

Supplementary Information for Deep-learning-enabled temporally super-resolved multiplexed fringe projection profilometry: high-speed kHz 3D imaging with low-speed camera

Wenwu Chen^{1,2,3}, Shijie Feng^{1,2,3,*}, Wei Yin^{1,2,3}, Yixuan Li^{1,2,3}, Jiaming Qian^{1,2,3}, Qian Chen^{3,*}, and Chao Zuo^{1,2,3,*}

¹Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu Province, China

²Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing 210019, Jiangsu Province, China

³Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu Province, China

*shijiefeng@njust.edu.cn

*chenqian@njust.edu.cn

*zuochao@njust.edu.cn

ABSTRACT

This document provides the information about the system synchronization, network architecture of MultiResUnet, training dataset generation, network training, loss functions design, system calibration and augmented 3D reconstruction, comparative study of different deep learning-based methods & different compression rates, discussion of motion blur, supplementary experiment of a running fascia gun to the main manuscript "Deep-learning-enabled temporally super-resolved multiplexed fringe projection profilometry: high-speed kHz 3D imaging with low-speed camera".

Contents

Supplementary Note 1. System synchronization

Supplementary Note 2. Network architecture of MultiResUnet

Supplementary Note 3. Training dataset generation

Supplementary Note 4. Network training

Supplementary Note 5. Loss functions design

Supplementary Note 6. System calibration and augmented 3D reconstruction (A3DR)

Supplementary Note 7. Comparative study of different deep learning-based methods & different compression rates

Supplementary Note 8. Discussion of motion blur

Supplementary Note 9. Supplementary experiment of a running fascia gun

Supplementary Note 1. System synchronization

We use a digital light processing (DLP) Discovery 4100 development kit from Texas Instruments as the basis of our projection system. The DLP kit includes a digital micro-mirror device (DMD) chip and a DMD driver board. By running the DLP projector in binary (1-bit) mode, quasi-sinusoidal fringe patterns can be created at high frame rates with lens defocusing and binary dithering techniques^{1,2}. With this technique, we drive the DMD at a refresh rate of 1,000 frame per second (fps), and the active time at 1,000 μs , as in **Fig. S1**. DLP development hardware is further used for precisely triggering to ensure the signal synchronization between the projector and the cameras. To capture multiplexed images, the projector outputs a trigger signal (Trigger1) for every nine frames (#1 to #9) to trigger the cameras. The cameras capture the multiplexed image by a long exposure time of 8,999.5 μs (the corresponding camera frame rate is ~ 111.11 Hz). In addition, for the preparation of training datasets, to generate the ground truth for deep neural networks 1 (DNN1) and DNN2, the cameras need to capture each separate fringe image and corresponding 12-step phase-shifting (PS)³ images (detailed in **Supplementary Note 3**). The projector was set to output a trigger signal (Trigger2) every one frame, while the camera frame rate is the same as the DMD at 1,000 Hz and the exposure time is 999.5 μs .

Moreover, to calibrate and verify the system synchronization, we use the projector to repeatedly project two complementary patterns. Inaccurate synchronization will lead to a crosstalk between two patterns in the captured image sequence. Then, a well-designed intensity detection algorithm can be used to validate the system synchronization. Details can be found in the reference⁴.

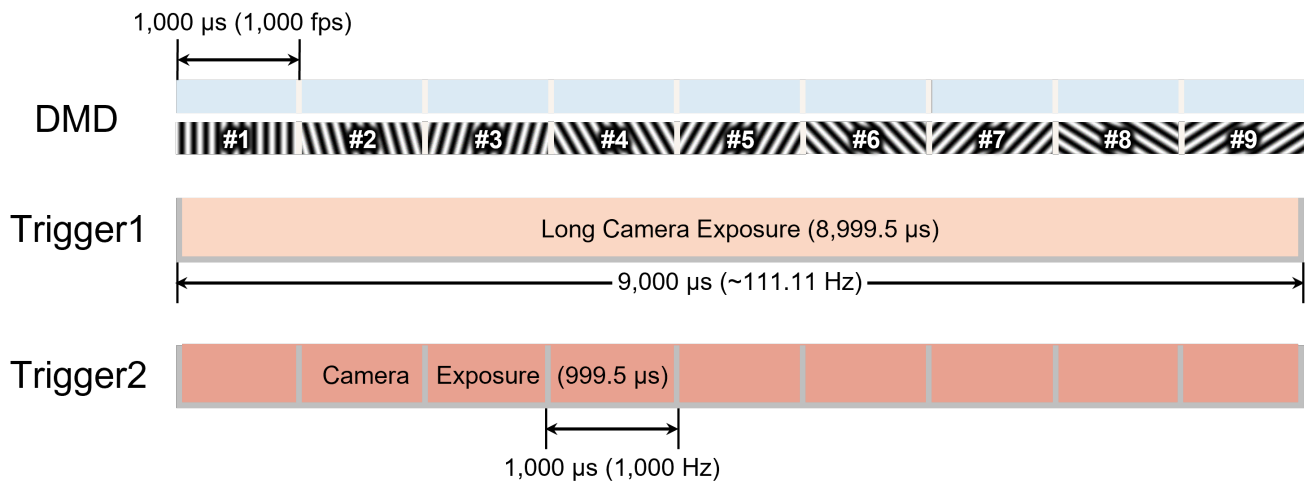


Figure S1. Timing diagram of different exposure and trigger settings (Trigger 1 and 2) of the DLMFPP system.

Supplementary Note 2. Network architecture of MultiResUnet

MultiResUnet is a novel architecture developed as the potential successor to the well-known U-Net framework^{5,6}. The diagram of MultiResUnet architecture is shown in **Fig. S2(a)**. In this model, the sequences of two convolutional layers in the classical U-Net architectures are replaced with MultiRes blocks, as shown in **Fig. S2(b)**. The number of filters is increased in the successive three layers gradually and a residual connection is added (along with 1×1 filter for conserving dimensions). The customized MultiRes block factorizes the bigger and more expensive 5×5 and 7×7 filters as a succession of 3×3 filters, thus having the advantage to reconcile features from different context size, save memory and speed up network training. In addition, instead of combining the features of encoders and decoders directly, residual path (Res-path) are constructed (shown in **Fig. S2(c)**), where several 1×1 and 3×3 convolutional layers along the shortcut connections (like in U-Net) are incorporated to alleviate the disparity between the encoder-decoder features.

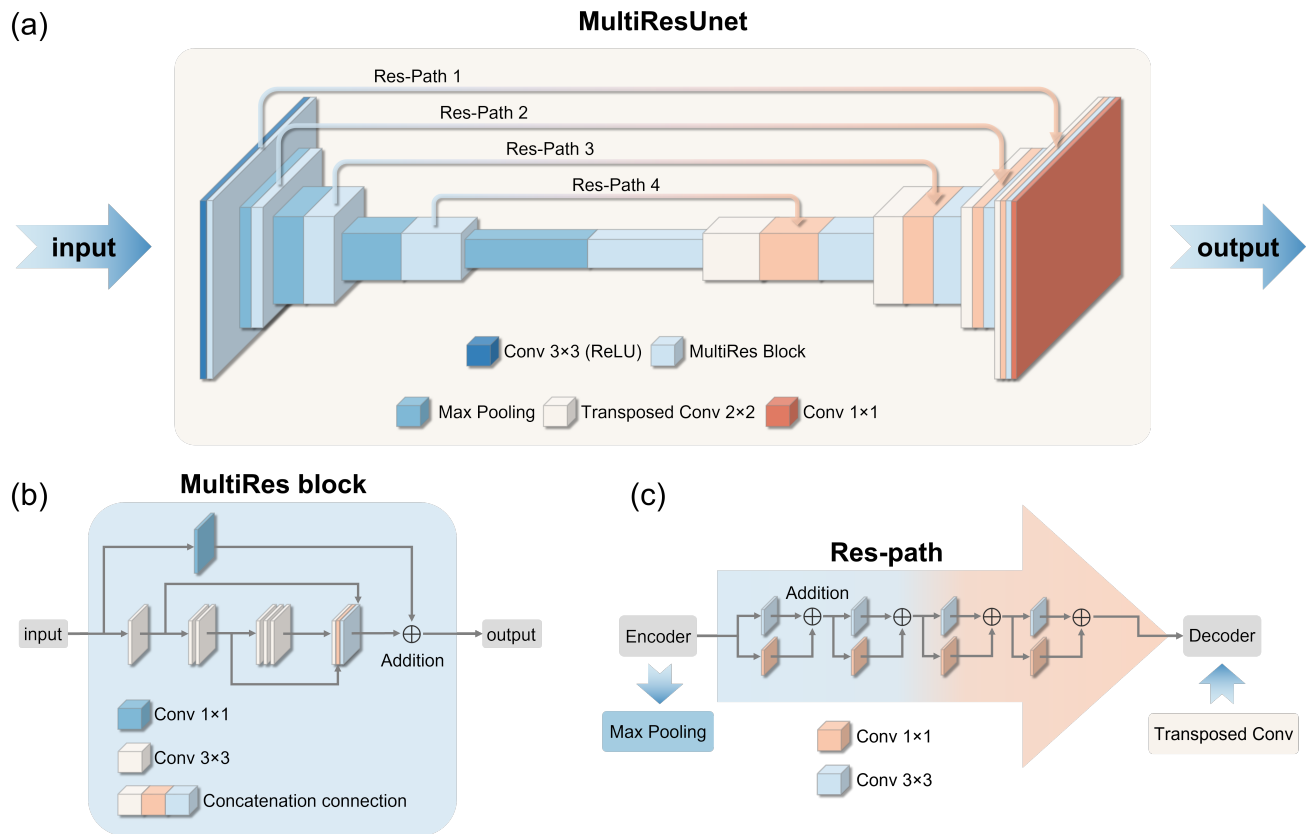


Figure S2. Diagram of the MultiResUnet. (a) Network architecture. (b) Structure of the MultiRes block. The customized block factorizes the bigger and more expensive 5×5 and 7×7 filters as a succession of 3×3 filters. (c) Structure of the Res-path, where several 1×1 and 3×3 convolutional layers along the shortcut connections are incorporated to alleviate the disparity between the encoder-decoder features.

Supplementary Note 3. Training dataset generation

To generate the dataset, the projector needs to sequentially project a series of fringe patterns as shown in **Fig. S3** to diverse static objects. In detail, step 1, the trigger mode of the camera is set to Trigger1. The projector sequentially projects fringe patterns I_m^p which can be represented as Eq. (1-3) of the main manuscript. In this work, the frequency f_x^p and f_y^p are selected as $f_x^p = 1/16$, $f_y^p = 1/12$, and θ is set to 14° . Then the camera captures the multiplexed image I_{LE} by a long exposure time. Step 2, the trigger mode of the camera is set to Trigger2. The projector sequentially projects nine sets of 12-step PS fringe patterns I_{mn}^p corresponding to I_m^p , which can be represented as

$$I_{mn}^p(x^p, y^p) = a^p + b^p \cos[\phi_m^p(x^p, y^p) - 2\pi n/N], \quad (S1)$$

where n represents the PS index $n = 0, 1, 2, \dots, N-1$, and in this work $N = 12$; m denotes the fringe pattern index $m = 1, 2, 3, \dots, M$, M is the total number of the patterns, and in this work $M = 9$. Specifically, I_{m0}^p is identical to I_m^p . In Trigger2 mode, the camera can capture separate fringe images I_m (i.e., I_{m0}) modulated by the object surface and their corresponding 12-step PS fringe images I_{mn} , which can be represented as:

$$I_m(x, y) = A_m(x, y) + B_m(x, y) \cos \phi_m(x, y), \quad (S2)$$

$$I_{mn}(x, y) = A_m(x, y) + B_m(x, y) \cos[\phi_m(x, y) - 2\pi n/N]. \quad (S3)$$

Then the numerator M_m and denominator D_m can be calculated by the following equation:

$$M_m(x, y) = \sum_{n=0}^{N-1} I_{mn}(x, y) \sin(2\pi n/N), \quad (S4)$$

$$D_m(x, y) = \sum_{n=0}^{N-1} I_{mn}(x, y) \cos(2\pi n/N). \quad (S5)$$

At the period of training dataset preparation, we photographed a variety of objects made of different materials (plastic, plaster, metal, ceramic, etc.) to generate diverse datasets, while a larger number of datasets is employed for the network training. In this work, 1,200 groups of images were captured, of which 800 groups were used for training and 400 groups for validation. This allows the network to see as many different scenes and conditions as possible, thereby improving the network's ability to generalize to various applications.

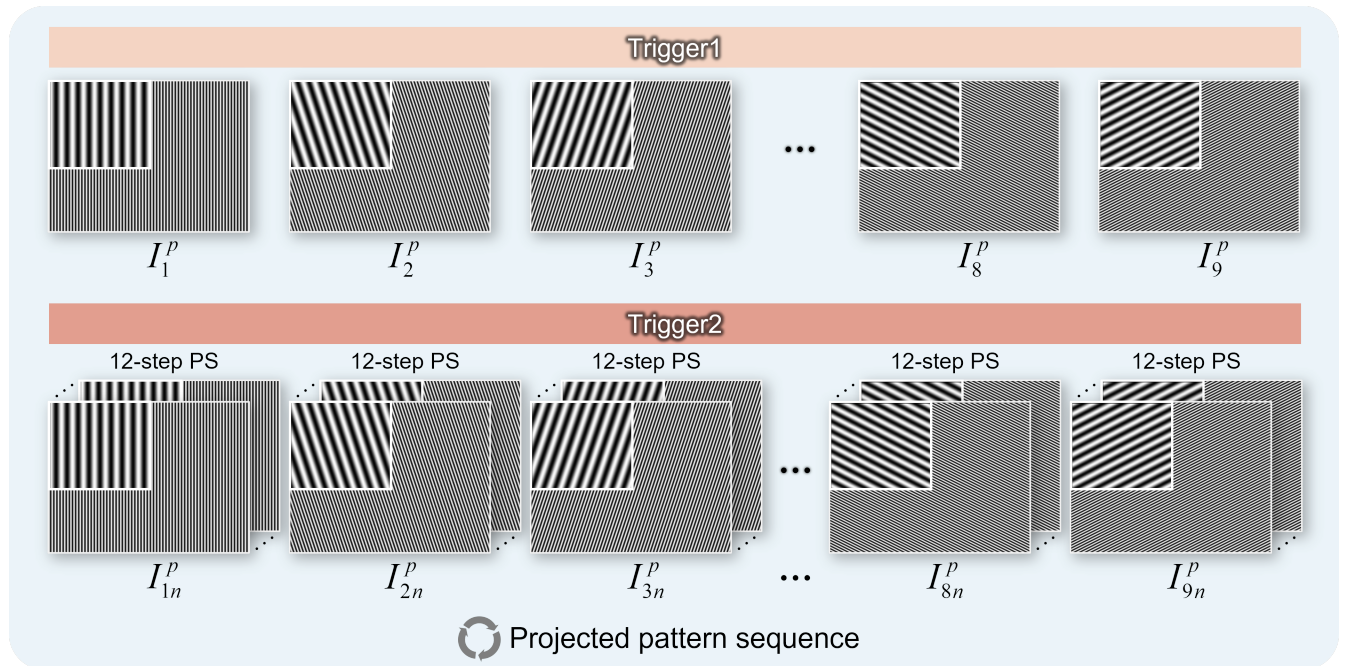


Figure S3. Projected pattern sequence consists of (1) separate fringe patterns I_m^P , $m = 1, 2, 3, \dots, 9$, in Trigger1 mode and (2) nine sets of 12-step PS fringe patterns I_{mn}^P corresponding to I_m^P , $n = 0, 1, 2, \dots, 11$, in Trigger2 mode.

Supplementary Note 4. Network training

Figure 2 in the main manuscript illustrates the flowchart of the proposed deep-learning-enabled multiplexed fringe projection profilometry (DLMFPP) method, where the spatial decomposing (SD), frequency decomposing (FD), and feature ensemble (FE) branches in DNN1 and the augmented fringe pattern analysis (AFPA) module in DNN2 are trained separately by supervised learning. The training of SD and FD branches is shown in **Fig. S4(a)**. Taking a plaster geometry as an example, the camera in Trigger1 mode captures the multiplexed image I_{LE} ($H \times W \times 1$) as the input to the SD branch. In Trigger2 mode, the camera captures separate fringe images $I_1 - I_9$ ($H \times W \times 9$) as the label. Then, I_{LE} is Fourier transformed (FT) to obtain the spectrum $\mathcal{F}_{LE}$. The real and imaginary components of $\mathcal{F}_{LE}$ are placed in two channels ($H \times W \times 2$) together as input of the FD branch⁷. Correspondingly, the same operation is performed for $I_1 - I_9$ to obtain the label $\mathcal{F}_1 - \mathcal{F}_9$ ($H \times W \times 18$). The training of the FE branch with the idea of ensemble learning⁸ is shown in **Fig. S4(b)**. The input consists of two parts: one part is the predicted separate fringe images $\hat{I}_1 - \hat{I}_9$ ($H \times W \times 9$) output from the trained SD branch; the other part is the inverse FT (iFT) of the spectrums predicted by the trained FD branch, that is $\hat{I}'_1 - \hat{I}'_9$ ($H \times W \times 9$). Finally, both two parts ($H \times W \times 18$) are fed together into the FE branch. The label of the FE branch is the same as that of the SD branch, i.e., the separate fringe images $I_1 - I_9$ ($H \times W \times 9$). The training of the AFPA module is shown in **Fig. S4(c)**. The AFPA module's input is a separate fringe image ($H \times W \times 1$) with any direction. For a better understanding, I_5 is taken as an example in **Fig. S4(c)**. The numerator term M_5 and denominator term D_5 are calculated from the 12-step PS fringe images I_{5n} and Eq. (S4-S5). M_5 and D_5 together ($H \times W \times 2$) are used as the label of AFPA.

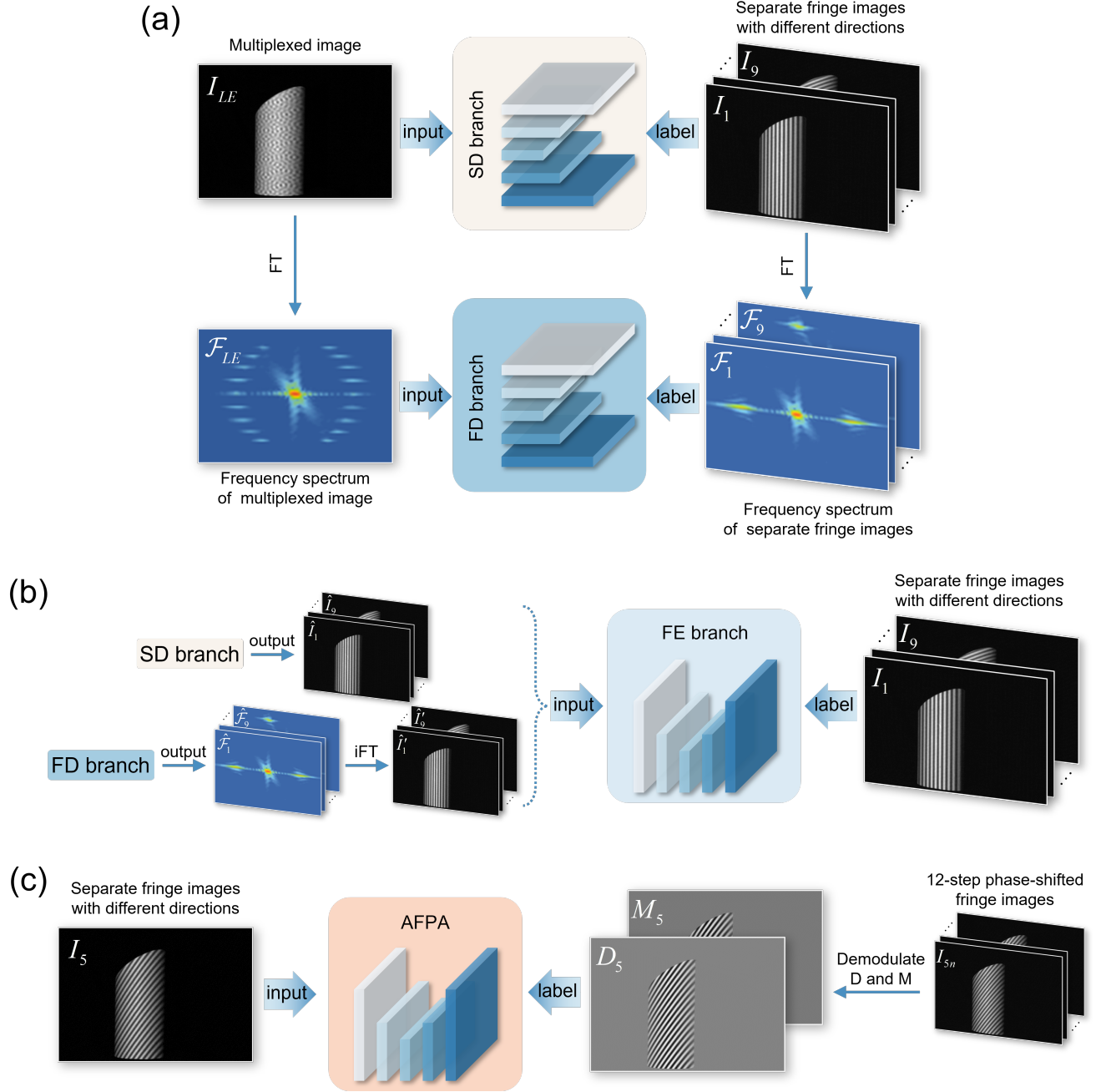


Figure S4. Network training of DLMFPP. (a) Training of SD and FD branches. In the SD branch, the input is the multiplexed image I_{LE} , and the label is separate fringe images $I_1 - I_9$. In the FD branch, the input is the spectrum $\mathcal{F}_{LE}$ corresponding to I_{LE} , and the label is spectrums $\mathcal{F}_1 - \mathcal{F}_9$ corresponding to $I_1 - I_9$. (b) Training of FE branch. The input consists of two parts: one part is the predicted separate fringe images $\hat{I}_1 - \hat{I}_9$ output from the trained SD branch; the other part is the iFT of the spectrums predicted by the trained FE branch, that is $\hat{\mathcal{F}}_1 - \hat{\mathcal{F}}_9$. The label is the same as that of the SD branch. (c) Training of AFPA module. The input is a separate fringe image with any direction, illustrated by I_5 , and the label is the numerator term M_5 and denominator term D_5 calculated from 12-step PS fringe images I_{5n} and Eq. (S4-S5).

Supplementary Note 5. Loss functions design

Loss functions are shown in **Fig. S5**. Specifically, the objective functions of the SD and FD branches are joint losses defined as:

$$Loss_{SD} = Loss_{SD}^{data} + \alpha Loss_{SD}^{physics} = \sum_{m=1}^9 \text{MSE}(\hat{I}_m, I_m) + \alpha \text{MSE}(\sum_{m=1}^9 \hat{I}_m, I_{LE}), \quad (\text{S6})$$

$$Loss_{FD} = Loss_{FD}^{data} + \alpha Loss_{FD}^{physics} = \sum_{m=1}^9 \text{MAE}(\hat{\mathcal{F}}_m, \mathcal{F}_m) + \alpha \text{MAE}(\sum_{m=1}^9 \hat{\mathcal{F}}_m, \mathcal{F}_{LE}), \quad (\text{S7})$$

where $\hat{I}_m$ and I_m denote the predicted and ground truth of m th separate fringe image, $m = 1, \dots, 9$, I_{LE} is the input multiplexed image, $\hat{\mathcal{F}}_m$, $\mathcal{F}_m$, and $\mathcal{F}_{LE}$ are the FT of $\hat{I}_m$, I_m , and I_{LE} , respectively, and α is the weighting scalar to balance the mean squared error (MSE) or mean absolute error (MAE) of the data-based loss $Loss_{SD/FD}^{data}$ and the physics-based loss $Loss_{SD/FD}^{physics}$. Differently, the objective functions in the FE branch and the AFPA module contain only the data-based loss, which are defined as:

$$Loss_{FE} = Loss_{FE}^{data} = \text{MSE}(\hat{I}_m, I_m), \quad (\text{S8})$$

$$Loss_{AFPA} = Loss_{AFPA}^{data} = \text{MSE}(\hat{Y}_m, Y_m), \quad (\text{S9})$$

where $\hat{Y}_m = (\hat{M}_m, \hat{D}_m)$ is the AFPA module's output and $Y_m = (M_m, D_m)$ is the ground truth obtained by 12-step PS.

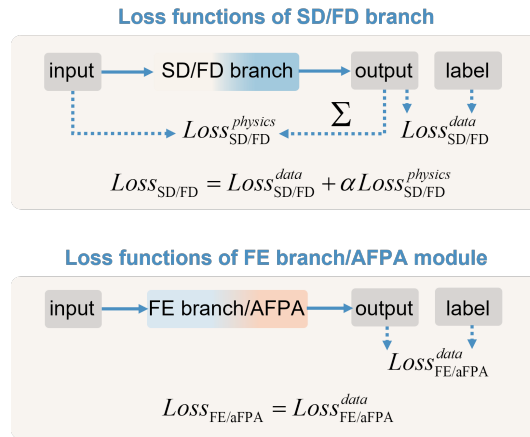


Figure S5. Loss functions. The SD and FD branches use joint losses containing data-based and physics-based loss, as defined as Eq. (S6)-(S7). The FE branch and AFPA module use only the data-based loss, as defined as Eq. (S8)-(S9).

Supplementary Note 6. System calibration and augmented 3D reconstruction

In the system calibration procedure, a calibration board with a white 9×11 circle array distributed uniformly on a black background is used. The distance between the centers of each circle is 25 mm. We capture 9 images of the calibration board with different poses to get a calibration volume of $\sim 400 \text{ mm} \times 280 \text{ mm} \times 200 \text{ mm}$. The projector is calibrated as an inverse camera with the help of the calibrated camera. We project two orthogonal sets of frequency-varying (fringe periods 1, 8, 64 in the x direction and 1, 8, 48 in the y direction) and PS (16-frame for the highest frequency) sinusoidal fringe patterns onto the calibration board. The absolute phase recovered by PS and multi-frequency temporal phase unwrapping techniques is used to establish a one-to-one mapping between a camera pixel and a DMD pixel. The intrinsic and extrinsic of the projector and camera are calibrated based on MATLAB Calibration toolbox, and optimized with bundle adjustment^{9,10} in which both measurement errors and the fabrication errors of the calibration board are taken into account. Once the projector and camera are calibrated in the same world coordinate system, the 3D world coordinates can be obtained.

However, in a horizontally configured fringe projection profilometry (FPP) system, the stereo phase unwrapping (SPU) and the mapping from phase to 3D coordinates are generally designed for vertical fringes. To cope with the case of arbitrarily oriented fringes in this work, we propose the augmented 3D reconstruction (A3DR) method. It is worth mentioning that in the process of SPU for arbitrarily oriented fringes, only the 3D reconstruction algorithm differs from the classical SPU methods¹¹. Thereby, A3DR can also solve the SPU problem for arbitrarily oriented fringes. The details of A3DR are as follows.

Assuming that all fringe orders are known, the final unwrapped phase image can be obtained. Then we can create an equation for every camera pixel (x^c, y^c) :

$$\phi_m(x^c, y^c) = 2\pi(f_x^p x^p \cos\theta_m + f_y^p y^p \sin\theta_m). \quad (\text{S10})$$

The relationship between the 3D world coordinates of the measured object (x^w, y^w, z^w) and the 2D camera pixel coordinates (x^c, y^c) can be described as:

$$\begin{aligned} s \begin{bmatrix} x^c \\ y^c \\ 1 \end{bmatrix} &= \mathbf{A}^c [\mathbf{R}^c, \mathbf{t}^c] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix} = \mathbf{P}^c \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix} \\ &= \begin{pmatrix} p_{11}^c & p_{12}^c & p_{13}^c & p_{14}^c \\ p_{21}^c & p_{22}^c & p_{23}^c & p_{24}^c \\ p_{31}^c & p_{32}^c & p_{33}^c & p_{34}^c \end{pmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}. \end{aligned} \quad (\text{S11})$$

The superscript c represents the camera and s is a scaling factor. $[\mathbf{R}^c, \mathbf{t}^c]$ is the 3×4 extrinsic parameter matrix of the rotation-translation from the world coordinate system to the camera coordinate system. $\mathbf{A}^c$ is the 3×3 intrinsic parameter matrix of the camera. $\mathbf{P}^c$ is the 3×4 perspective matrix, which is the

product of the intrinsic parameter matrix and the extrinsic parameter matrix. Considering the projector as an inverse camera model, a similar relationship can be established between the 3D world coordinates of the measured object (x^w, y^w, z^w) and the 2D projector pixel coordinates (x^p, y^p) :

$$\begin{aligned}
 s \begin{bmatrix} x^p \\ y^p \\ 1 \end{bmatrix} &= \mathbf{A}^p [\mathbf{R}^p, \mathbf{t}^p] \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix} = \mathbf{P}^p \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix} \\
 &= \begin{pmatrix} p_{11}^p & p_{12}^p & p_{13}^p & p_{14}^p \\ p_{21}^p & p_{22}^p & p_{23}^p & p_{24}^p \\ p_{31}^p & p_{32}^p & p_{33}^p & p_{34}^p \end{pmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix}.
 \end{aligned} \tag{S12}$$

The superscript p represents the projector, and the meaning of the parameters in Eq. (S12) are the same as those in Eq. (S11), but for the projector. If the projector and camera are calibrated in the same world coordinate system, the 3D world coordinates (x^w, y^w, z^w) of each camera pixel (x^c, y^c) can be obtained:

$$\begin{bmatrix} x^w \\ y^w \\ z^w \end{bmatrix} = \begin{pmatrix} p_{11}^c - p_{31}^c x^c & p_{12}^c - p_{32}^c x^c & p_{13}^c - p_{33}^c x^c \\ p_{21}^c - p_{31}^c y^c & p_{22}^c - p_{32}^c y^c & p_{23}^c - p_{33}^c y^c \\ p_{11}^p \cos \theta_m + p_{21}^p \alpha^p \sin \theta_m - p_{31}^p \beta^p & p_{12}^p \cos \theta_m + p_{22}^p \alpha^p \sin \theta_m - p_{32}^p \beta^p & p_{13}^p \cos \theta_m + p_{23}^p \alpha^p \sin \theta_m - p_{33}^p \beta^p \end{pmatrix}^{-1} \begin{bmatrix} p_{34}^c x^c - p_{14}^c \\ p_{34}^c y^c - p_{24}^c \\ p_{34}^p \beta^p - p_{14}^p \cos \theta_m - p_{24}^p \alpha^p \sin \theta_m \end{bmatrix}, \tag{S13}$$

where $\alpha^p = f_y^p / f_x^p$, $\beta^p = x^p \cos \theta_m + (f_y^p / f_x^p) y^p \sin \theta_m$. Since f_x^p , f_y^p , and θ_m are known (see Note 3), a unique correspondence value β^p for every camera pixel coordinate (x^c, y^c) can be established from Eq. (S10), and 3D reconstruction can be performed for every (x^c, y^c) from Eq. (S13).

Supplementary Note 7. Comparative study of different deep learning-based methods & different compression rates

To demonstrate the progressiveness of DLMFPP in achieving high-speed, high-resolution 3D imaging with low-speed cameras, we compare our proposed DLMFPP method with two state-of-the-art deep learning approaches^{12–15} based on the architectures of Multi-path convolutional neural network (CNN) and U-Net, respectively, in **Fig. S6**. **Figure S6(a)** shows a multiplexed image modulated by the scene and the corresponding spatial spectrum (locally zoomed in for better visibility). We captured the ground truths of separate fringe images by setting the camera frame rate to 1,000 Hz (same as the DMD refresh rate), which is shown in **Fig. S6(b)**. And the ground truths of phase in **Fig. S6(c)** were obtained by 12-step PS (refer to **Supplementary Note 3** for more details). **Figures. S6(d)–(f)** present the phase errors of Multi-path CNN, U-Net and DLMFPP, respectively. From **Figs. S6(d) and (e)**, we can see substantial phase errors on the regions with large height variations and sharp edges, with average MAEs of 0.5045 rad and 0.2022 rad in **(d)** and **(e)**, respectively. It is noticed that both of the state-of-the-art methods fail to handle regions with large height variations of the object. Moreover, these methods provide inconsistent phase recovery performance for fringe images with different directions (it seems that the accuracy is poor for near-vertical fringes and high for near-horizontal fringes). Different from the state-of-the-art approaches, DLMFPP divides the whole phase recovery process into two steps, which are multiplexed pattern decomposition (DNN1) and AFPA (DNN2). DNN1 incorporates the physical model of FT and the idea of ensemble learning⁸ to adaptively merge features from both the spatial and frequency domains, yielding high-quality reconstructed fringe images. Then, DNN2 introduces the physical model of PS¹⁶ to demodulate high-accuracy phase information from fringe images with different directions. As shown in **Fig. S6(f)**, DLMFPP achieves the lowest phase error with the average MAE of 0.0495 rad. The comparative study demonstrates that DLMFPP solves the dilemma of the state-of-the-art deep learning-based methods in handling regions with large height variations and providing consistent phase recovery performance for fringe images with different directions, revealing the superior performance from DLMFPP's advanced network design.

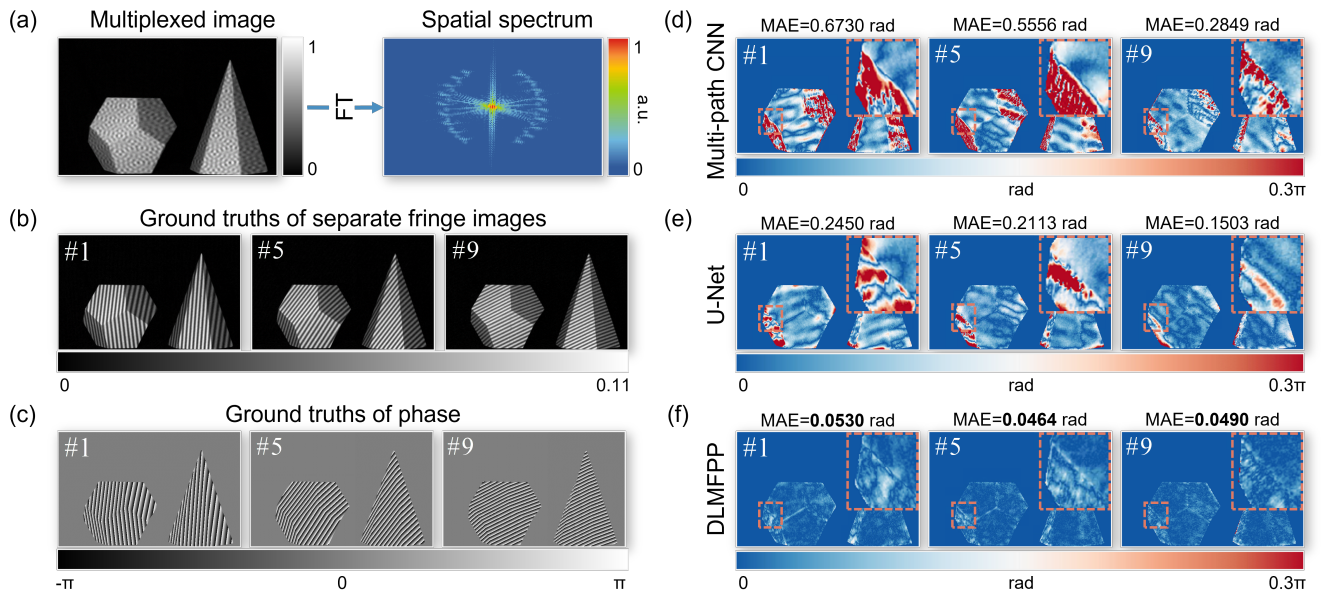


Figure S6. Comparative study of DLMFPP and two conventional end-to-end deep learning-based methods: (a) Multiplexed image modulated by the scene, and the corresponding spatial spectrum (locally zoomed in); (b) ground truths of separate fringe images, obtained by setting the camera frame rate same as the DMD refresh rate (1,000 Hz); (c) ground truths of phase, obtained by 12-step PS; (d-f) phase errors demodulated by Multi-path CNN, U-Net and DLMFPP, respectively.

The number of overlapping fringe images in a multiplexed image is referred as compression rate (CR) in this work. To determine the CR, we performed experiments with different CRs using the proposed DLMFPP. **Figure S7(a)** shows four multiplexed images, corresponding to CR = 1, 5, 9, and 13. The spatial spectrums of the multiplexed images are shown in **Fig. S7(b)** (locally zoomed in for better visibility). It can be seen that as the CR increases, the spectral aliasing of fundamental frequency components becomes more critical. **Figure S7(c)** plots the phase error of DLMFPP at different CRs. The plot shows that the phase error (characterized by the average MAE of demodulated phases) keeps increasing as the CR raises. It is easily noticed that there is a clear inflection point at CR = 9. As the CR further raises from 9 to 11, the slope of the line increases dramatically, and the phase error significantly raises from 0.0519 to 0.1085. Experiments show that the marginal benefit between CR and phase error is highest at CR = 9. Therefore, in this work, we overlap nine fringe images in a multiplexed image, which allows DLMFPP to achieve 9x temporal super-resolution while preserving spatial resolution. Practically, our approach is flexible to tradeoff temporal resolution and spatial resolution accuracy. If higher phase accuracy is required, the CR can be reduced appropriately, and vice versa.

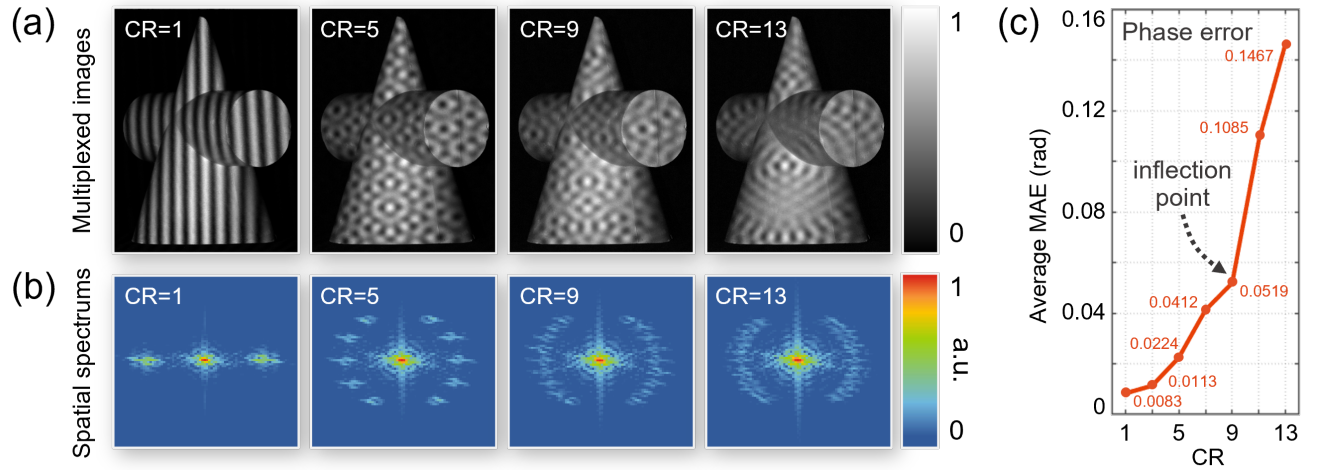


Figure S7. Comparative study of DLMFPP with different numbers of overlapping fringe images in a multiplexed image [the number is referred as compression rate (CR)]: (a) Multiplexed images, corresponding to CR = 1, 5, 9, and 13; (b) spatial spectrums of (a) (locally zoomed in); (c) phase errors of DLMFPP at different CRs.

Supplementary Note 8. Discussion of motion blur

High-speed imaging often faces the challenge of rapid dynamic scene changes and the motion blur caused by them. The proposed DLMFPP is no exception. Taking the measurement of a rotating fan (presented in Fig. 4 of the main manuscript) as an example, it can be clearly noticed in **Fig. S8(a)** that the multiplexed image I_{LE} directly captured by the camera (with an exposure time of 8,999.5 μs) suffers from an obvious motion blur caused by the rapidly rotating blades. However, this blurred image is actually generated by overlapping nine fringe patterns at different times. Due to the mechanism of acquiring long exposure multiplexed images first and then decomposing them in DLMFPP, the actual exposure time of separate fringe image is determined by the projection time of each pattern, not the camera exposure time. That is, motion blur in our method is not determined by the camera exposure time, but by the projection time, which is near-one-order of magnitude-lower than the exposure time of a single camera frame. Therefore, when the multiplexed image is decomposed by the proposed multiplexed pattern decomposition module (DNN1), we can obtain nine separate fringe images free of motion blur [I_1, I_3, I_5, I_7 and I_9 shown in **Fig. S8(b)**]. The exposure time of each fringe image is approximately 1/9 (999.5 μs) of that of the multiplexed image, and this greatly reduced actual exposure time effectively handles the challenges of motion blur of rotating blades, thus ensuring accurate 3D reconstruction [see Fig. 4(d) in the main manuscript].

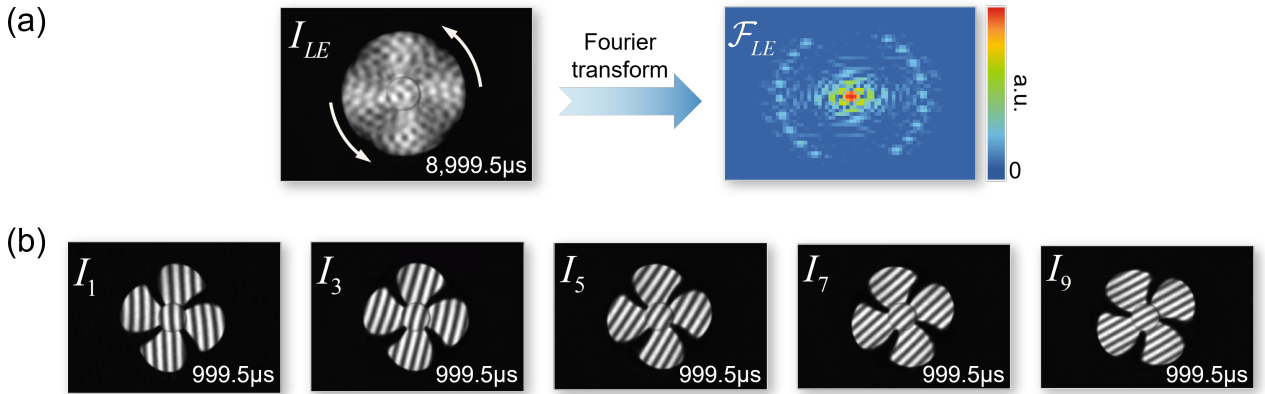


Figure S8. Motion blur in the measurement of a rotating fan. (a) The multiplexed image I_{LE} with an exposure time of 8,999.5 μs suffers from an obvious motion blur caused by the rapidly rotating blades, and corresponding spectrum $\mathcal{F}_{LE}$ (locally zoomed in). (b) Five decomposed separate fringe images free of motion blur (I_1, I_3, I_5, I_7 and I_9) with an exposure time of 999.5 μs . This greatly reduced exposure time effectively handles the challenges of motion blur of rotating blades.

Specific techniques contribute to the robust performance of DLMFPP is the physical priori, i.e., the fringes with different directions are separable in the frequency domain. The separability can be referred to $\mathcal{F}_{LE}$ in **Fig. S8(a)**. It is this separable nature that guarantees the physical feasibility of decomposing multiplexed image using deep learning. Moreover, we introduce the physical models of FT and PS method¹⁶ in DLMFPP to integrate features from both the spatial and frequency domains with the idea of ensemble learning⁸, which further contribute to the high accuracy and robust performance of DLMFPP in different scenarios.

Supplementary Note 9. Supplementary experiment of a running fascia gun

We applied DLMFPP to image a running commercially available fascia gun. **Figure S9(a)** shows the multiplexed image I_{LE} captured by the ~ 111.11 Hz camera (in Trigger1 mode) and the corresponding spectrum $\mathcal{F}_{LE}$ (locally zoomed for better visibility). The point A on the gun is chosen to demonstrate the cyclic displacement of the fascia gun head. The z (out-of-plane) direction displacement at the point A location is plotted as a function of time over a 90 ms period in **Fig. S9(b)**. The plot shows that the cyclic movement of the gun head has a period of about 35 ms, which corresponds to a speed of 1,714 rotations per minute (rpm) of the rotary motor inside the fascia gun. **Figure S9(c)** presents three fringe images (I_1 , I_5 and I_9 , corresponding to $T = 0, 4$, and 8 ms) decoded from the multiplexed image I_{LE} and the corresponding 3D model reconstructed by the proposed method. The 3D reconstruction of the whole dynamic process can be found in Supplementary Movie S3.

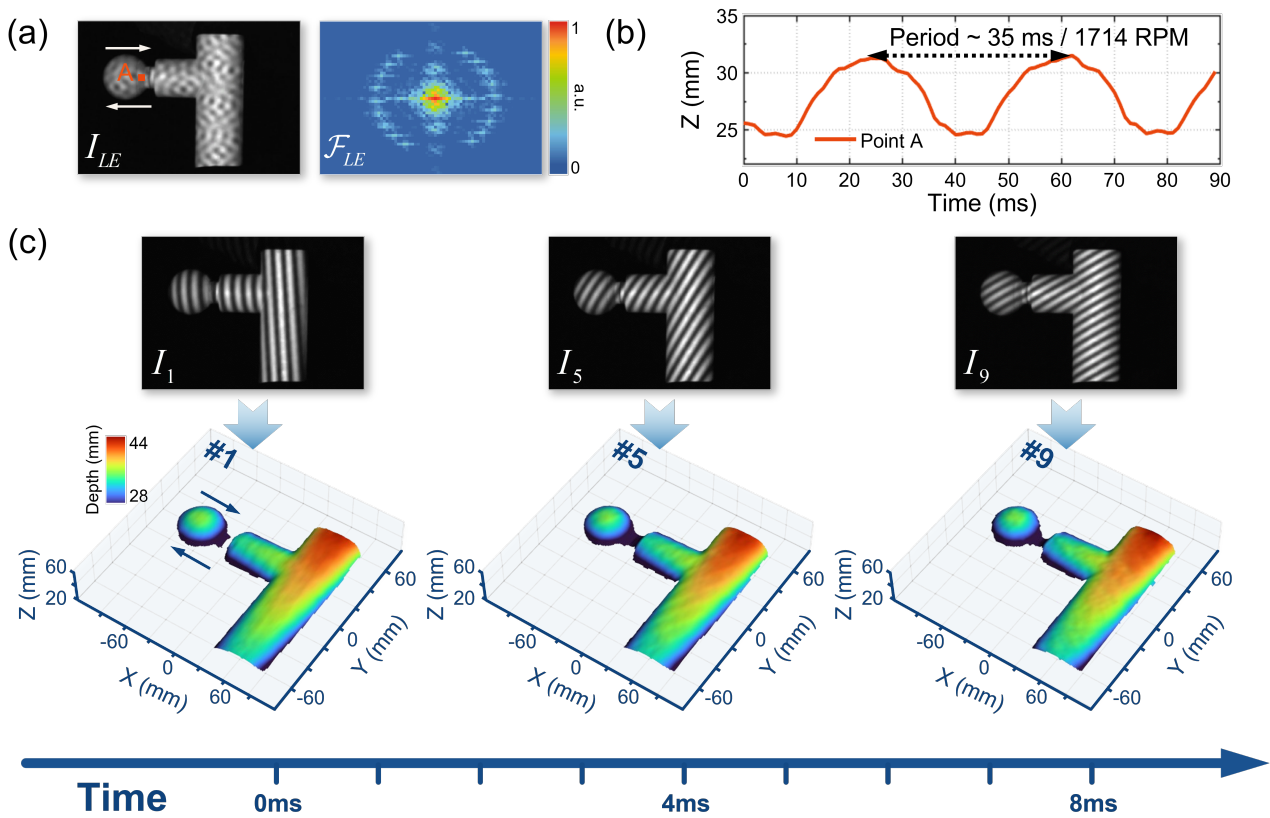


Figure S9. Measurement of a running fascia gun. (a) The multiplexed image I_{LE} and the corresponding spectrum $\mathcal{F}_{LE}$ (locally zoomed). (b) Displacement in the z direction at a chosen point location [A shown in (a)] as a function of time over a 90 ms period. (c) Three fringe images (I_1 , I_5 and I_9 , corresponding to $T = 0, 4$, and 8 ms) decoded from I_{LE} , and the corresponding 3D model reconstructed by the proposed method.

References

1. Li, B., Wang, Y., Dai, J., Lohry, W. & Zhang, S. Some recent advances on superfast 3d shape measurement with digital binary defocusing techniques. *Opt. Lasers Eng.* **54**, 236–246 (2014).
2. Zuo, C. *et al.* Optimized pulse width modulation pattern strategy for three-dimensional profilometry with projector defocusing. *Appl. optics* **51**, 4477–4490 (2012).
3. Zuo, C. *et al.* Phase shifting algorithms for fringe projection profilometry: A review. *Opt. Lasers Eng.* **109**, 23–59 (2018).
4. Zuo, C. *et al.* Micro fourier transform profilometry (μ ftp): 3d shape measurement at 10,000 frames per second. *Opt. Lasers Eng.* **102**, 70–91 (2018).
5. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, 234–241 (Springer, 2015).
6. Ibtehaz, N. & Rahman, M. S. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks* **121**, 74–87 (2020).
7. Yin, W. *et al.* Physics-informed deep learning for fringe pattern analysis. *Opto-Electronic Adv.* **7**, 230034–1 (2024).
8. Feng, S. *et al.* Fringe-pattern analysis with ensemble deep learning. *Adv. Photonics Nexus* **2**, 036010–036010 (2023).
9. Triggs, B., Zisserman, A. & Szeliski, R. Vision algorithms: Theory and practice: International workshop on vision algorithms corfu, greece, september 21–22, 1999 proceedings. In *International Workshop on Vision Algorithms* (1999).
10. Huang, L., Zhang, Q. & Asundi, A. Camera calibration with active phase target: improvement on feature detection and optimization. *Opt. letters* **38**, 1446–1448 (2013).
11. Tao, T. *et al.* Real-time 3-D shape measurement with composite phase-shifting fringes and multi-view system. *Opt. Express* **24**, 20253–20269 (2016).
12. Qian, J. *et al.* Single-shot absolute 3D shape measurement with deep-learning-based color fringe projection profilometry. *Opt. Lett.* **45**, 1842–1845 (2020).
13. Qian, J. *et al.* Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement. *Apl Photonics* **5**, 046105 (2020).
14. Li, Y., Qian, J., Feng, S., Chen, Q. & Zuo, C. Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement. *Opto-Electronic Adv.* **5**, 210021 (2022).
15. Li, Y., Qian, J., Feng, S., Chen, Q. & Zuo, C. Composite fringe projection deep learning profilometry for single-shot absolute 3D shape measurement. *Opt. Express* **30**, 3424–3442 (2022).

16. Feng, S. *et al.* Fringe pattern analysis using deep learning. *Adv. Photonics* **1**, 1 (2019).