

<https://doi.org/10.1038/s44310-025-00096-z>

# Digital-twin-driven unambiguous structured light 3D imaging with physics-aware learning

Check for updates

Yiheng Liu<sup>1,2</sup>, Wenwu Chen<sup>1,2</sup>, Jinyang Jiang<sup>1,2</sup>, Shengqi Yu<sup>1,2</sup>, Ziheng Jin<sup>1,2</sup>, Xinsheng Li<sup>1,2</sup>, Shijie Feng<sup>1,2,3</sup>✉, Qian Chen<sup>1,2,3</sup> & Chao Zuo<sup>1,2,3</sup>

Temporal phase unwrapping (TPU) plays a pivotal role in resolving phase ambiguities in fringe projection profilometry (FPP) caused by surface discontinuities or spatially isolated features. Although recent AI-based TPU methods significantly outperform traditional algorithms in processing noisy wrapped phases, they often depend on large-scale manually collected real-world datasets, which are time-consuming and labor-intensive. Moreover, these methods typically assume that training and testing data follow the same distribution, leading to dramatic accuracy degradation when applied to fringe images from unseen measurement systems. To overcome these limitations, we propose a digital-twin-driven, physics-aware framework for unambiguous structured-light 3D imaging. This framework leverages digital twin technology to generate vast amounts of realistic synthetic fringe images for training, while incorporating Fourier-domain consistency constraints and TPU physical models as priors. It establishes a generalized solution that supports multi-frequency (MF), multi-wavelength (MW), and number-theoretic (NT) TPU approaches. Experimental results show that the proposed network demonstrates exceptional generalization capabilities across unseen measurement systems. It achieves over 94% phase unwrapping accuracy for high-frequency fringes where conventional networks fail, performing comparably to models trained on real-world data. This research provides a promising pathway toward low-cost, high-precision, and highly generalizable intelligent optical metrology systems.

Three-dimensional (3D) imaging technologies play a crucial role in fields such as industrial inspection, medical diagnostics, and self-driving. In general, optical 3D measurement systems can be categorized into four major types: laser scanning, time-of-flight (ToF), stereo vision, and fringe projection. Laser scanning provides high precision but unsuitable for dynamic measurements<sup>1</sup>. ToF methods enable high-speed imaging but often suffer from low spatial resolution<sup>2</sup>. Stereo vision leverages multi-view information to enable rapid surface reconstruction, yet its reliability degrades in textureless or repetitive regions<sup>3</sup>. In contrast, fringe projection systems use a single camera and a projector to cast pre-coded patterns onto the object's surface. By analyzing the pattern deformation, these systems achieve dense, high-accuracy 3D reconstruction, making them effective solutions for full-field optical metrology<sup>4–8</sup>. Typically, the arctangent function is used to calculate the object's phase information in fringe projection systems. However,

due to the inherent periodicity of the arctangent, the phase is restricted to the principal value range  $(-\pi, \pi]$ , introducing  $2\pi$  discontinuities and resulting in phase ambiguity. To recover a continuous and unambiguous phase map, phase unwrapping algorithms are essential.

In the field of fringe projection profilometry (FPP), existing phase unwrapping methods can be broadly categorized into two types: spatial phase unwrapping (SPU)<sup>9–13</sup> and temporal phase unwrapping (TPU)<sup>14</sup>. SPU methods rely on the assumption of spatial continuity, whereby adjacent pixels are presumed to differ by no more than an integer multiple of  $2\pi$ . While this approach performs well for smooth and continuous surfaces, it often fails in scenarios involving geometric discontinuities, surface occlusions, or steep gradients, where error propagation and phase inconsistency become significant. In contrast, TPU methods introduce globally encoded auxiliary information to perform pixel-wise, unambiguous absolute phase

<sup>1</sup>Smart Computational Imaging Laboratory (SCILab), Nanjing University of Science and Technology, Nanjing, China. <sup>2</sup>Jiangsu Key Laboratory of Visual Sensing & Intelligent Perception, Nanjing, China. <sup>3</sup>State Key Laboratory of Extreme Environment Optoelectronic Dynamic Measurement Technology and Instrument, Taiyuan, China. ✉e-mail: [shijiefeng@njust.edu.cn](mailto:shijiefeng@njust.edu.cn)

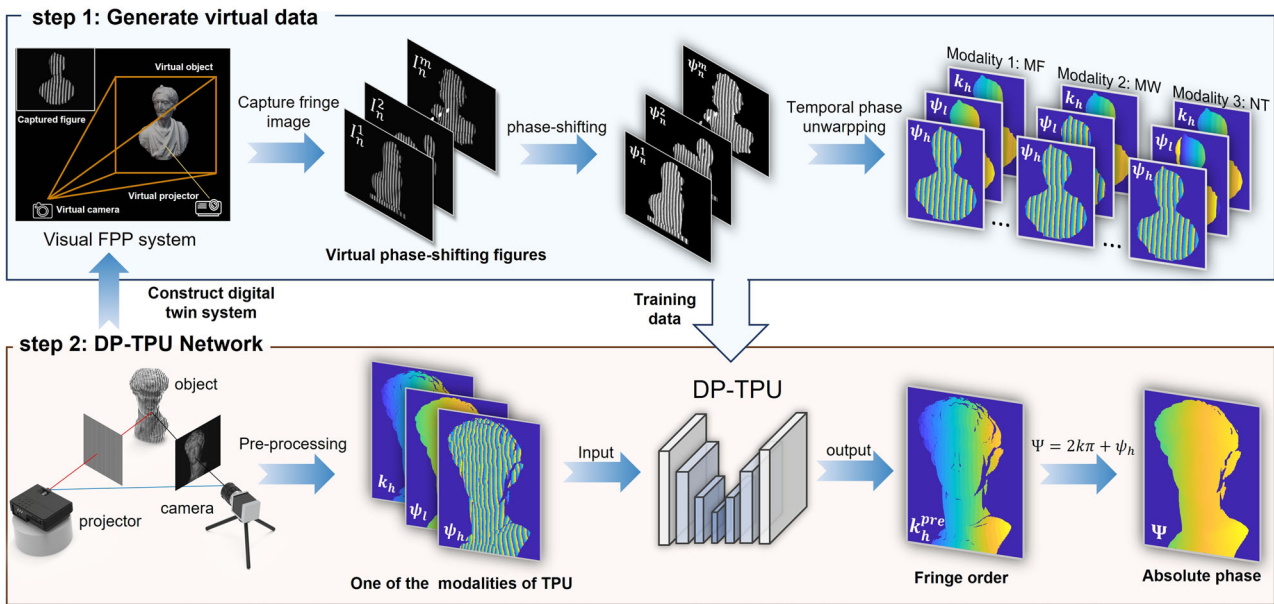


Fig. 1 | Workflow of the DP-TPU network for training and inference.

recovery, making them more suitable for complex or disconnected surfaces. Representative TPU approaches include multi-frequency (MF)<sup>15–17</sup> methods, multi-wavelength (MW)<sup>18–21</sup> methods, and number-theoretic (NT)<sup>22–24</sup> methods, as well as fringe encoding methods such as binary Gray codes<sup>25,26</sup>, spatial binary coding<sup>27–29</sup>, and phase encoding<sup>30</sup>.

In recent years, deep learning has been widely applied in the field of structured-light 3D imaging. Feng et al.<sup>31,32</sup> proposed a neural network-based fringe analysis method that predicts the numerator and denominator of the arctangent function, enabling high-precision wrapped phase recovery from a single image. Li et al.<sup>33</sup> introduced a cross-domain adaptive learning (CDL) framework that integrates a Mixture of Experts (MoE) model with a gating mechanism. By training multiple expert networks under domain randomization and dynamically fusing their features, this approach improves generalization and robustness across diverse imaging systems and measurement conditions. For phase unwrapping, deep learning has also been applied to SPU tasks. Wang et al.<sup>34</sup> demonstrated the effectiveness of simulation-driven training using synthetic datasets, achieving robust phase unwrapping in dynamic scenes such as live osteoblast imaging and candle flame tracking, with strong resilience to noise and aliasing artifacts. Zhang et al.<sup>35</sup> proposed an improved SegFormer-based network (SFNet) trained on large-scale synthetic datasets containing various noise patterns and phase discontinuities. Their method exhibits stable and accurate phase recovery in complex scenes. Despite the promising performance of SPU under simulation-driven paradigms, its inherent limitations in handling geometrically discontinuous or occluded scenarios make TPU, which exploits temporal information, a more suitable choice. To this end, Yin et al.<sup>36</sup> introduced deep learning into TPU by training deep neural networks on real data to enhance MF phase unwrapping accuracy. Guo et al.<sup>37</sup> proposed FOA-Net, a multi-scale residual network unifying MF, MW, and NT methods into a unifying TPU framework with enhanced noise suppression and robust performance in complex scenes. Liu et al.<sup>38</sup> further developed a multimodal adaptive TPU framework that combines deep learning with physical priors, enabling high-precision phase unwrapping for previously unseen frequencies and systems. Despite achieving improved generalization, this method still heavily relies on real-world data.

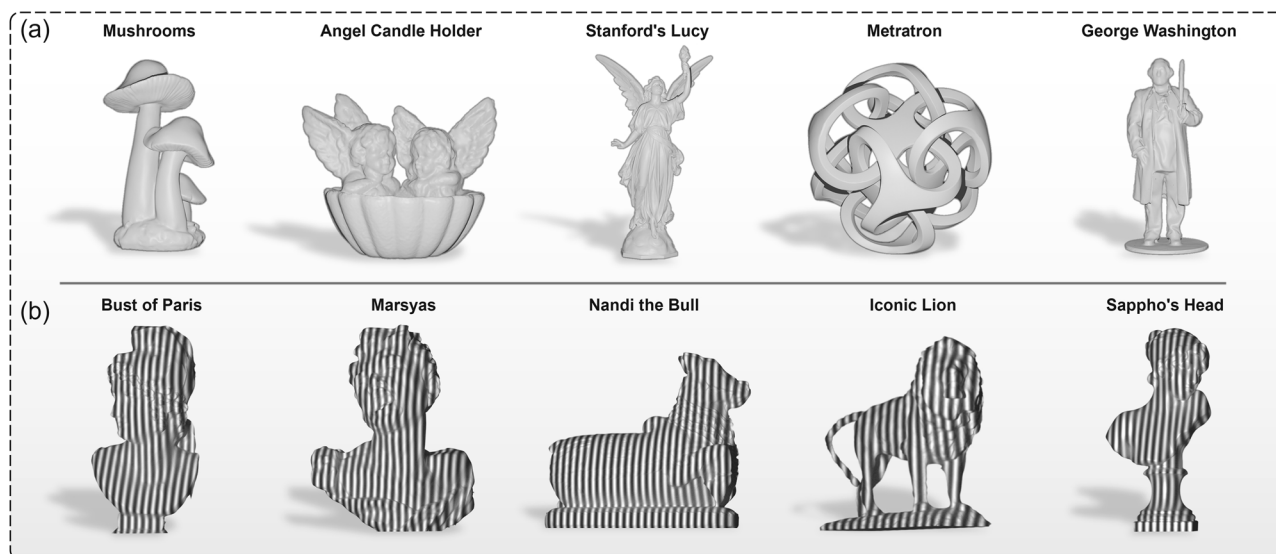
Although deep learning has shown great potential in phase unwrapping tasks, most existing methods still depend heavily on large-scale real-world datasets, resulting in high data acquisition costs and complex experimental workflows. Although Li et al.<sup>39</sup> proposed a simulation-driven deep learning approach for TPU, its applicability remains limited to multi-

frequency temporal phase unwrapping after training. Moreover, there often exists a significant domain gap between synthetic data generated within a specific system and real data captured by different measurement systems. Even with meticulous modeling, it remains challenging to fully replicate the intricate characteristics of real-world measurements. To address this issue, researchers have explored domain adaptation strategies<sup>40</sup> to narrow the performance gap between synthetic and real domains. However, such methods usually require intricate training procedures and careful parameter tuning tailored to specific real systems, compromising their generalization capability in unseen conditions. Overall, current methods are caught in a dilemma between real and synthetic data, highlighting the urgent need for a structured-light 3D imaging approach that fully leverages the advantages of synthetic data while effectively mitigating domain discrepancies to achieve high accuracy, low cost, and strong adaptability.

To overcome these limitations, we propose a Digital-twin-driven unambiguous structured light 3D imaging with physics-aware learning (DP-TPU) framework for unambiguous structured-light 3D imaging, as illustrated in Fig. 1. This framework leverages digital twin technology to construct a large-scale, high-fidelity virtual fringe image dataset, which effectively replacing manually collected training data and significantly reducing data acquisition costs. Furthermore, Fourier-domain consistency constraints and physical priors derived from MF, MW, and NT TPU models are integrated to overcome domain gaps, enhancing the model’s generalization capability and reconstruction accuracy across diverse system parameters and unseen scenarios. Experiments based on fringe images of varying frequencies and from diverse imaging systems show that our method achieves phase unwrapping accuracy comparable to models trained on real-world data. Remarkably, it maintains over 94% unwrapping accuracy even under high-frequency fringe conditions where conventional methods fail. We believe this work provides a promising solution for the development of low-cost, highly generalizable, and intelligent 3D imaging techniques.

## Results Setup

We constructed a structured light projection system consisting of a projector (DLP 4500, Texas Instruments) and an industrial camera (Basler acA640-750 μm). After performing system calibration, we obtained the following parameters: a focal length of the camera of 12 mm, a baseline distance of 0.17 m between the camera and the projector, an object-to-baseline



**Fig. 2 | Representative training 3D models rendered in a single Blender scene. a** 3D models rendered under white light. **b** 3D models rendered under fringe projection.

perpendicular distance of 0.4 m, and a defined angular separation of  $6.21^\circ$  between the optical axes of the camera and projector. Detailed calibration parameters for both the real-world and digital twin systems are provided in Supplementary Material 1. Based on these parameters, we constructed a digital twin of the FPP system in Blender<sup>41</sup> to generate synthetic training datasets.

The virtual camera is a Blender perspective camera and the virtual projector is a calibrated spot-light projector, which projects the fringe image as an emission texture. Their intrinsics and poses are set to match the calibrated values. To construct a highly generalizable training dataset, we adopted the Thing10K<sup>42</sup> dataset. This dataset contains 10,000 models, covering a wide range of object categories and geometric complexities. All synthetic data were generated using the “Cycles” path-tracing renderer in Blender, which accurately simulates the physical propagation of light paths to produce highly realistic images. To enhance the diversity of the virtual dataset, all virtual objects were assigned Physically Based Rendering (PBR) materials using the “Principled BSDF” shader in Blender. Key material properties—including roughness and index of refraction (IOR)—were randomized within physically plausible ranges to simulate a wide variety of real-world surfaces. To improve generalization while keeping objects within the camera’s field of view, we randomized their absolute depth to 0.8–1.5 m relative to the camera plane and scaled all models to an approximate height of 0.2 m. Representative samples of the resulting synthetic data are shown in Fig. 2. Moreover, the Blender scripts and some auxiliary code used for this study are available at <https://github.com/nomineee/DP-TPU>.

To enhance the model’s ability to analysis wrapped phase features, the wrapped phase maps were normalized to the range (0, 1] by dividing by  $2\pi$  before being fed into the network, thus improving training convergence. The network was implemented using PyTorch and trained on a NVIDIA RTX 4090 GPU. We employed the AdamW<sup>43</sup> optimizer with a batch size of 6, an initial learning rate of 0.001, and a total of 300 training epochs. It took about 11 h for training. A loss function composed of MSE and Fourier-domain consistency constraints was designed to jointly optimize unwrapping accuracy and structural consistency, with weighting parameters  $\lambda_1 = 0.9$  and  $\lambda_2 = 0.1$ , respectively. The effectiveness of this loss design is further validated through ablation experiments detailed in Supplementary Material 1. In particular, all test objects were excluded from the training set to ensure unbiased performance evaluation.

### Evaluation of the digital twin’s accuracy

To validate the accuracy of the geometric and optical modeling in the proposed digital twin system, we designed a cross-validation test using a

standard spherical object to quantitative evaluation of the physical fidelity between the digital twin and real measurement systems. In this experiment, a standard ceramic sphere with a diameter of 50.8082 mm was used as the measurement target. Its absolute phase map was obtained using the twelve-step phase-shifting method and a three-frequency TPU algorithm. A 3D reconstruction was then performed using the real system’s calibration parameters to serve as the ground-truth reference.

Subsequently, the sphere’s center coordinates were estimated using the robust sphere fitting algorithm proposed by Torr and Zisserman<sup>44</sup>. Based on these world coordinates, a virtual standard sphere with identical coordinates and diameter was created in the digital twin system. Finally, the calibration parameters of the real and digital twin systems were cross-applied to the image data acquired from both systems to reconstruct corresponding 3D point clouds. The similarity between the two systems was then compared. Detailed information on the calibration method of the digital twin system is provided in Supplementary Material 1.

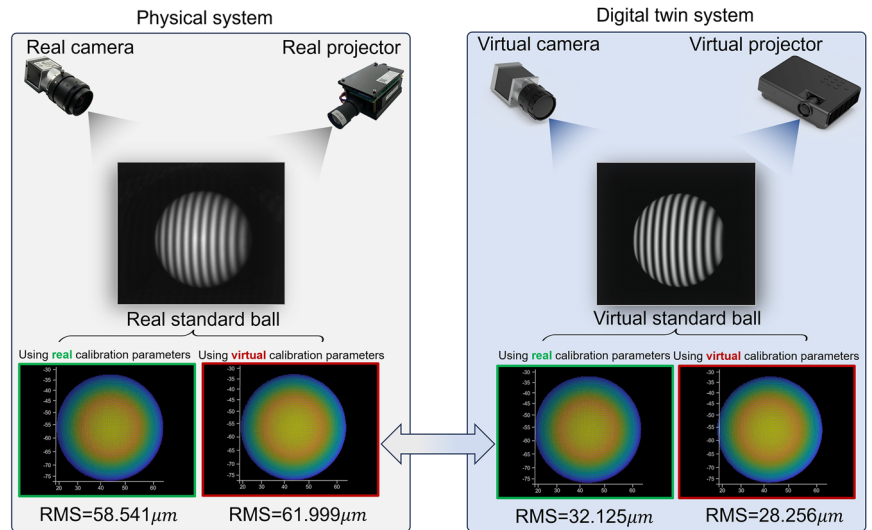
Figure 3 illustrates one of the phase-shifting fringe images of the real standard sphere and its corresponding virtual sphere in the digital twin system, along with their respective 3D reconstruction results using real and virtual calibration parameters. It can be observed that when the two spheres are placed at identical positions in both systems, the generated fringe images exhibit high consistency in spatial distribution, fringe width, and periodicity.

For each scene, the reconstructed 3D point clouds were fitted to a sphere, and the root mean square (RMS) error relative to the ground-truth sphere was calculated as the evaluation metric. Experimental results show that for the real sphere, the RMS reconstruction error using the real system’s calibration parameters is 58.541  $\mu\text{m}$ , while that using the digital twin’s calibration parameters slightly increases to 61.999  $\mu\text{m}$  when using digital twin calibration parameters, with an error margin below 4  $\mu\text{m}$ . This confirms that the calibration parameters in the digital twin system accurately approximate those of the real system. For the virtual sphere, where noise is negligible, the RMS values obtained using real and digital twin calibration parameters are 32.125  $\mu\text{m}$  and 28.256  $\mu\text{m}$ , respectively. These results confirm that the digital twin system can faithfully replicate the geometric and optical characteristics of the real measurement system, ensuring high physical fidelity in the generated virtual data for subsequent tasks.

### Selection of training fringe images

To ensure that our model can generalize to diverse unseen fringe images while relying on a minimal number of training samples, we tested different combinations of virtual fringe patterns in this experiment. The target frequency domain was defined as  $\{f_h | f_h^1 \leq f_h^m \leq f_h^N\}$ , where  $f_h^1 = 16$ ,  $f_h^N = 96$ ,

**Fig. 3 |** Cross-reconstruction validation of physical fidelity between the digital twin and real measurement systems.



and  $f_h^n$  represents the  $n$ -th frequency in the domain ( $n = 1, 2, \dots, N$ ). Nine representative frequencies were selected for analysis:  $f_h^n = \{16, 32, 36, 48, 56, 64, 76, 80, 96\}$ . The network’s performance was evaluated under five distinct training frequency combinations ( $S_1 = \{56\}$ ,  $S_2 = \{16, 96\}$ ,  $S_3 = \{16, 56, 96\}$ ,  $S_4 = \{32, 48, 64, 80\}$ ,  $S_5 = \{16, 36, 56, 76, 96\}$ ), as illustrated in Fig. 4a. These combinations were designed to assess the impact of frequency diversity on the network’s generalization ability and phase unwrapping accuracy.

As shown in Fig. 4b, under low-frequency testing conditions ( $f_h^n \leq 56$ ), all five training combinations yielded high phase unwrapping accuracy, with all models achieving over 98% and showing negligible differences across strategies. However, as the testing frequency increased, models trained with single-frequency  $S_1$  began to deteriorate rapidly in performance. For example, at  $f_h^n = 96$ , the accuracy of the multi-frequency TPU network trained using combination  $S_1$  dropped to 85.83%, significantly lower than that of traditional TPU algorithms. This highlights the inadequacy of single-frequency training for achieving robust frequency generalization.

When multiple frequency training strategies were applied, the network demonstrated significant improvements in both frequency generalization and high-frequency unwrapping accuracy. For example, under the number-theoretic TPU method with the dual-frequency training set  $S_2 = \{16, 96\}$ , the unwrapping accuracy at  $f_h^n = 96$  reached 93.96%, representing an 8.1% improvement over the single-frequency strategy  $S_1$ . Expanding to a triple-frequency training set  $S_3 = \{16, 56, 96\}$  further increased accuracy to 94.92%, demonstrating the benefit of exposing the network to a broader range of training frequencies for improved generalization. However, increasing the number of training frequencies beyond this point resulted in diminishing marginal returns. For instance, extending the frequency set from  $S_4 = \{32, 48, 64, 80\}$  to  $S_5 = \{16, 36, 56, 76, 96\}$  improved accuracy at  $f_h^n = 96$  by a maximum gain of 0.1% for three TPU methods, while incurring a 25% increase in training cost. This indicates that while increasing frequency diversity can improve performance, it also significantly raises training costs, highlighting the need to balance accuracy gains with computational resource consumption when selecting training combinations. Figure 4c statistically illustrates the overall accuracy distribution of different combination strategies under the MF, MW, and NT training mechanisms. The boxplots visualize the distribution of accuracy across nine testing frequencies for each combination. Evidently, the single-frequency strategy  $S_1$  exhibits the widest interquartile range and the lowest mean accuracy, indicating poor generalization and high variability in high-frequency scenes. In contrast, the dual-frequency combination  $S_2$  significantly improves both average accuracy and stability. For example, under the MF strategy, mean accuracy increases from 94.92% ( $S_1$ ) to approximately 97.17% ( $S_2$ ), though

some performance fluctuation remains. As the number of training frequencies increases from  $S_3$  to  $S_5$ , the overall accuracy and stability further improve. The interquartile range narrows progressively, and the gap between upper and lower quartiles diminishes, indicating consistent performance across most test frequencies. However, when extending the frequency combination from  $S_4$  to  $S_5$ , improvements in both accuracy and stability approach saturation.

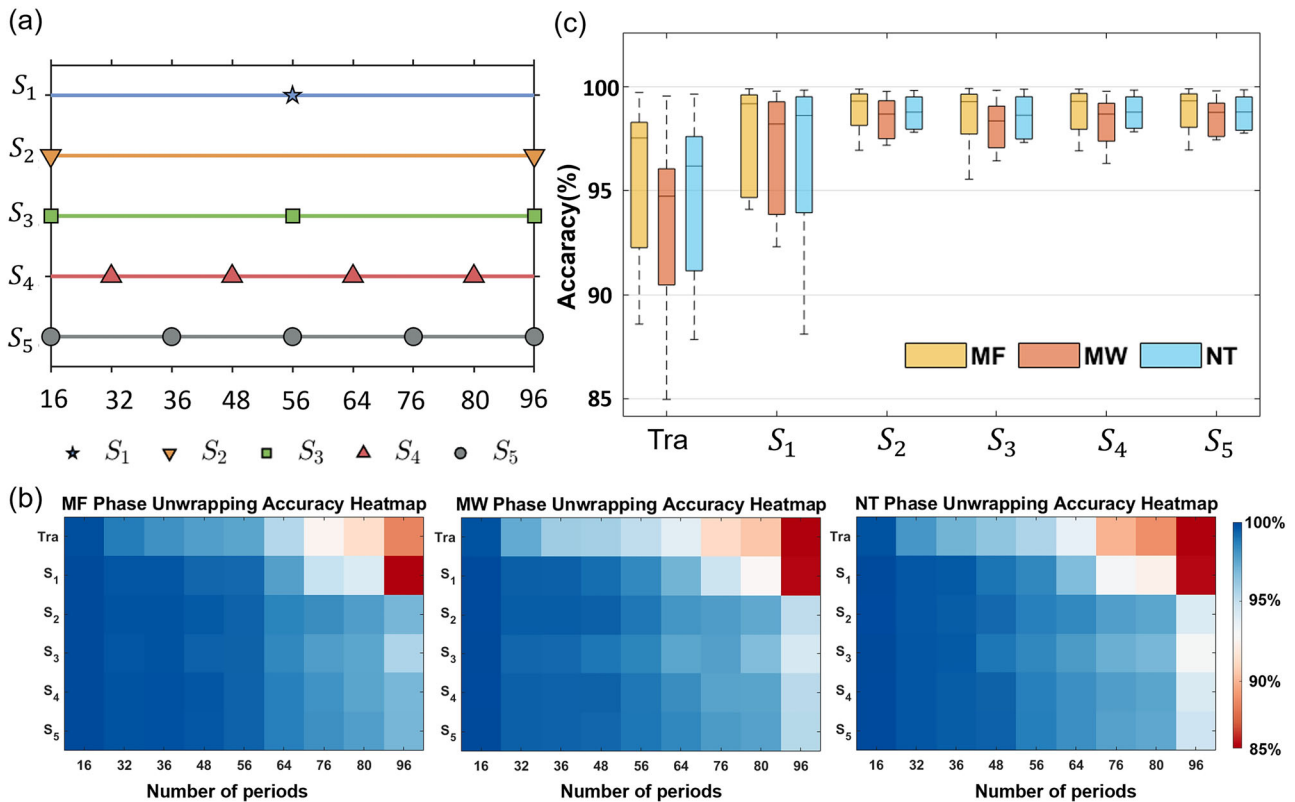
Therefore, considering the trade-off between performance and computational cost, we selected the frequency set  $S_4 = \{32, 48, 64, 80\}$  for use in subsequent experiments. This combination ensures high and stable unwrapping accuracy across the target frequency range while maintaining efficient training, achieving an optimal balance between performance and resource usage.

### Unambiguous 3D reconstruction of static objects

To evaluate the adaptability of the proposed method in real measurement systems, we collected real-world fringe images across the target frequency domain  $\{f_h | 16 \leq f_h^n \leq 96\}$  with an interval of 4 frequencies. In addition, we compared the performance of the network trained using synthetic data with that of a network trained on real data. During training, we adopted the optimized frequency combination  $S_4 = \{32, 48, 64, 80\}$ , as determined in Section “Selection of training fringe images”. For the MF method, the auxiliary grating frequencies were set to  $f_l = \{1, 1, 1, 1\}$ ; for the MW method,  $f_l = \{31, 47, 63, 79\}$ ; and for the NT method,  $f_l = \{10.9, 10.9, 10.9, 10.9\}$ .

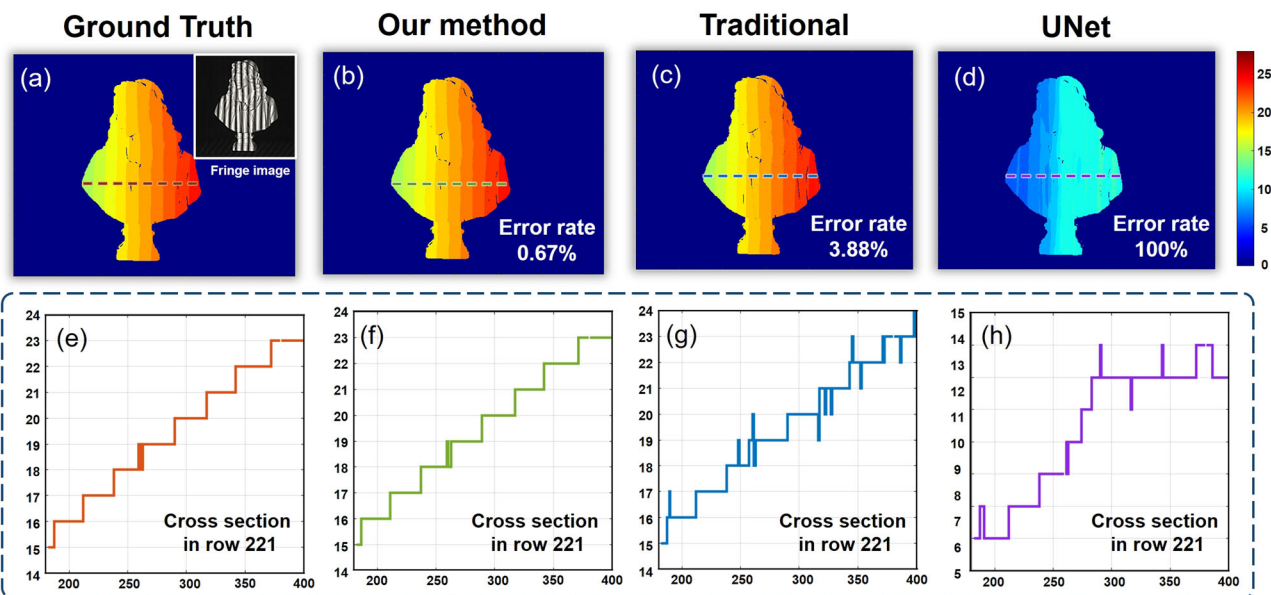
For each of these three TPU methods and their corresponding frequency combinations, we used the digital twin FPP system to generate 800 sets of dual-frequency twelve-step phase-shifted virtual fringe images across diverse scenes. Additionally, 300 real-world dual-frequency twelve-step fringe image sets were collected to train the real-data-driven model. As a result, the real dataset contained 3600 sets, while the virtual dataset included 9,600 sets in total. To assess the performance gap between the digital-twin-driven and real-data-driven approaches, a separate test set comprising 540 groups of real dual-frequency three-step fringe images was used for comparative analysis. Notably, conventional U-Net models without physical priors completely failed when tested on previously unseen high-frequency fringes (e.g.,  $f_h^n = 96$ ), yielding 100% error rates (see Supplementary Material 1 for detailed analysis).

We first compared the fringe-order estimates produced by different methods. As shown in Fig. 5, we compared three approaches under the MF method: our DP-TPU method, the traditional TPU method (providing coarse fringe order maps), and a conventional UNet model trained only at a single frequency ( $f_h = 32$ ). Figure 5a–d presents the ground truth and results



**Fig. 4 | Comparison of frequency generalization capabilities under different training strategies.**  $S_1 - S_5$  denote the five training frequency combinations, and “Tra” represents traditional TPU method. **a** Schematic of high-frequency grating

periods used in each training combination; **b** heatmap of phase unwrapping accuracy for the three TPU methods (MF, MW, and NT); **c** boxplots of accuracy distributions across different combinations for each TPU method.

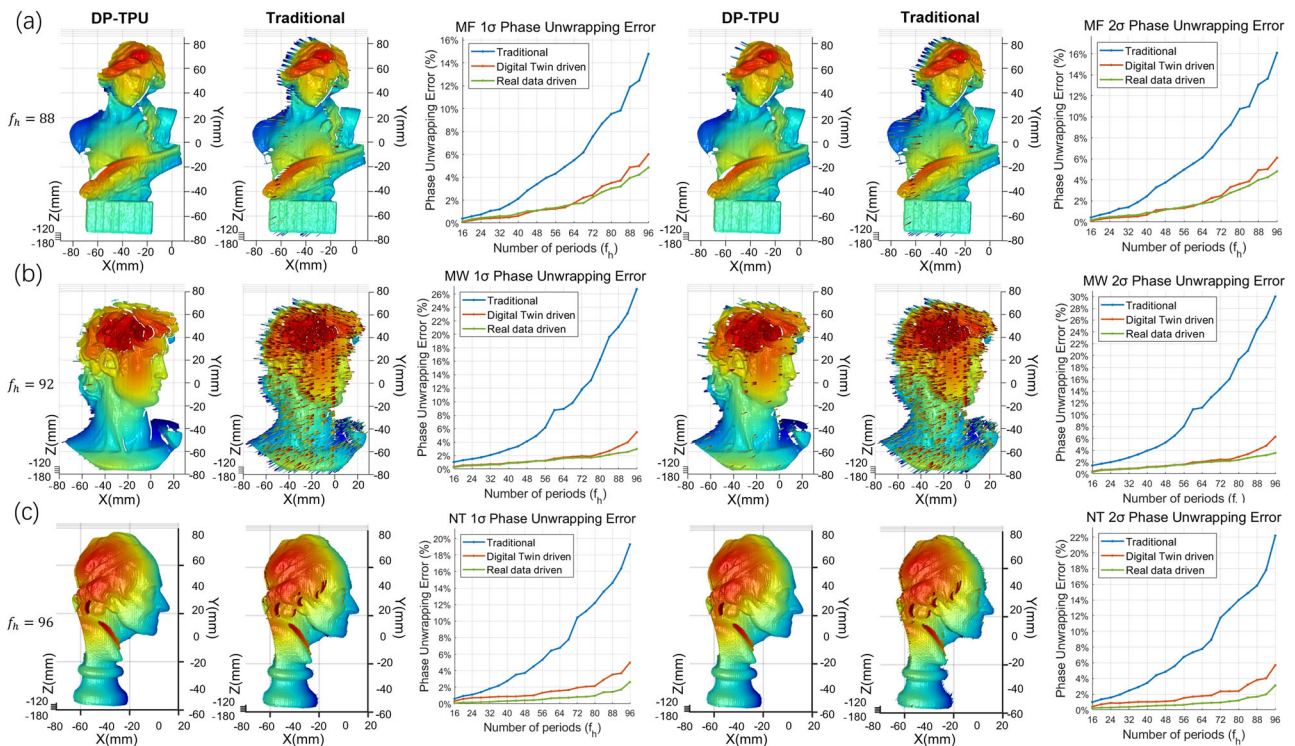


**Fig. 5 | Fringe order maps and cross-sectional comparisons for MF methods.** **a-d** Correspond to ground truth, our method, traditional TPU (coarse fringe order map), and a dual-input UNet model trained only at a single frequency ( $f_i = 32$ ), respectively. **e-h** Show the cross-sections along row 221 for each method.

of the three methods respectively, where the UNet completely failed (100% error rate) when facing the unseen test frequency ( $f_i = 52$ ), while our DP-TPU achieved the best phase unwrapping accuracy (0.67% error rate) - outperforming the traditional TPU (3.88% error rate). Figure 5e-h shows the cross-sectional profiles along row 221 of these methods. The traditional TPU suffered from obvious phase jump errors, and the UNet lost the ability

for absolute depth estimation. In contrast, our DP-TPU produced a profile closely matching the ground truth, achieving fairly good fringe order recovery results.

Subsequently, we performed 3D reconstruction on the results obtained by the three TPU methods. Figure 6a-c illustrates the phase unwrapping error versus frequency and 3D reconstruction performance for the MF,



**Fig. 6 | Comparison of 3D reconstruction quality and phase unwrapping accuracy among traditional TPU, real-data-driven, and digital-twin-driven methods.** a–c Shows the phase unwrapping error versus frequency and 3D reconstruction

results for MF, MW, and NT methods under different noise levels. Left: 3D reconstruction comparison between DP-TPU and traditional TPU algorithms. Right: Phase unwrapping error curves for different training strategies across frequencies.

MW, and NT methods under real and virtual data-driven conditions, respectively. The results show that our method achieves excellent adaptability across all three TPU methods, maintaining low unwrapping error rates over a wide frequency range. In the low-to-medium frequency domain ( $f_h^n \leq 56$ ), both the digital-twin-driven and real-data-driven models achieve stable unwrapping accuracy above 98%, with a maximum performance gap of 0.5%, indicating equivalent effectiveness under moderate conditions. Even under high-noise ( $2\sigma$ ) and high-frequency conditions ( $56 \leq f_h^n \leq 92$ ), the digital-twin-driven network remains robust, with accuracy gaps relative to the real data-driven network consistently below 1%. For instance, at  $f_h^n = 84$ , the virtual data-driven MF method achieves an accuracy of 96.29%, only 0.5% lower than its real data-driven counterpart. Similarly, the MW and NT methods achieve 95.24% and 95.98% accuracy, respectively, with performance gaps of less than 1%. These results highlight the method’s exceptional sim-to-real transfer capability in high-noise, high-frequency scenes. Remarkably, in scenes where traditional TPU methods performed poorly, the proposed method consistently maintains unwrapping accuracy exceeding 94%. This demonstrates that, despite relying solely on virtual data for training, our method achieves performance comparable to real data-driven models. Further experiments validating the effectiveness of the digital twin strategy and physical priors are detailed in Supplementary Material 1.

**Unambiguous 3D reconstruction of dynamic objects**

To further evaluate the adaptability of the proposed method in dynamic scenes, we built a new FPP system that is composed of a high-speed CMOS camera (Vision Research Phantom V611) and a customized projection system with an XGA-resolution ( $1024 \times 768$ ) digital micromirror device (DMD). The DMD operated in binary (1-bit) mode to achieve a refresh rate of 1000 fps. The camera was equipped with an 18.7 mm focal length lens. The baseline distance between the camera and projector was set to 0.12 m, the object-to-baseline vertical distance was 0.6 m, and an angle of  $10.07^\circ$  between the camera and projector.

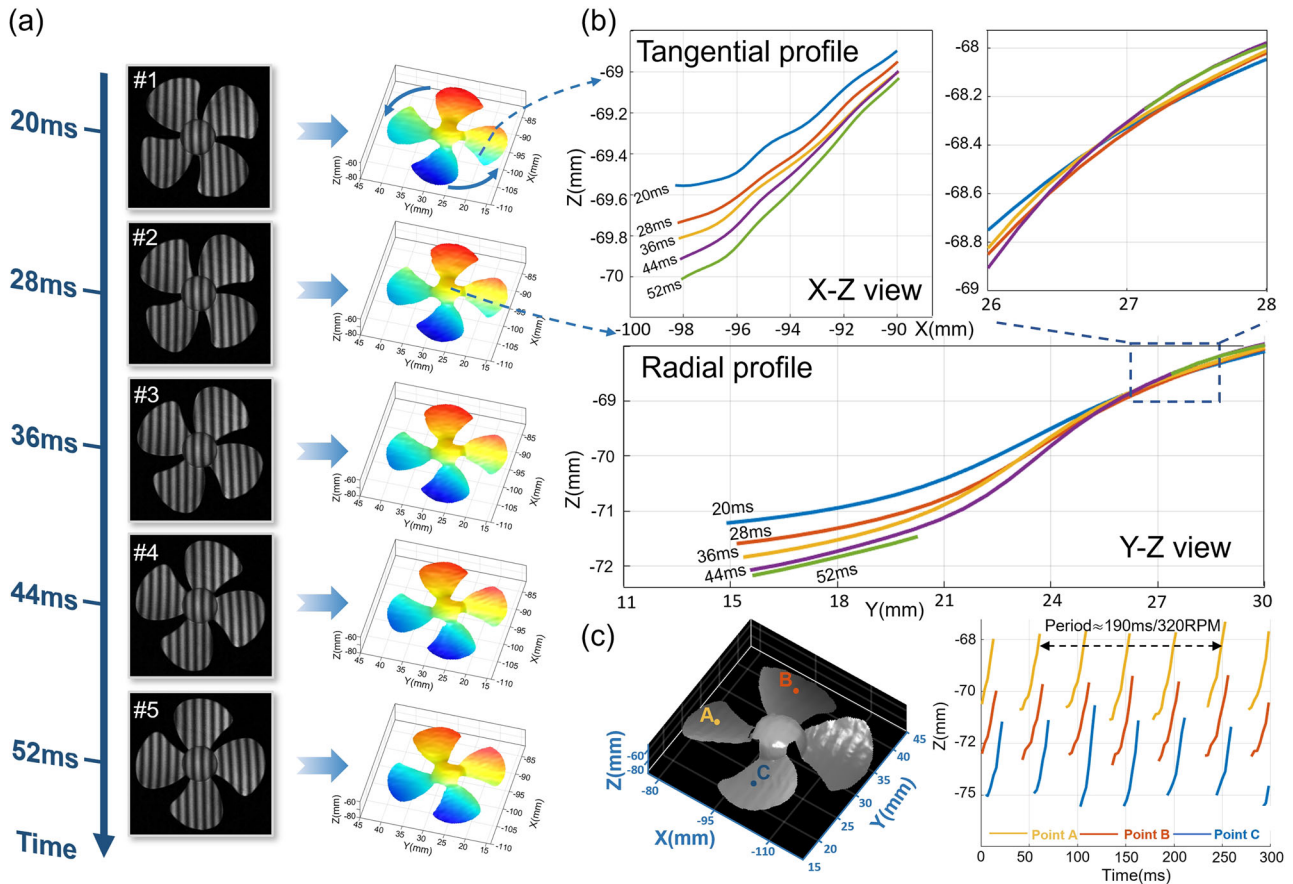
A motor-driven four-blade plastic fan was selected as the dynamic target. To rigorously evaluate the generalization capability of our method, all

dynamic experiments involved fringe frequencies and systems entirely unseen during training. To mitigate motion artifacts, the proposed DP-TPU method was integrated with a deep learning-based single-frame fringe analysis technique<sup>45</sup>.

The experimental results are shown in Fig. 7. We evaluated the MF method of our method on a rotating four-blade plastic fan using an unseen system configuration and an unseen fringe frequency ( $f_h = 96$ ). Figure 7a shows the 3D reconstruction results of our proposed DP-TPU at a representative frame. The reconstruction results demonstrate that the method successfully captures the fine geometric structures of the blades while effectively suppressing phase jumps caused by motion blur and noise. Figure 7b presents cross-sectional views of the 3D reconstruction results at five consecutive time points, highlighting the local structural details captured by the proposed method. These results demonstrate that our approach can accurately recover the continuous geometric deformation of the fan blades during rotation, faithfully representing dynamic changes in blade shape and thickness, which reflects the method’s high 3D reconstruction precision under dynamic conditions. Figure 7c shows the temporal depth variations of three fixed points located on the fan blades. The resulting depth trajectories exhibit smooth and clearly periodic patterns, accurately reflecting the blades’ rotational motion over time. Notably, the periodicity of the depth variation reveals a rotation period of approximately 190 ms per revolution (equivalent to 320 rotation per minute), demonstrating the method’s ability to suppress noise and motion artifacts in high-speed dynamic scenes. A complete visualization of the reconstruction sequence is provided in Virtualization 1. These results demonstrate that by combining the efficient feature extraction capabilities of deep learning with physical priors from traditional TPU models, the proposed method exhibits strong generalization and robustness in high-speed dynamic 3D measurement tasks.

**Discussions**

This paper proposes a digital-twin-driven unambiguous structured light 3D imaging with physics-aware learning (DP-TPU), enabling high-precision



**Fig. 7 | Performance of DP-TPU in high-speed dynamic 3D reconstruction.** **a** 3D reconstruction results at five consecutive time points. **b** Cross-sectional reconstruction accuracy and local detail comparison. **c** Depth variation over time for selected points on the fan blades.

execution of multi-frequency, multi-wavelength, and number-theoretic TPU tasks without requiring real-world training data. By constructing a highly physically faithful FPP digital twin system, the method generates abundant virtual data, significantly reducing data acquisition costs and bridging the domain gap between synthetic and real-world data. To enhance the network’s perception of fringe hierarchy, a Fourier-domain consistency constraint is introduced into the loss function. This constraint enforces alignment between predicted and ground-truth phase distributions in the frequency domain, improving structural fidelity in high-frequency regions.

To further overcome the limited generalization of networks trained solely on simulation data, we incorporate the physical models of TPU as priors. These priors guide the network to learn the intrinsic relationship between wrapped phase and fringe order, enabling robust cross-domain generalization. Specifically, the incorporation of the fringe order as a physical prior provides the network with an explicit representation of absolute depth, substantially simplifying the learning process. This approach not only improves unwrapping accuracy under high-frequency patterns where conventional methods fail but also greatly enhances the model’s adaptability to different system configurations and fringe frequencies.

The proposed DP-TPU framework not only supports a wide range of fringe frequencies but also adapts to distribution shifts caused by varying imaging systems. Moreover, it generalizes three TPU methods into a single training pipeline, enabling phase unwrapping using a single model and significantly enhancing practical applicability and measurement efficiency. Experimental results demonstrate that the proposed method achieves superior performance across diverse frequencies and system conditions. Even in high-frequency scenes where traditional methods fail, the proposed network consistently

maintains over 94% phase unwrapping accuracy, matching the performance of real-data-trained models. These results affirm that the integration of digital twin technology and physical priors greatly enhances both the generalization and robustness of deep learning-based phase unwrapping. This work provides a promising pathway toward low-cost, high-precision, and highly generalizable intelligent optical 3D measurement technologies. Nonetheless, a last-mile gap may persist between the real system and its digital twin when effects such as defocus blur, optical aberrations, or photometric nonlinearities are present. Future work in this area would benefit from leveraging differentiable rendering<sup>46</sup> to model these residual effects, enabling the digital twin to more precisely mimic specific hardware configurations and further narrow the sim-to-real discrepancy.

## Methods

### Phase calculation

In a typical FPP system, the projector casts computer-generated sinusoidal fringe patterns onto the surface of the target object. As the object surface varies in height, the projected fringes become distorted. According to the N-step phase-shifting method, the captured intensity distribution  $I_n^c(x, y)$  can be expressed as

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

where  $A^c(x, y)$  denotes the background illumination,  $B^c(x, y)$  represents the modulation of the sinusoidal fringe,  $n$  is the phase-shifting index ( $n = 0, 1, \dots, N - 1$ ),  $N$  is the total number of phase steps, and  $\psi(x, y)$  is the wrapped phase. The wrapped phase  $\psi(x, y)$  is typically recovered

through a least-squares estimation approach

$$\psi(x, y) = \arctan \frac{\sum_{n=1}^N I_n^c(x, y) \sin\left(\frac{2\pi n}{N}\right)}{\sum_{n=1}^N I_n^c(x, y) \cos\left(\frac{2\pi n}{N}\right)}. \quad (2)$$

### Temporal phase unwrapping

As shown in Eq. (2), the arc tangent function confines the wrapped phase  $\psi(x, y)$  to the range  $(-\pi, \pi]$ , introducing  $2k\pi$  ambiguities. To recover the absolute phase, temporal phase unwrapping (TPU) methods are employed. The fundamental principle of TPU can be expressed as

$$\Psi(x, y) = \psi(x, y) + 2\pi k(x, y), \quad (3)$$

where  $\Psi(x, y)$  is the absolute phase, and  $k(x, y)$  is the fringe order ( $k \in \mathbb{Z}$ ). The key challenge lies in accurately determining  $k(x, y)$  for each pixel. This work focuses on three TPU methods: the MF method, MW method, and NT method. All three methods utilize at least one set of low-frequency wrapped phases to determine  $k(x, y)$ . Based on Eq. (3) and the phase ratio relationship between high- and low- frequency wrapped phases, we derive

$$\begin{cases} \Psi_h(x, y) = \psi_h(x, y) + 2\pi k_h(x, y), \\ \Psi_l(x, y) = \psi_l(x, y) + 2\pi k_l(x, y), \\ \Psi_h(x, y) = \frac{f_h}{f_l} \Psi_l(x, y), \end{cases} \quad (4)$$

where  $k_h(x, y)$  and  $k_l(x, y)$  are the fringe orders corresponding to the high- and low- frequency wrapped phases, respectively. This system contains three linearly independent equations but four unknowns ( $k_h(x, y)$ ,  $k_l(x, y)$ ,  $\Psi_h(x, y)$ , and  $\Psi_l(x, y)$ ), resulting in underdetermination. To resolve this, additional constraints are introduced through carefully designed high- and low-frequency fringe patterns.

One common strategy is to set the frequency of the auxiliary grating to  $f_l = 1$ , which ensures that the associated fringe order  $k_l(x, y)$  becomes zero. This method is known as multi-frequency TPU<sup>47</sup>. Thus,  $k_h(x, y)$  can be derived as

$$k_h^{MF}(x, y) = \text{Round} \left[ \frac{f_h \Psi_l(x, y) - f_l \Psi_h(x, y)}{2\pi f_l} \right]. \quad (5)$$

Similarly, wrapped phase can also be unwrapped using the equivalent phase generated by subtracting the low-frequency wrapped phase from the high-frequency wrapped phase. This method is called multi-wavelength TPU<sup>48</sup>. The equivalent phase  $\psi_{eq}$  and equivalent frequency  $f_{eq}$  are defined as

$$\begin{cases} \psi_{eq}(x, y) = \text{mod}(\psi_h(x, y) - \psi_l(x, y), 2\pi), \\ f_{eq} = f_h - f_l. \end{cases} \quad (6)$$

In order to ensure the unambiguity of phase unwrapping, a suitable low-frequency grating must be selected so that the equivalent frequency  $f_{eq}$  meets  $f_{eq} \leq 1$ , which can also ensure that  $k_{eq}(x, y) = 0$ . Therefore, with the aid of the equivalent phase,  $k_h(x, y)$  can be expressed as

$$k_h^{MW}(x, y) = \text{Round} \left[ \frac{f_h \psi_{eq}(x, y) - f_{eq} \psi_h(x, y)}{2\pi f_{eq}} \right]. \quad (7)$$

Since  $k_h(x, y)$  and  $k_l(x, y)$  must be positive integers, it is proposed that the fringe order pair  $(k_h, k_l)$  of two groups of coprime sinusoidal fringe patterns can be determined by using the wavelengths  $\lambda_h$  and  $\lambda_l$  of the two groups. This method is called Number-theoretic (NT) TPU<sup>49</sup>. Deforming

Eq. (4), we can derive

$$\frac{\lambda_l \Psi_l - \lambda_h \Psi_h}{2\pi} = \lambda_h k_h - \lambda_l k_l. \quad (8)$$

In order to ensure the uniqueness of the fringe order pair  $(k_h, k_l)$ , For two coprime wavelengths  $\lambda_h$  and  $\lambda_l$ , their least common multiple (LCM) must satisfy  $LCM(\lambda_h, \lambda_l) \geq W^{50}$ . The fringe order pair  $(k_h, k_l)$  can be determined using a precomputed lookup table (LUT)<sup>51</sup>

$$(k_h^{NT}, k_l^{NT}) = \text{LUT} \left[ \text{Round} \left( \frac{\lambda_l \Psi_l - \lambda_h \Psi_h}{2\pi} \right) \right]. \quad (9)$$

### Development of a digital twin system

We employed digital twin technology to construct a precise computational simulation system that replicates the physical characteristics and operational principles of a real FPP measurement system, enabling the generation of highly realistic virtual data<sup>52,53</sup>. Specifically, the calibration parameters of the real measurement system are computed and mapped to a digital twin FPP system within a computer rendering environment. This study utilizes the open-source CG software Blender to build the digital twin system and generate virtual fringe patterns.

For a structured light projection system, the mapping between pixel coordinates  $(u, v)$  on the camera and 3D spatial coordinates  $(x^w, y^w, z^w)$  in the world coordinate system can be described as

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R \quad T] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}, \quad (10)$$

where  $s$  denotes the scaling factor,  $K$  represents the camera's intrinsic matrix, while  $R$  and  $T$  correspond to the rotation and translation matrices, respectively.  $R$  and  $T$  are collectively termed the extrinsic matrix, defining the camera's pose relative to the world coordinate system. The intrinsic matrix  $K$  is further expressed as

$$K = \begin{bmatrix} f_u & \lambda & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (11)$$

where  $f_u$  and  $f_v$  are the camera's focal lengths along the  $u$ - and  $v$ - axes,  $\lambda$  is the skew factor between axes, and  $(u_0, v_0)$  are the coordinates of the optical center on the imaging plane.

Assuming the world coordinate origin is  $[0, 0, 0]^T$  and the camera coordinates are located at point  $P = [x_0^c, y_0^c, z_0^c]^T$ , we derive

$$0 = RP + T. \quad (12)$$

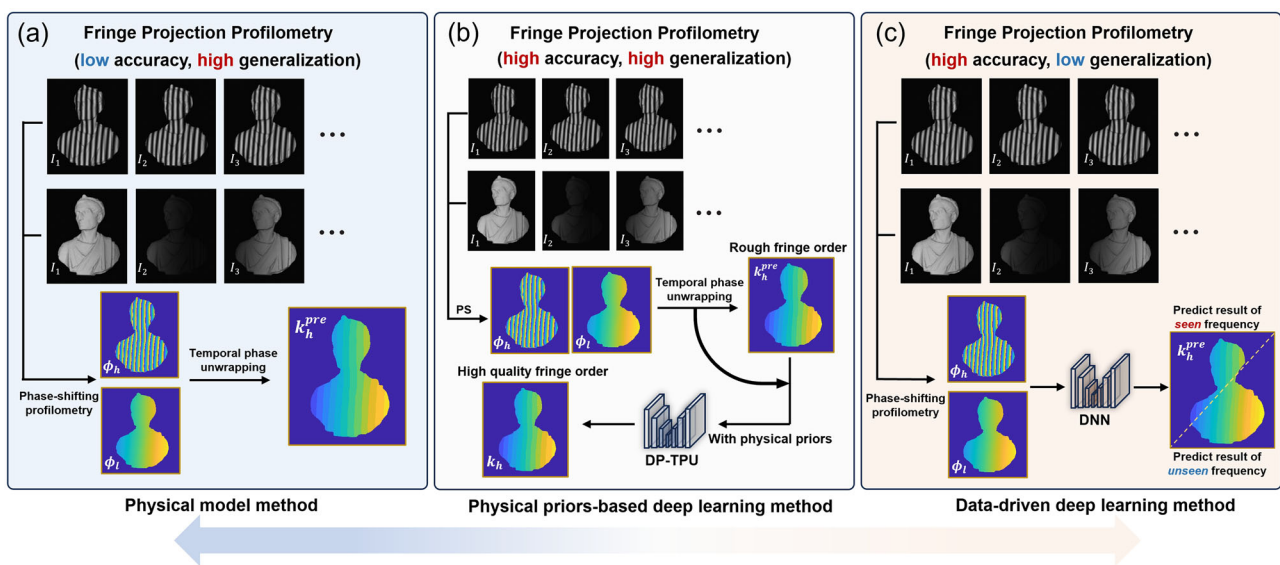
Since  $R$  is an orthogonal matrix ( $R^T = R^{-1}$ ), the camera's position is calculated as

$$P = -R^T T. \quad (13)$$

The camera's orientation is further defined by Euler angles  $\phi$ ,  $\theta$ , and  $\psi$  (rotations around the  $x$ -,  $y$ -, and  $z$ -axes, respectively), derived from  $R$  using

**Table 1 | Mapping between real-system calibration matrices and digital twin system parameters**

	Virtual camera	Virtual projector
<b>Intrinsic matrix</b>		
Focal length	$f_u^c, f_v^c$	$f_u^p, f_v^p$
Principle Points	$(u_0^c, v_0^c)$	$(u_0^p, v_0^p)$
Skew factor	$\lambda_c$	$\lambda_p$
<b>Extrinsic matrix</b>		
Location	$-R_c^T T_c$	$-R_p^T T_p$
Rotated angles	$\psi^c = \arctan\left(\frac{r_{32}^c}{r_{33}^c}\right)$	$\psi^p = \arctan\left(\frac{r_{32}^p}{r_{33}^p}\right)$
	$\theta^c = \arctan\left(\frac{-r_{31}^c}{\sqrt{(r_{32}^c)^2 + (r_{33}^c)^2}}\right)$	$\theta^p = \arctan\left(\frac{-r_{31}^p}{\sqrt{(r_{32}^p)^2 + (r_{33}^p)^2}}\right)$
	$\phi^c = \arctan\left(\frac{r_{21}^c}{r_{11}^c}\right)$	$\phi^p = \arctan\left(\frac{r_{21}^p}{r_{11}^p}\right)$



**Fig. 8 | Diagrams of three FPP methodologies. a** Physics-based modeling, **b** physics-prior guided, and **c** data-driven approaches.

the formula proposed by Slabaugh<sup>54</sup>

$$\begin{cases} \psi = \arctan(r_{32}, r_{33}), \\ \theta = \arctan\left(-r_{31}, \sqrt{r_{32}^2 + r_{33}^2}\right), \\ \phi = \arctan(r_{21}, r_{11}). \end{cases} \quad (14)$$

Thus, we complete the mapping from the real calibration matrices of the camera to the digital twin system configuration. To calibrate the intrinsic and extrinsic parameters of the camera, Zhang’s camera calibration algorithm<sup>55</sup> can be employed. For the projector, it is treated as an inverse camera<sup>56</sup>, and thus follows the same mathematical model. Consequently, a digital twin system with identical parameters to the real FPP system is constructed<sup>57</sup>. The specific mapping relationships are summarized in Table 1, where superscripts *c* and *p* denote the calibration parameters of the camera and projector, respectively.

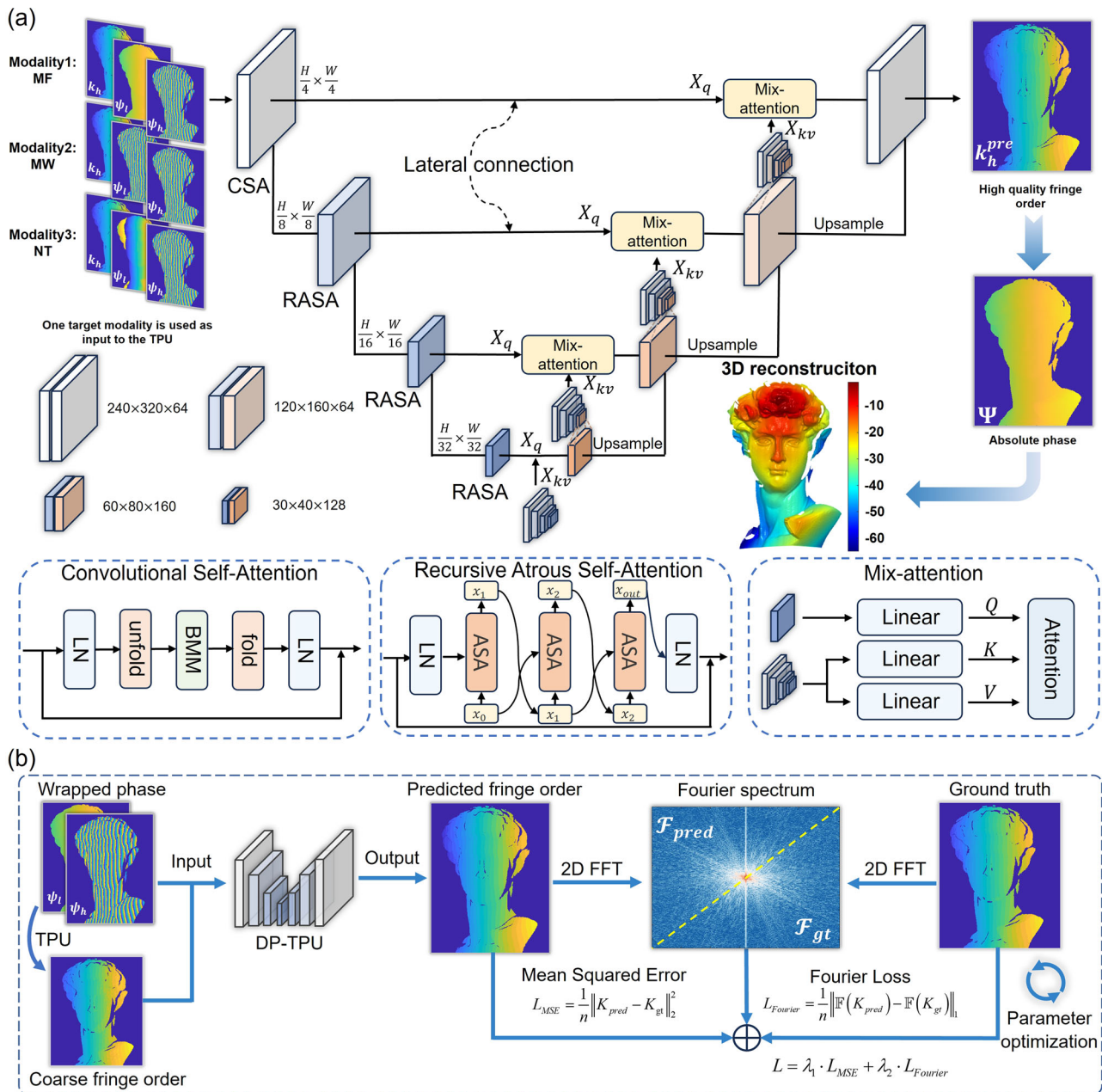
**Physics-aware generalized temporal phase unwrapping deep neural network**

As illustrated in Fig. 8a, c, traditional TPU algorithms rely on simplified physical models. While they offer strong theoretical generalization, their performance degrades significantly in noisy environments. In contrast, purely data-driven deep learning methods achieve higher unwrapping

precision by mining latent patterns from training data but suffer from poor generalization in unseen frequency or system scenarios due to the absence of theoretical constraints. To harmonize the strengths of both approaches, this work integrates the mathematical model of TPU as physical priors, constructing a hybrid physics-data-driven framework named DP-TPU, as illustrated in Fig. 8b. The network inputs include the high-frequency wrapped phase, auxiliary low-frequency wrapped phase, and a coarse fringe order map computed using traditional TPU algorithms. This design enables the network to refine predictions by combining physical priors with data-driven insights.

Training phase is conducted entirely on virtual data generated by the Blender-based digital twin system, eliminating reliance on real-world data and addressing annotation challenges. Ground-truth labels are computed from dual-frequency, 12-step phase-shifting fringe images using the same TPU formulas as in our real-data pipeline. Detailed calculations of synthesis fringe patterns can be found in Supplementary Material 1. After training, the network can be directly applied to real measurements for inference tasks<sup>58,59</sup>. The specific structure of the network is shown in Fig. 9a. To enhance global feature extraction, the network adopts a lightweight U-shaped Vision Transformer (ViT) architecture<sup>60</sup> to balance computational efficiency and performance.

Specifically, the proposed network integrates the deep learning architectures of Lite Vision Transformer (LVT)<sup>61</sup> and U-MixFormer<sup>62</sup> to



**Fig. 9 | The neural network used in our proposed method. a** The architecture of DP-TPU network. **b** Joint spatial-Fourier loss for fringe-order prediction.

efficiently encode and decode features. The network accepts high-frequency and low-frequency wrapped phases, along with the coarse fringe orders generated from any TPU algorithm as inputs. During forward propagation, the input data is first downsampled by a factor of four and subsequently fed into the encoder. The encoder adopts a four-stage hierarchical structure based on LVT. At the initial stage, a Convolutional Self-Attention (CSA) module is used to dynamically extract local features, while higher stages employ Recursive Atrous Self-Attention (RASA) modules to efficiently capture multi-scale contextual information with fewer parameters, enhancing the network’s representational capacity. The decoder adopts the U-MixFormer structure, which takes multi-scale features from the encoder as query vectors (Q), and combines them with key (K) and value (V) vectors generated from fused multi-scale representations via a mixed-attention mechanism. This design enables efficient integration and progressive refinement of both local and global information during decoding. Finally, the decoder outputs are upsampled by a factor of four to restore the original

resolution, yielding high-precision predictions of fringe orders. In our implementation, the network predicts a continuous-valued fringe order, which is subsequently rounded to the nearest integer during post-processing.

The joint spatial-Fourier loss for fringe-order prediction is shown in Fig. 9b. In constructing the loss function, we first adopt the Mean Squared Error (MSE) loss, which focuses on pixel-wise prediction accuracy for TPU. However, as the fringe orders inherently exhibit a staircase distribution, they contain significant high-frequency components at the edges, reflecting essential physical information. Therefore, neglecting frequency-domain features might degrade the network’s performance in accurately reconstructing step transitions. To address this, we introduce a Fourier-domain consistency constraint into the loss function via a Fourier Loss term, guiding the network to learn frequency-domain characteristics of fringe orders during training. The final loss is formulated as a weighted combination of spatial-domain and frequency-domain consistency, simultaneously

ensuring pixel-wise accuracy and physical coherence, and significantly improving both precision and robustness in phase unwrapping. The loss function is expressed as

$$L_{\text{loss}} = \lambda_1 L_{\text{MSE}} + \lambda_2 L_{\text{Fourier}}, \quad (15)$$

where  $\lambda_1$  and  $\lambda_2$  represent the weights for different loss components. Specifically, the expression for  $L_{\text{MSE}}$  is given as

$$L_{\text{MSE}} = \frac{1}{HW} \sum_{n=1}^N \frac{(k_n^{\text{pred}}(x, y) - k_n^{\text{true}}(x, y))^2}{N}, \quad (16)$$

where  $k_n^{\text{pred}}(x, y)$  denotes the fringe orders predicted by the network for the  $n$ -th data sample in the training set, and  $k_n^{\text{true}}(x, y)$  denotes the corresponding ground-truth fringe orders.  $N$  represents the size of the training set, and  $H$  and  $W$  denote the height and width of the image, respectively. The  $L_{\text{Fourier}}$  term represents the Fourier Loss function, enforcing consistency between the frequency-domain values of the prediction and the ground-truth. The expression for  $L_{\text{Fourier}}$  is defined as

$$L_{\text{Fourier}} = \frac{\sum_{n=1}^N |\mathcal{F}(k_n^{\text{pred}}(x, y)) - \mathcal{F}(k_n^{\text{true}}(x, y))|}{N}, \quad (17)$$

where,  $\mathcal{F}(\ast)$  represents the Fourier transform operation. After training, by simply feeding the network with the high-frequency and low-frequency wrapped phases and the corresponding coarse fringe orders from any specific temporal phase unwrapping method (MF, MW, or NT), one can obtain high-quality fringe order predictions corresponding to the selected TPU algorithm.

## Data availability

The Blender scripts and some auxiliary code used for this study are available at <https://github.com/nominee/DP-TPU>.

## Code availability

The Blender scripts and some auxiliary code used for this study are available at <https://github.com/nominee/DP-TPU>.

Received: 17 May 2025; Accepted: 27 October 2025;

Published online: 03 December 2025

## References

- Baltsavias, E. P. A comparison between photogrammetry and laser scanning. *ISPRS J. Photogramm. Remote Sens.* **54**, 83–94 (1999).
- Ganapathi, V., Plagemann, C., Koller, D. & Thrun, S. Real time motion capture using a single time-of-flight camera. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 755–762 (IEEE, 2010).
- Scharstein, D. & Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **47**, 7–42 (2002).
- Leach, R. *Optical measurement of surface topography*, vol. 8 (Springer, 2011).
- Zhang, S. *Handbook of 3D machine vision: Optical metrology and imaging* (CRC Press, 2013).
- Lu, L. et al. Generative deep-learning-embedded asynchronous structured light for three-dimensional imaging. *Adv. Photonics* **6**, 046004–046004 (2024).
- Wu, Z. et al. Three-dimensional nanoscale reduced-angle ptychotomographic imaging with deep learning (rapid). *eLight* **3**, 7 (2023).
- Saba, A., Gigli, C., Ayoub, A. B. & Psaltis, D. Physics-informed neural networks for diffraction tomography. *Adv. Photonics* **4**, 066001–066001 (2022).
- Su, X. & Chen, W. Reliability-guided phase unwrapping algorithm: a review. *Opt. Lasers Eng.* **42**, 245–261 (2004).
- Goldstein, R. M., Zebker, H. A. & Werner, C. L. Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Sci.* **23**, 713–720 (1988).
- Lim, H., Xu, W. & Huang, X. Two new practical methods for phase unwrapping. In *1995 International Geoscience and Remote Sensing Symposium, IGARSS'95. Quantitative Remote Sensing for Science and Applications*, vol. 1, 196–198 (IEEE, 1995).
- Flynn, T. J. Two-dimensional phase unwrapping with minimum weighted discontinuity. *JOSAA* **14**, 2692–2701 (1997).
- Ghiglia, D. C. & Romero, L. A. Minimum lp-norm two-dimensional phase unwrapping. *JOSAA* **13**, 1999–2013 (1996).
- Zuo, C., Huang, L., Zhang, M., Chen, Q. & Asundi, A. Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review. *Opt. Lasers Eng.* **85**, 84–103 (2016).
- Tian, J., Peng, X. & Zhao, X. A generalized temporal phase unwrapping algorithm for three-dimensional profilometry. *Opt. Lasers Eng.* **46**, 336–342 (2008).
- Kinell, L. & Sjödal, M. Robustness of reduced temporal phase unwrapping in the measurement of shape. *Appl. Opt.* **40**, 2297–2303 (2001).
- Peng, X., Yang, Z. & Niu, H. Multi-resolution reconstruction of 3-d image with modified temporal unwrapping algorithm. *Opt. Commun.* **224**, 35–44 (2003).
- Wyant, J. Testing aspherics using two-wavelength holography. *Appl. Opt.* **10**, 2113–2118 (1971).
- Alcock, A. & Ramsden, S. Two wavelength interferometry of a laser-induced spark in air. *Appl. Phys. Lett.* **8**, 187–188 (1966).
- Polhemus, C. Two-wavelength interferometry. *Appl. Opt.* **12**, 2071–2074 (1973).
- Dändliker, R., Thalmann, R. & Prongué, D. Two-wavelength laser interferometry using superheterodyne detection. *Opt. Lett.* **13**, 339–341 (1988).
- Burke, J., Bothe, T., Osten, W. & Hess, C. F. Reverse engineering by fringe projection. In *Interferometry XI: Applications*, vol. 4778, 312–324 (SPIE, 2002).
- Towers, C. E., Towers, D. P. & Jones, J. D. Time efficient chinese remainder theorem algorithm for full-field fringe phase analysis in multi-wavelength interferometry. *Opt. Express* **12**, 1136–1143 (2004).
- Gushov, V. & Solodkin, Y. N. Automatic processing of fringe patterns in integer interferometers. *Opt. Lasers Eng.* **14**, 311–324 (1991).
- Wu, Z., Guo, W., Li, Y., Liu, Y. & Zhang, Q. High-speed and high-efficiency three-dimensional shape measurement based on gray-coded light. *Photonics Res.* **8**, 819–829 (2020).
- He, X., Zheng, D., Kemao, Q. & Christopoulos, G. Quaternary gray-code phase unwrapping for binary fringe projection profilometry. *Opt. Lasers Eng.* **121**, 358–368 (2019).
- Wang, Y., Liu, L., Wu, J., Chen, X. & Wang, Y. Spatial binary coding method for stripe-wise phase unwrapping. *Appl. Opt.* **59**, 4279–4285 (2020).
- Wu, H., Cao, Y., Dai, Y. & Zhang, H. Ultra-fast 3d imaging by a big codewords space division multiplexing binary coding. *Opt. Lett.* **48**, 2793–2796 (2023).
- Wu, H., Cao, Y., Dai, Y. & Wei, Z. Orthogonal spatial binary coding method for high-speed 3d measurement. *IEEE Transactions on Image Processing* (2024).
- Wang, Y. & Zhang, S. Novel phase-coding method for absolute phase retrieval. *Opt. Lett.* **37**, 2067–2069 (2012).
- Feng, S. et al. Fringe-pattern analysis with ensemble deep learning. *Adv. Photonics Nexus* **2**, 036010–036010 (2023).
- Feng, S., Zuo, C., Hu, Y., Li, Y. & Chen, Q. Deep-learning-based fringe-pattern analysis with uncertainty estimation. *Optica* **8**, 1507–1510 (2021).
- Li, X. et al. Adaptive structured-light 3d surface imaging with cross-domain learning. *Laser & Photonics Reviews* 2401609 (2025).
- Wang, K., Li, Y., Kemao, Q., Di, J. & Zhao, J. One-step robust deep learning phase unwrapping. *Opt. Express* **27**, 15100–15115 (2019).

35. Zhang, Z. et al. Efficient and robust phase unwrapping method based on sfnet. *Opt. Express* **32**, 15410–15432 (2024).
36. Yin, W. et al. Temporal phase unwrapping using deep learning. *Sci. Rep.* **9**, 20175 (2019).
37. Guo, X. et al. Unifying temporal phase unwrapping framework using deep learning. *Opt. Expr.* **31**, 16659–16675 (2023).
38. Liu, Y. et al. Multimodal adaptive temporal phase unwrapping using deep learning and physical priors. *APL Photonics* **10** (2025).
39. Li, Z. et al. Dual-frequency phase unwrapping based on deep learning driven by simulation dataset. *Opt. Lasers Eng.* **178**, 108168 (2024).
40. Singhal, P., Walambe, R., Ramanna, S. & Kotecha, K. Domain adaptation: challenges, methods, datasets, and applications. *IEEE access* **11**, 6973–7020 (2023).
41. Blender, O. Blender—a 3d modelling and rendering package. blender foundation, stichting blender foundation, amsterdam (2018).
42. Zhou, Q. & Jacobson, A. Thingi10k: A dataset of 10,000 3d-printing models. *arXiv* (2016).
43. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
44. Torr, P. H. & Zisserman, A. Mlesac: A new robust estimator with application to estimating image geometry. *Comput. Vis. image Underst.* **78**, 138–156 (2000).
45. Feng, S. et al. Fringe pattern analysis using deep learning. *Adv. Photonics* **1**, 025001–025001 (2019).
46. Kato, H. et al. Differentiable rendering: A survey. *CoRR* **2020**, abs/2006.12057 <https://arxiv.org/abs/2006.12057> (2020).
47. Zhao, H., Chen, W. & Tan, Y. Phase-unwrapping algorithm for the measurement of three-dimensional object shapes. *Appl. Opt.* **33**, 4497–4500 (1994).
48. Cheng, Y.-Y. & Wyant, J. C. Two-wavelength phase shifting interferometry. *Appl. Opt.* **23**, 4539–4543 (1984).
49. Takeda, M., Gu, Q., Kinoshita, M., Takai, H. & Takahashi, Y. 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).
50. Zuo, C. et al. 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).
51. Zhong, J. & Wang, M. Phase unwrapping by lookup table method: application to phase map with singular points. *Opt. Eng.* **38**, 2075–2080 (1999).
52. Liu, M., Fang, S., Dong, H. & Xu, C. Review of digital twin about concepts, technologies, and industrial applications. *J. Manuf. Syst.* **58**, 346–361 (2021).
53. Liu, X. et al. Digital twin modeling and controlling of optical power evolution enabling autonomous-driving optical networks: a bayesian approach. *Adv. Photonics* **6**, 026006–026006 (2024).
54. Slabaugh, G. G. Computing euler angles from a rotation matrix. *Retrieved August* **6**, 39–63 (1999).
55. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. pattern Anal. Mach. Intell.* **22**, 1330–1334 (2002).
56. Zhang, S. & Huang, P. S. Novel method for structured light system calibration. *Opt. Eng.* **45**, 083601–083601 (2006).
57. Zheng, Y., Wang, S., Li, Q. & Li, B. Fringe projection profilometry by conducting deep learning from its digital twin. *Opt. express* **28**, 36568–36583 (2020).
58. Wang, F., Wang, C. & Guan, Q. Single-shot fringe projection profilometry based on deep learning and computer graphics. *Opt. Express* **29**, 8024–8040 (2021).
59. Zhu, X., Zhang, Z., Hou, L., Song, L. & Wang, H. Light field structured light projection data generation with blender. In *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*, 1249–1253 (IEEE, 2022).
60. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
61. Yang, C. et al. Lite vision transformer with enhanced self-attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11998–12008 (2022).
62. Yeom, S.-K. & Von Klitzing, J. U-mixerformer: Unet-like transformer with mix-attention for efficient semantic segmentation. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1–10 (IEEE, 2025).

## Acknowledgements

This work was supported by National Key Research and Development Program of China (2022YFB2804603, 2022YFB2804605), National Natural Science Foundation of China (U21B2033, 62205147, 62522508, 62571249), Fundamental Research Funds for the Central Universities (2023102001, 2024202002), National Key Laboratory of Shock Wave and Detonation Physics (JCKYS2024212111), China Postdoctoral Science Fund (2023T160318), and Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (JSGP202105, JSGP202201).

## Author contributions

Y.L. provided the original idea. Y.L. and S.F. designed and performed the experiments. Y.L. analyzed the data. Y.L. prepared the figures. Y.L. wrote the manuscript. Z.J. and X.L. provided partial code support and suggestions for figure design. J.J. contributed partial data support. S.F., W.C., and S.Y. supervised the overall projects. All the authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44310-025-00096-z>.

**Correspondence** and requests for materials should be addressed to Shijie Feng.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025