaluate the adaptability of our method in dynamic scenes, another FPP system with a measurement speed of 100 Hz was employed to project sinusoidal fringe patterns with a high frequency f_h = 64 onto three rotating isolated objects with complex surfaces, which were not present in the training and validation datasets. Three traditional TPU algorithms and our MA-TPU network (only trained on a static scene dataset) were applied to perform phase unwrapping on the dynamic data. The test results are shown in Fig. 6.

As depicted, at low rotational speeds (two rotations per minute), traditional methods exhibit significant performance degradation in reconstruction quality compared to static scenarios, primarily due to the combined effects of motion artifacts and noise interference. In contrast, our approach demonstrates effective error mitigation under low rotational speed conditions, achieving superior reconstruction quality and demonstrating robust noise suppression capabilities. However, since our method requires capturing six phase-shifted fringe patterns to obtain a single 3D reconstruction result, severe motion artifacts can occur in high rotational speed scenes (40 rotations per minute). As demonstrated, both traditional methods and our approach struggle to handle motion artifacts in such high rotational speed scenarios, leading to numerous errors at the image edges. Under these conditions, both methods struggle to maintain high-quality reconstruction due to the rapid changes in the scene.

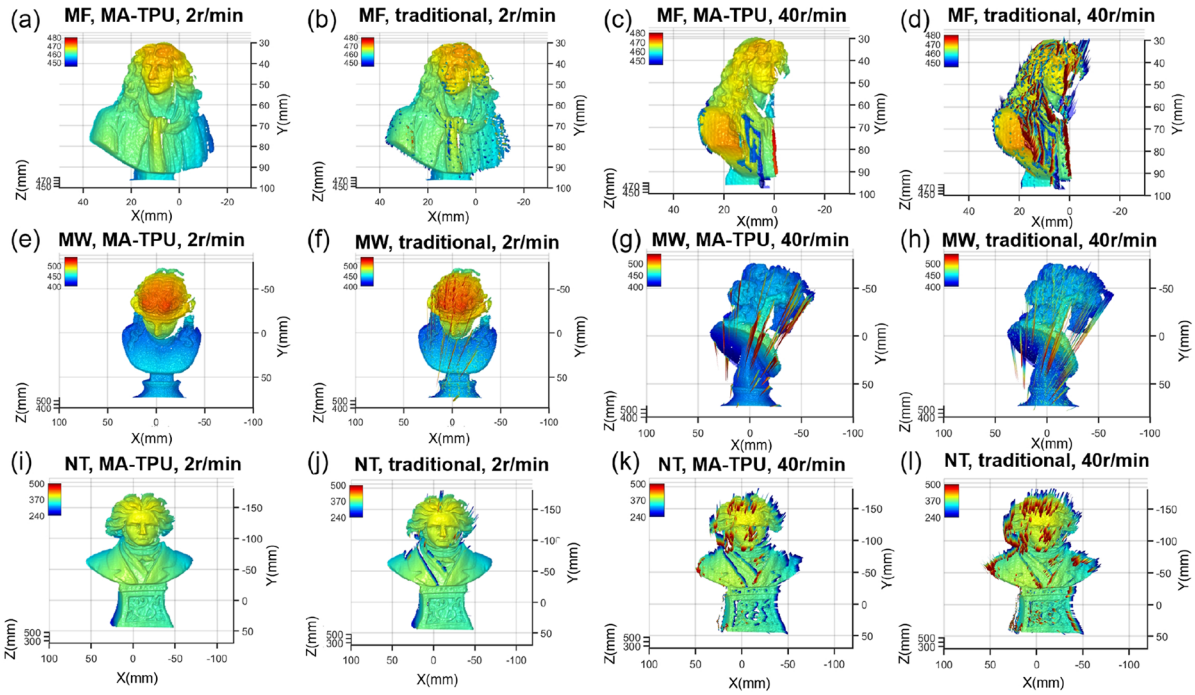

FIG. 6. Measurement results of three dynamic objects under varying conditions. (a)–(d) Multi-frequency TPU results, (e)–(h) multi-wavelength TPU results, and (i)–(l) number-theoretical TPU results at different rotation speeds.(Multimedia available online).

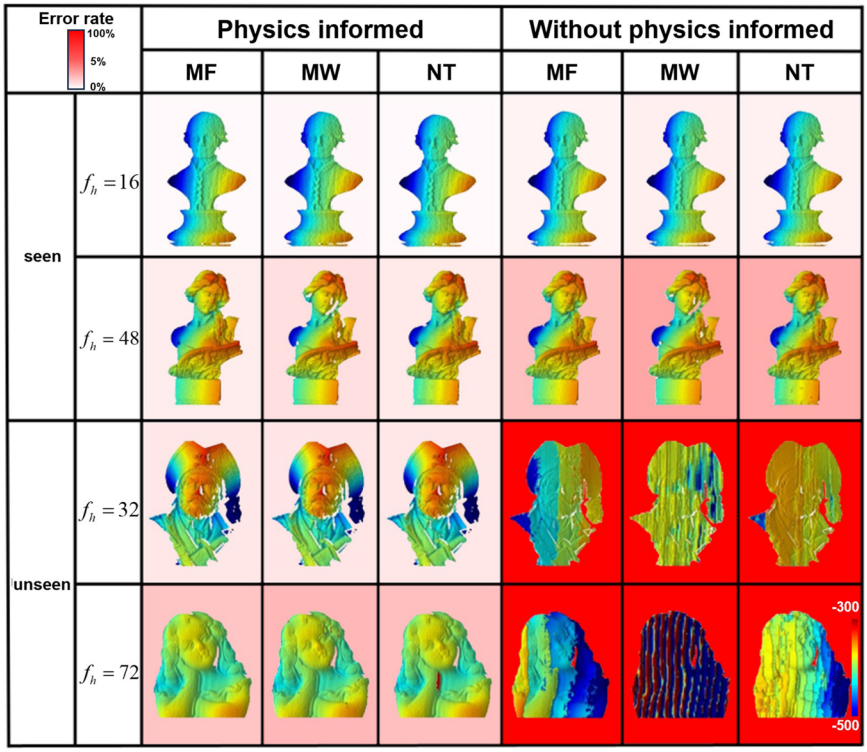

FIG. 7. Matrix of 3D reconstructions that indicate the necessity of physical priors on the phase unwrapping error rate. The background color represents the average error rate of phase unwrapping for 50 scenes.

E. Ablation study

To verify the necessity of incorporating physical prior information to enhance the frequency adaptation of our method, we conducted ablation experiments. In these experiments, we trained one network with physical prior information (MA-TPU) and another without it by using fringe patterns with spatial frequencies of $f_h = \{16, 48\}$. The results are shown in Fig. 7. For the previously seen frequencies (i.e., $f_h = \{16, 48\}$), both models demonstrated relatively stable performance with low error rates, indicating that the DNN adapted well to the fringe frequencies encountered during training. However, as the frequency increased to $f_h = 48$, while the overall performance remained good, the model with physical priors exhibited relatively lower error rates. This improvement arises from explicitly incorporating fringe order as a physical prior, providing direct guidance for the network in learning the mapping from wrapped phase to fringe order. The key reason is that the valid range of fringe orders varies significantly across different frequencies and FPP systems. By explicitly including fringe order as an input to the network, the possible range of predicted values becomes clearly constrained, reducing uncertainty and, thus, enhancing the accuracy of phase unwrapping. In contrast, when using only wrapped phase as input, the inherent ambiguity of the wrapped phase forces the network to simultaneously infer the fringe order and suppress noise, increasing the complexity of the learning task and consequently limiting accuracy. Therefore, integrating this physical prior not only reduces inference errors in phase unwrapping but also substantially improves network stability and efficiency, especially under complex or noisy measurement conditions.

For unseen frequencies, we observed a significant disparity between the two methods. Under frequency conditions $f_h = 32$ and $f_h = 72$, the error rate of the model without physical priors rose dramatically, with the background turning deep red, indicating very high errors for TPU tasks and insufficient generalization ability. In contrast, the model incorporating physical prior information maintained strong robustness under unseen frequency conditions, showing significantly lower error rates compared to the model without physical priors. This demonstrates that physical priors facilitate better generalization for the model when confronted with unseen scenarios.

IV. CONCLUSION

In this work, we have presented a novel multimodal adaptive TPU deep neural network model, MA-TPU, which incorporates the TPU physical model as prior knowledge to guide the DNN in learning the relationship between wrapped phases and fringe order. The model effectively processes fringe images at different frequencies and adapts to variations in data distribution caused by different imaging systems. In addition, it unifies the three TPU modalities MF, MW, and NT, enabling the completion of multimodal adaptive phase unwrapping tasks within a single training process, significantly enhancing efficiency. Experimental results demonstrate that the proposed method performs excellently on images collected from various unseen spatial frequencies and systems, achieving unwrapping accuracy nearly 96% even in scenarios where traditional deep learning methods fail. Furthermore, due to its lightweight network design, the proposed method also attains a higher phase unwrapping speed. These results underscore the strong generalization ability and robustness of the proposed method.

Despite these promising outcomes, several limitations and opportunities for improvement remain. For instance, the proposed method remains dependent on real data. In our experiments, the construction of a total of 3000-sample datasets demanded meticulous data collection procedures, a process requiring no fewer than 15 h. Our future research will explore the use of simulated data as an alternative to real data or investigate unsupervised training methods to reduce manpower and time costs. Moreover, although the proposed method has demonstrated strong generalization capability across varying frequencies and projection systems in TPU tasks, it currently requires capturing six phase-shifted fringe images to obtain absolute phase information. As demonstrated in our dynamic experiments, this multi-frame requirement can lead to motion artifacts, reducing its effectiveness in high-speed dynamic scenarios. Recently, Li *et al.*[55] have proposed a cross-domain learning (CDL) approach to enhance the generalization capability of single-frame fringe analysis, demonstrating potential to improve the efficiency of wrapped phase calculation in TPU and consequently reduce its sensitivity to motion. Future research could explore integrating CDL-based strategies with the MA-TPU framework to mitigate motion artifacts, potentially enabling more efficient and robust phase unwrapping with enhanced generalization and accuracy in dynamic environments. We believe that the proposed method offers a new direction for developing versatile and highly adaptive AI-driven optical metrology technologies, with the potential to expand the application of optical 3D measurement technologies across a broader range of fields.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (Grant Nos. 2022YFB2804603 and 2022YFB2804605), National Natural Science Foundation of China (Grant Nos. 62075096, 62005121, and U21B2033), Leading Technology of Jiangsu Basic Research Plan (Grant No. BK20192003 and BK20192005), "333 Engineering" Research Project of Jiangsu Province (Grant No. BRA2016407), and Fundamental Research Funds for the Central Universities (Grant Nos. 30921011208, 30920032101, 2023102001, and 2024202002).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Yiheng Liu: Conceptualization (equal); Data curation (lead); Methodology (equal); Software (equal); Validation (lead); Writing – original draft (lead). **Xinsheng Li**: Conceptualization (supporting). **Ziheng Jin**: Software (supporting). **Wenwu Chen**: Visualization (supporting). **Edmund Y. Lam**: Writing – review & editing (supporting). **Shijie Feng**: Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing – review & editing (equal). **Qian Chen**:

Funding acquisition (equal); Resources (equal). **Chao Zuo**: Funding acquisition (equal); Resources (equal).

DATA AVAILABILITY

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

REFERENCES

[1] R. Leach, *Optical Measurement of Surface Topography* (Springer, 2011), Vol. 8.
[2] S. Zhang, *Handbook of 3D Machine Vision: Optical Metrology and Imaging* (CRC Press, 2013).
[3] Z. Wu, H. Wang, F. Chen, X. Li, Z. Chen, and Q. Zhang, "Dynamic 3D shape reconstruction under complex reflection and transmission conditions using multi-scale parallel single-pixel imaging," Light: Adv. Manuf. **5**, 373 (2024).
[4] Z. Wu, I. Kang, Y. Yao, Y. Jiang, J. Deng, J. Klug, S. Vogt, and G. Barbastathis, "Three-dimensional nanoscale reduced-angle ptycho-tomographic imaging with deep learning (RAPID)," eLight **3**, 7 (2023).
[5] A. Saba, C. Gigli, A. B. Ayoub, and D. Psaltis, "Physics-informed neural networks for diffraction tomography," Adv. Photonics **4**, 066001 (2022).
[6] S. S. Gorthi and P. Rastogi, "Fringe projection techniques: Whither we are?," Opt. Lasers Eng. **48**, 133–140 (2010).
[7] J. Xu and S. Zhang, "Status, challenges, and future perspectives of fringe projection profilometry," Opt. Lasers Eng. **135**, 106193 (2020).
[8] D. Malacara, *Optical Shop Testing* (Wiely Interscience, 2007).
[9] W. Chen, S. Feng, W. Yin, Y. Li, J. Qian, Q. Chen, and C. Zuo, "Deep-learning-enabled temporally super-resolved multiplexed fringe projection profilometry: High-speed kHz 3D imaging with low-speed camera," PhotoniX **5**, 25 (2024).
[10] B. Wang, W. Chen, J. Qian, S. Feng, Q. Chen, and C. Zuo, "Single-shot super-resolved fringe projection profilometry (SSSR-FPP): 100,000 frames-per-second 3D imaging with deep learning," Light: Sci. Appl. **14**, 70 (2025).
[11] C. Zuo, S. Feng, L. Huang, T. Tao, W. Yin, and Q. Chen, "Phase shifting algorithms for fringe projection profilometry: A review," Opt. Lasers Eng. **109**, 23–59 (2018).
[12] M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," Appl. Opt. **22**, 3977–3982 (1983).
[13] Q. Kemao, "Windowed Fourier transform for fringe pattern analysis," Appl. Opt. **43**, 2695–2702 (2004).
[14] X. Su and W. Chen, "Reliability-guided phase unwrapping algorithm: A review," Opt. Lasers Eng. **42**, 245–261 (2004).
[15] C. Zuo, L. Huang, M. Zhang, Q. Chen, and A. Asundi, "Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review," Opt. Lasers Eng. **85**, 84–103 (2016).
[16] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," Radio Sci. **23**, 713–720 (1988).
[17] H. Lim, W. Xu, and X. Huang, "Two new practical methods for phase unwrapping," in *1995 International Geoscience and Remote Sensing Symposium, IGARSS'95. Quantitative Remote Sensing for Science and Applications* (IEEE, 1995), Vol. 1, pp. 196–198.
[18] T. J. Flynn, "Two-dimensional phase unwrapping with minimum weighted discontinuity," J. Opt. Soc. Am. A **14**, 2692–2701 (1997).
[19] D. C. Ghiglia and L. A. Romero, "Minimum L^p-norm two-dimensional phase unwrapping," J. Opt. Soc. Am. A **13**, 1999–2013 (1996).
[20] X. He and Q. Kemao, "A comparative study on temporal phase unwrapping methods in high-speed fringe projection profilometry," Opt. Lasers Eng. **142**, 106613 (2021).
[21] Z. Wu, W. Guo, Y. Li, Y. Liu, and Q. Zhang, "High-speed and high-efficiency three-dimensional shape measurement based on gray-coded light," Photonics Res. **8**, 819–829 (2020).
[22] X. He, D. Zheng, Q. Kemao, and G. Christopoulos, "Quaternary gray-code phase unwrapping for binary fringe projection profilometry," Opt. Lasers Eng. **121**, 358–368 (2019).
[23] Y. Wang, L. Liu, J. Wu, X. Chen, and Y. Wang, "Spatial binary coding method for stripe-wise phase unwrapping," Appl. Opt. **59**, 4279–4285 (2020).
[24] H. Wu, Y. Cao, Y. Dai, and H. Zhang, "Ultra-fast 3D imaging by a big codewords space division multiplexing binary coding," Opt. Lett. **48**, 2793–2796 (2023).
[25] H. Wu, Y. Cao, Y. Dai, and Z. Wei, "Orthogonal spatial binary coding method for high-speed 3d measurement," IEEE Trans. Image Process. **33**, 2703 (2024).
[26] Y. Wang and S. Zhang, "Novel phase-coding method for absolute phase retrieval," Opt. Lett. **37**, 2067–2069 (2012).
[27] J. M. Huntley and H. Saldner, "Temporal phase-unwrapping algorithm for automated interferogram analysis," Appl. Opt. **32**, 3047–3052 (1993).
[28] J. Tian, X. Peng, and X. Zhao, "A generalized temporal phase unwrapping algorithm for three-dimensional profilometry," Opt. Lasers Eng. **46**, 336–342 (2008).
[29] H. Zhao, W. Chen, and Y. Tan, "Phase-unwrapping algorithm for the measurement of three-dimensional object shapes," Appl. Opt. **33**, 4497–4500 (1994).
[30] L. Kinell and M. Sjödahl, "Robustness of reduced temporal phase unwrapping in the measurement of shape," Appl. Opt. **40**, 2297–2303 (2001).
[31] X. Peng, Z. Yang, and H. Niu, "Multi-resolution reconstruction of 3-D image with modified temporal unwrapping algorithm," Opt. Commun. **224**, 35–44 (2003).
[32] K. Creath, "Step height measurement using two-wavelength phase-shifting interferometry," Appl. Opt. **26**, 2810–2816 (1987).
[33] Y.-Y. Cheng and J. C. Wyant, "Two-wavelength phase shifting interferometry," Appl. Opt. **23**, 4539–4543 (1984).
[34] J. C. Wyant, "Testing aspherics using two-wavelength holography," Appl. Opt. **10**, 2113–2118 (1971).
[35] A. J. Alcock and S. A. Ramsden, "Two wavelength interferometry of a laser-induced spark in air," Appl. Phys. Lett. **8**, 187–188 (1966).
[36] C. Polhemus, "Two-wavelength interferometry," Appl. Opt. **12**, 2071–2074 (1973).
[37] R. Dändliker, R. Thalmann, and D. Prongué, "Two-wavelength laser interferometry using superheterodyne detection," Opt. Lett. **13**, 339–341 (1988).
[38] V. I. Gushov and Y. N. Solodkin, "Automatic processing of fringe patterns in integer interferometers," Opt. Lasers Eng. **14**, 311–324 (1991).
[39] J. Burke, T. Bothe, W. Osten, and C. F. Hess, "Reverse engineering by fringe projection," Proc. SPIE **4778**, 312–324 (2002).
[40] M. Takeda, Q. Gu, M. Kinoshita, H. Takai, and Y. Takahashi, "Frequency-multiplex Fourier-transform profilometry: A single-shot three-dimensional shape measurement of objects with large height discontinuities and/or surface isolations," Appl. Opt. **36**, 5347–5354 (1997).
[41] M. Wang and M. Wang, "Phase unwrapping by a lookup table method: Application to phase maps with singular points," Opt. Eng. **38**, 2075–2080 (1999).
[42] C. E. Towers, D. P. Towers, and J. D. C. Jones, "Time efficient Chinese remainder theorem algorithm for full-field fringe phase analysis in multi-wavelength interferometry," Opt. Express **12**, 1136–1143 (2004).
[43] S. Feng, Q. Chen, G. Gu, T. Tao, L. Zhang, Y. Hu, W. Yin, and C. Zuo, "Fringe pattern analysis using deep learning," Adv. Photonics **1**, 025001 (2019).
[44] S. Feng, Y. Xiao, W. Yin, Y. Hu, Y. Li, C. Zuo, and Q. Chen, "Fringe-pattern analysis with ensemble deep learning," Adv. Photonics Nexus **2**, 036010 (2023).
[45] S. Feng, C. Zuo, Y. Hu, Y. Li, and Q. Chen, "Deep-learning-based fringe-pattern analysis with uncertainty estimation," Optica **8**, 1507–1510 (2021).
[46] T. Zhang, S. Jiang, Z. Zhao, K. Dixit, X. Zhou, J. Hou, Y. Zhang, and C. Yan, "Rapid and robust two-dimensional phase unwrapping via deep learning," Opt. Express **27**, 23173–23185 (2019).
[47] K. Wang, Y. Li, Q. Kemao, J. Di, and J. Zhao, "One-step robust deep learning phase unwrapping," Opt. Express **27**, 15100–15115 (2019).
[48] G. E. Spoorthi, R. K. Sai Subrahmanyam Gorthi, and S. Gorthi, "PhaseNet 2.0: Phase unwrapping of noisy data based on deep learning approach," IEEE Trans. Image Process. **29**, 4862–4872 (2020).
[49] J. Qian, S. Feng, T. Tao, Y. Hu, Y. Li, Q. Chen, and C. Zuo, "Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement," APL Photonics **5**, 046105 (2020).
[50] W. Yin, Q. Chen, S. Feng, T. Tao, L. Huang, M. Trusiak, A. Asundi, and C. Zuo, "Temporal phase unwrapping using deep learning," Sci. Rep. **9**, 20175 (2019).

[51]X. Guo, Y. Li, J. Qian, Y. Che, C. Zuo, Q. Chen, E. Y. Lam, H. Wang, and S. Feng, "Unifying temporal phase unwrapping framework using deep learning," Opt. Express **31**, 16659–16675 (2023).
[52]C. Zuo, Q. Chen, G. Gu, S. Feng, F. Feng, R. Li, and G. Shen, "High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width-modulation fringe projection," Opt. Lasers Eng. **51**, 953–960 (2013).
[53]C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Springer, 2018), pp. 325–341.
[54]N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," Neural Networks **121**, 74–87 (2020).
[55]X. Li, S. Feng, W. Chen, Z. Jin, Q. Chen, and C. Zuo, "Adaptive structured-light 3D surface imaging with cross-domain learning," Laser Photonics Rev. 2401609 (published online, 2025).