

Article

Single-Shot Three-Dimensional Measurement by Fringe Analysis Network

Mingzhu Wan ^{1,*}, Lingbao Kong ^{1,*}  and Xing Peng ^{2,3,4} ¹ Shanghai Engineering Research Center of Ultra-Precision Optical Manufacturing, School of Information Science and Technology, Fudan University, Shanghai 200438, China² College of Intelligent Science and Technology, National University of Defense Technology, Changsha 410073, China³ Hunan Provincial Key Laboratory of Ultra-Precision Machining Technology, Changsha 410073, China⁴ Laboratory of Science and Technology on Integrated Logistics Support, Changsha 410073, China

* Correspondence: lkong@fudan.edu.cn

Abstract: Fringe projection profilometry (FPP) has been broadly applied in three-dimensional (3D) measurements, but the existing multi-shot methods, which mostly utilize phase-shifting techniques, are heavily affected by the disturbance of vibration and cannot be used in dynamic scenes. In this work, a single-shot 3D measurement method using a deep neural network named the Fringe Analysis Network (FrANet) is proposed. The FrANet is composed of a phase retrieval subnetwork, phase unwrapping subnetwork, and refinement subnetwork. The combination of multiple subnetworks can help to recover long-range information that is missing for a single U-Net. A two-stage training strategy in which the FrANet network is pre-trained using fringe pattern reprojection and fine-tuned using ground truth phase maps is designed. Such a training strategy lowers the number of ground truth phase maps in the data set, saves time during data collection, and maintains the accuracy of supervised methods in real-world setups. Experimental studies were carried out on a setup FPP system. In the test set, the mean absolute error (MAE) of the refined absolute phase maps was 0.0114 rad, and the root mean square error (RMSE) of the 3D reconstruction results was 0.67 mm. The accuracy of the proposed method in dynamic scenes was evaluated by measuring moving standard spheres. The measurement of the sphere diameter maintained a high accuracy of 84 μm at a speed of 0.759 m/s. Two-stage training only requires 8800 fringe images in data acquisition, while supervised methods require 96,000 fringe images for the same number of iterations. Ablation studies verified the effectiveness of two training stages and three subnetworks. The proposed method achieved accurate single-shot 3D measurements comparable to those obtained using supervised methods and has a high data efficiency. This enables the accurate 3D shape measurement of moving or vibrating objects in industrial manufacturing and allows for further exploration of network architecture and training strategy with few training samples for single-shot 3D measurement.

Keywords: 3D measurement; fringe projection; structured light; deep learning; single-shot



Citation: Wan, M.; Kong, L.; Peng, X. Single-Shot Three-Dimensional Measurement by Fringe Analysis Network. *Photonics* **2023**, *10*, 417. <https://doi.org/10.3390/photonics10040417>

Received: 23 February 2023

Revised: 3 April 2023

Accepted: 5 April 2023

Published: 7 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Three-dimensional (3D) optical measurement has been broadly applied in various fields, such as biomedicine, computer vision, and industrial manufacturing, due to its non-contact feature, low cost, and high efficiency. Structured light techniques are the most popular 3D measurement methods, due to their accuracy and versatility. Fringe projection profilometry (FPP) utilizes a structured light system to project sinusoidal fringe patterns onto the measured objects and capture modulated patterns to calculate the height maps. Most fringe analysis methods retrieve the phase from a set of phase-shifting fringe patterns and perform temporal phase unwrapping using multi-frequency fringe patterns [1,2].

Research on FPP has mostly focused on the improvement of phase-shifting techniques and temporal phase unwrapping. Yu et al. [3] proposed a 3D measurement method based

on the unequal-period combination of shifting Gray code and dual-frequency phase-shifting fringes. This method adopts a strategy of additionally projecting a set of low-frequency phase shifting fringe patterns to effectively correct the period jump error. The experimental results show that the proposed method can effectively reduce the period jump error, and can measure multiple isolated objects, as well as objects with drastic changes in surface height, such as plaster heads. Peng et al. [4] produced sinusoidal fringes for high-speed three-dimensional shape measurement using a phase-shifting algorithm. In this method, the sinusoidal fringes are generated by expanding the inverted image of a filled binary sinusoidal pattern in a specified direction, and the phase shift is realized by switching the backlight sources of a set of binary sinusoidal patterns formed on a slide. Hu et al. [5] adopted a multi-frequency phase-shifting scheme for the microscopic 3D measurement of shiny surfaces. The multi-frequency phase-shifting scheme can improve the integrity of the final phase map of the shiny surface, on which basis a complete and high-accuracy 3D reconstruction combined with a microscopic telecentric stereo system can be achieved. Wu et al. [6] utilized a phase-shifting profilometry (PSP) to realize temporal phase unwrapping simultaneously with the lowest number of fringe patterns. This method does not use other additional structural patterns to achieve full field phase unwrapping and can measure discontinuous surfaces or multiple isolated objects simultaneously. Li et al. [7] proposed a high-accuracy temporal phase unwrapping method based on super-grayscale multi-frequency grating projection by using the time-division multiplexing characteristic of a projector and the integral characteristic of a CCD camera. In this method, the super-grayscale grating is designed instead of the traditional 256 grayscale multi-frequency gratings to reduce the digital error. Li et al. [8] adopted a dynamic 3D reconstruction framework based on a modified three-wavelength phase unwrapping algorithm and phase error compensation method to acquire a sufficient number of 3D results, ensure the continuity of the dynamic 3D shape measurement, and reduce the phase error and measurement error introduced by object motion. Li et al. [9] used redundant data for self-correction in multi-frequency phase unwrapping. Pistellato et al. [10] proposed to eliminate phase ambiguity using probabilistic consensus where phase values were modeled from Wrapped Gaussian distribution. Lilienblum and Michaelis [11] derived a phase unwrapping method by applying pattern sequences and reduced calculation errors caused by discontinuity, occlusion, and reflection.

As deep learning techniques evolve, the deep convolutional neural network (CNN) has been applied to 3D measurements. Qi et al. [12] proposed an absolute phase measurement method with limited patterns. The proposed method combines the object reflectivity correction and the half-period gray-coded phase unwrapping algorithm and can obtain a large number of codewords for a fringe order without reducing the intensity level of each stair. Yao et al. [13] first proposed a multi-purpose neural network combined with code-based patterns to recover the absolute phase. The multi-purpose network can learn the principle of extracting absolute phase from a small number of patterns and greatly decrease the number of patterns with a high accuracy.

These abovementioned methods require multi-shot phase-shifting fringe images in multiple frequencies. Therefore, 3D measurements based on these methods are time-consuming and inaccurate when applied to dynamic scenes. Although multi-shot methods provide highly accurate measurement data for static objects by capturing multiple fringe pattern images, the accuracy may be decreased due to the disturbance from vibration and movement between the gaps in the image shots, rendering these methods vulnerable to temporal noises.

Single-shot methods extract phase maps from one fringe image and are robust against movement. Thus, single-shot methods are desirable for dynamic 3D measurement. Traditional single-shot 3D measurement methods utilize spatial demodulation, including Fourier transform profilometry (FTP), windowed Fourier transform profilometry (WFTP), and wavelet transform profilometry (WTP) [14–16]. Single-shot methods using specifically designed patterns have also been proposed. In addition to the methods with complex

patterns, Kawasaki et al. [17] utilized a simple grid pattern to achieve dense shape reconstruction. Single-shot methods based on deep learning are more accurate and robust than traditional methods, thus being more viable for practical applications. These methods were first introduced into FPP to retrieve phase maps. Feng et al. [18] demonstrated that deep learning can improve the accuracy of phase demodulation from a single fringe pattern. They used twelve-step phase-shifting techniques to generate ground truth and trained two convolutional neural networks (CNNs) for fringe analysis, where CNN1 predicted the background image and CNN2 predicted the numerator and denominator for phase calculation. Their results were superior to those of FTP and WFTP. Qiao et al. [19] presented a single-shot phase retrieval method that can reconstruct the phase distribution of specular objects by using deep learning. The deep networks are built on ideas of deep separable convolution and inverted residual. The method can estimate result closer to ground truth and effectively retain details on the measured surface. Yu et al. [20] designed the FPTNet for fringe transformation rather than setting phase maps as the output. The network only requires a single fringe as the input, and the FPTNet transformed one fringe image into multiple phase-shifting fringe images, and phase retrieval was achieved through phase-shifting techniques. Nguyen et al. [21] integrated a fringe-to-phase network with single-shot FPP to achieve 3D reconstructions. The proposed fringe-to-phase network has an architecture similar to that of U-Net and can directly retrieve three wrapped phase maps from a color image comprising three fringe patterns with designated frequencies. Zhang et al. [22] designed a convolutional neural network to accurately extract phase information in both the low signal-to-noise ratio (SNR) and saturation situations and increased the dynamic range of 3D measurement. Qian et al. [23] proposed a single-shot absolute 3D shape measurement method with deep-learning-based color FPP. Through learning on extensive data sets, the trained neural network can predict a high-resolution, motion-artifact-free, and crosstalk-free absolute phase directly from a single-color fringe image. Qian et al. [24] presented deep-learning-enabled geometric constraints and a phase unwrapping approach for the single shot absolute 3D shape measurement. This method generated more accurate absolute phase maps than spatial phase unwrapping.

End-to-end networks can also directly predict height maps from a single-shot fringe image. Van der Jeugh and Dirckx [25] designed an end-to-end neural network for single-shot measurement. They randomly generated a large number of height maps and collected fringe images using simulated fringe projection. These images, with the corresponding height maps as ground truth, compose the data set for network training. Nguyen et al. [26] adopted a four-step phase-shifting technique to produce ground truth height maps using a real-world FPP system. The input of the technique is a single fringe-pattern image, and the output is the corresponding depth map for 3D shape reconstruction. They compared different network architectures, including FCN, AEN, and U-Net, in terms of performance. U-Net obtained the most impressive results on their fringe projection data set. Machineni et al. [27] introduced an end-to-end deep-learning-based framework for FPP that does not require any frequency domain filtering or phase unwrapping. The framework reconstructs the depth profile of the object from the deformed fringe itself through multi-resolution similarity evaluation using a convolution neural network. Nguyen et al. [28] presented a 3D shape reconstruction technique that employs an end-to-end deep convolutional neural network to transform a single speckle-pattern image into its corresponding 3D point cloud. The deep network predicted the height map from the single-shot speckle image using the ground truth height map measured by FPP.

However, deep learning methods usually require large data sets to facilitate training. Data insufficiency results in overfitting and hinders performance. As in the case of FPP, the commonly used supervised learning methods require the ground truths of height maps or phase maps for training, which are measured using phase-shifting and temporal phase unwrapping. To produce training samples with accurate ground truth, twelve-step phase-shifting techniques with fringe patterns at four frequencies were adopted in some research, in which 48 fringe images were captured for one training sample. This is a

highly time-consuming process and is inconvenient for practical applications. To generate more training samples, previous works utilized computer graphics to generate synthetic data sets for supervised training or re-project fringe patterns for unsupervised training. Zheng et al. [29] proposed constructing a digital twin of the FPP system and conducting virtual scanning, which generated 7200 fringe images and 800 corresponding 3D scenes in 1.5 h. Wang et al. [30] presented a single-shot fringe projection profilometry method based on deep learning and computer graphics. They built a virtual FPP system and tested different parameters to construct a sufficient data set. To estimate the depth image from only one fringe image, a new loss function was designed and two network architectures, U-Net, and pix2pix, were compared. Fan et al. [31] developed an unsupervised training method for 3D reconstruction with dual-frequency fringe projection profilometry that does not require ground truth in the training set. They reprojected fringe patterns using the height maps predicted by the network and calculated the loss between the reprojected fringe images and real fringe images. The methods mentioned above that use computer graphics to generate synthetic data sets are less time-consuming in the data set preparation stage but cannot avoid accuracy loss when applied to real-world FPP systems. Synthetic data sets or reprojection processes are unable to account for various textures and noises in real-world fringe images. Thus, the domain discrepancy has a negative influence on the network performance.

To solve the dependence on static scenes and the time-consuming property of multi-shot methods, as well as the high data requirements of single-shot methods based on deep learning, in this work, we propose a single-shot 3D measurement method using a deep neural network named Fringe Analysis Network (FrANet). The FrANet consists of three subnetworks for fringe analysis, including the phase retrieval subnetwork, phase unwrapping subnetwork, and refinement subnetwork, replacing phase retrieval and phase unwrapping in traditional FPP. Phase unwrapping or height map prediction that integrates phase unwrapping in it is regarded as an ill-posed task that requires long-range information for the accurate prediction of absolute phase maps. U-Net is capable of extracting features and recovering the image resolution, thus being efficient in processing high-resolution images. However, generic CNN or a single U-Net cannot process long-range information due to the limitation of the small receptive field. Therefore, instead of using a single U-Net, in this work, we adopt two subnetworks to conduct phase unwrapping, where the phase unwrapping subnetwork with additional layers extracts long-range information, and the refinement subnetwork provides further refinement. Specifically, the phase retrieval subnetwork extracts wrapped phase maps from single-shot fringe images. The phase unwrapping subnetwork analyzes predicted wrapped phase maps and fringe images to yield absolute phase maps. The refinement subnetwork takes the wrapped phase maps, fringe images, and the primary prediction of absolute phase maps as inputs. Refined absolute phase maps are the final output of the FrANet. To solve the problems of overfitting and poor performance caused by insufficient samples, the FrANet is pre-trained using an unsupervised data set with fringe pattern reprojection and fine-tuned using a supervised data set with ground truth phase maps. Such a training strategy lowers the number of ground truth phase maps in the data set, saves time during data collection, and maintains the accuracy of supervised methods in real-world setups.

The contribution can be summarized as follows: A single-shot 3D measurement method based on deep learning is proposed, which enables accurate 3D measurements in dynamic scenes; A deep network named the FrANet with three subnetworks for fringe analysis is designed to improve the accuracy of 3D measurements; A two-stage training strategy for the FrANet is developed to reduce the number of supervised samples, which enables the efficient deployment of the proposed method in practical applications.

The rest of this paper is organized as follows: Section 2 describes the detailed architecture and training strategy of the FrANet; Section 3 presents the analysis of the experiments and the results; Section 4 presents the discussion, followed by the conclusion in Section 5.

2. Fringe Analysis Network for Single-Shot 3D Measurement

The proposed method utilizes the Fringe Analysis Network (FrANet) for single-shot 3D measurements. The FrANet consists of a phase retrieval subnetwork, phase unwrapping subnetwork, and refinement subnetwork. The phase unwrapping and height map prediction, which essentially involves phase unwrapping, are ill-posed tasks that require long-range information to remove phase ambiguity. A single U-Net cannot guarantee the acquisition of long-range information. The combination of multiple subnetworks can solve this problem by providing the phase unwrapping subnetwork with a large receptive field and the refinement subnetwork with a small receptive field, which renders both long-range and short-range information accessible. All subnetworks have an improved U-Net architecture. The phase retrieval subnetwork extracts the wrapped phase map from the single-shot fringe image. The phase unwrapping subnetwork generates the absolute phase map from the wrapped phase map and fringe image. The refinement subnetwork analyzes the wrapped phase map, fringe images, and the primary prediction of the absolute phase map to output the final refined absolute phase map. To solve the problem of insufficient samples for supervision training, in this work, we designed a two-stage training strategy where the FrANet is pre-trained using fringe pattern reprojection and fine-tuned using ground truth phase maps. The well-trained FrANet can process single-shot fringe images and predict accurate absolute phase maps, which can be converted into coordinates of measured objects for 3D reconstruction. We elaborate on the details of the network architecture and the training strategy in the following subsections.

2.1. Network Architecture

The FrANet replaces phase retrieval and phase unwrapping in traditional FPP. An architecture diagram of the FrANet is shown in Figure 1. The FrANet consists of three subnetworks, namely, the phase retrieval subnetwork, the phase unwrapping subnetwork, and the refinement subnetwork. All the subnetworks adopt an improved U-Net architecture. U-Net features an encoder–decoder structure, which increases the receptive field through down-sampling and recovers the resolution of the input through up-sampling. The previous layers are concatenated to up-sampled layers of the same resolution in order to avoid information loss in the up-sampling process. U-Net has also been proven effective in fringe analysis. Most fringe analysis networks adopt one U-Net or generic CNN to fulfill a task, such as phase retrieval, phase unwrapping, or height map prediction. Among these tasks, phase unwrapping is regarded as an ill-posed problem and requires long-range information to remove phase ambiguity. Height map prediction also integrates phase unwrapping into end-to-end networks. Thus, instead of using a single subnetwork, in this work, we adopted three subnetworks to conduct fringe analysis, where the phase retrieval subnetwork extracts the wrapped phase maps from single-shot fringe images, and the phase unwrapping subnetwork analyzes the predicted wrapped phase maps and fringe images to yield absolute phase maps. Finally, the refinement subnetwork takes wrapped phase maps, fringe images, and the primary prediction of the absolute phase maps as inputs. Refined absolute phase maps are the collective output of the FrANet.

The subnetwork used in the FrANet adopts an improved U-Net architecture, as shown in Figure 2. In the encoder part, down-sampling layers (orange) and convolutional layers (dark blue) are applied to extract features of the input at low resolutions in order to increase the computation efficiency. The down-sampling layers consist of convolutions with kernel size $k = 2$ and stride $s = 2$, batch normalization, and leaky ReLU as the activation function. The convolutional layers adopt convolutions with kernel size $k = 3$ and stride $s = 1$, batch normalization, and leaky ReLU. The numbers of filters are set to be identical for each subnetwork at the same resolution, being 16, 24, 24, 32, 32, and 48, corresponding to 1 , $1/2$, $1/4$, $1/8$, $1/16$, and $1/32$ of the input resolution. The residual blocks operate at the lowest resolution using four convolutional layers with skip connections. The number of filters is 48 for these layers. In the decoder part, the up-sampling layers (pink) utilize transposed convolutions with kernel size $k = 2$ and stride $s = 2$ to recover the features at a

high resolution, followed by batch normalization and leaky ReLU. The number of filters for each resolution are 16, 24, 24, 16, and 8 from left to right. The up-sampled features are concatenated to the output of the convolutional layers at the same resolution in the encoder part. After that, concatenated layers (bright blue) that include convolutions with kernel size $k = 1$ and stride $s = 1$, batch normalization, and leaky ReLU are designed to decrease the number of channels. These layers are followed by convolutional layers to conduct further feature extraction. The output layer (green) uses convolution and sigmoid functions to convert the features into the output dimension and value range of each subnetwork.

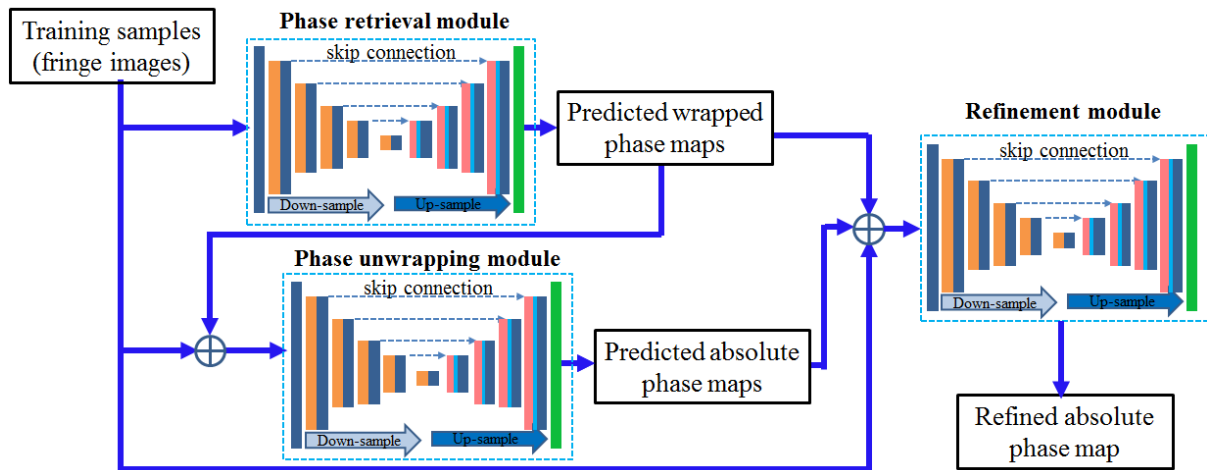


Figure 1. Architecture diagram of the FrANet.

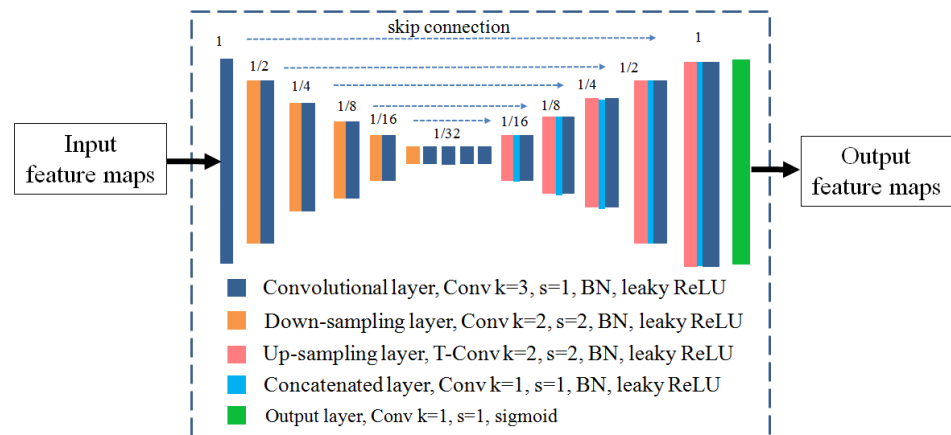


Figure 2. Improved U-Net architecture used in the FrANet.

In addition, to improve the accuracy of single-shot 3D measurements by FrANet, some special improvements were made according to the functions of each subnetwork. Since phase retrieval does not require long-range information to the same extent as phase unwrapping, residual blocks are removed from the phase retrieval subnetwork. The phase unwrapping subnetwork generates an absolute phase map which requires a large receptive field. Additional layers are added to the phase unwrapping subnetwork so as to down-sample the image into a $1/256$ input resolution, and the number of filters is kept at 48. The refinement subnetwork takes absolute phase maps, whose values vary between different image coordinates, as part of the input. Thus, batch normalization is omitted for all the convolutions of the refinement subnetwork to maintain the value scale.

2.2. Training Strategy

Deep networks are known for the high demand for training data, and the insufficient number of training samples leads to overfitting and poor network performance.

Reducing the complexity of training data sets with insufficient data could boost the performance of deep networks [32,33]. However, in the case of fringe analysis, critical information is contained in the intensity of fringe images, and it is extremely hard to lower the image complexity.

In this work, a two-stage training method combining unsupervised training and supervised training was designed to maintain the accuracy of the FrANet for single-shot 3D measurements while reducing the cost of obtaining supervised samples. Two-stage training has been applied in computer vision [34] and natural language processing [35]. Networks are pre-trained using self-supervised or supervised learning on large data sets and fine tuned using supervised learning on small data sets to complete different tasks or the same tasks in different domains. In the case of FPP, there has been little emphasis on two-stage training for deep neural networks. Therefore, here, a two-stage training method was designed to tackle the data requirement problem of deep learning methods. The two stages are an unsupervised pre-training stage followed by a supervised fine-tuning stage. Unsupervised training based on fringe pattern reprojection is used to calculate the loss between predicted and real single-shot fringe images. Supervised training utilizes phase shifting and temporal phase unwrapping to generate real phase maps as ground truth. Most training samples belong to the unsupervised data set without ground truth, being easier and faster to acquire. The supervised data set only contains a few samples with which to finetune the network with prior knowledge. This training strategy reduces the time required for data acquisition and bridges the gap between re-projected images and real-world fringe images to ensure high accuracy. The proposed two-stage training method can be viewed as a semi-supervised [36] approach that conducts training using both labeled and unlabeled data. In computer vision, a common semi-supervised method utilizes deep networks to produce pseudolabels for unlabeled data, which improves the network performance trained on labeled data only. Another semi-supervised method adopts an autoencoder for representation learning so that unlabeled data can also contribute. In contrast to these methods, we simply adopted two-stage training without pseudolabels or autoencoders.

2.2.1. Unsupervised Training

An overview of the unsupervised training process of the FrANet is shown in Figure 3. In the unsupervised training stage, fringe patterns at two frequencies are projected onto the measured objects, and fringe images are captured using a real-world FPP system. A large number of images are collected to form the unsupervised data set, which meets the data requirements of deep learning algorithms and avoids overfitting in training. Since unsupervised training does not rely on real phase maps or height maps as ground truth and only needs two fringe images for each training sample, which are single-shot fringe images without phase shifting at each frequency. Thus, time-consuming phase-shifting techniques are avoided in data collection, and the memory required for data storage is reduced.

During training and testing, only high-frequency fringe images are taken as the input of the FrANet. The network output, including predicted wrapped and absolute phase maps, is used for fringe reprojection. The loss between real and re-projected fringe images at both frequencies is calculated, and the network parameters are updated. High-frequency fringe images are adopted to guide the network into generating accurate phase maps. On the other hand, low-frequency fringe images are necessary for the evaluation of absolute phase maps and ensuring proper phase unwrapping.

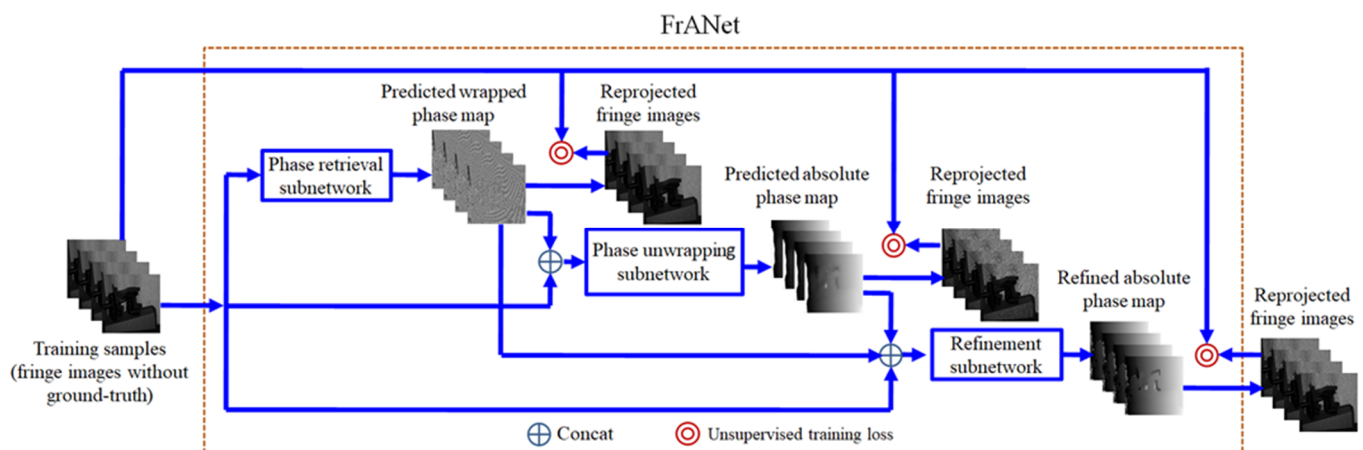


Figure 3. Overview of the unsupervised training process of the FrANet.

In FPP, captured fringe patterns can be expressed as:

$$I_f(x, y) = A(x, y) + B(x, y) \cos[\varphi_f(x, y)] = A(x, y) + B(x, y) \cos[\Phi_f(x, y)], \quad (1)$$

where A is the background intensity, B is the fringe amplitude, (x, y) is the coordinates of the fringe image, φ is the wrapped phase, Φ is the absolute phase, and f is the frequency. The background intensity and fringe amplitude of each pixel are simply estimated using maximum and minimum intensity values within a 21×21 window in the high-frequency fringe image. Since the unsupervised learning process only aims to pre-train the FrANet, the accuracy loss due to the reprojection error caused by this estimation will be compensated after supervised finetuning. Having determined the wrapped and absolute phase maps of high-frequency fringe images predicted by the network, fringe patterns can be reprojected using Equation (1). Absolute phase maps of high-frequency and low-frequency fringe images have the following relationship:

$$\frac{\Phi_h(x, y)}{f_h} = \frac{\Phi_l(x, y)}{f_l}, \quad (2)$$

where h is the high frequency and l is the low frequency. Low-frequency fringe images are reprojected using Equations (1) and (2) to ensure proper phase unwrapping. In contrast to the method developed by Fan et al. [31], our FrANet predicts phase maps instead of height maps, and thus, the configuration of the FPP system is not required in fringe pattern reprojection.

After the reprojection process, the unsupervised loss between the real and re-projected fringe images is calculated:

$$L_{\text{unsup}} = L(I_h, \bar{I}_h(\bar{\varphi}_h)) + L(I_h, \bar{I}_h(\bar{\Phi}_h)) + L(I_h, \bar{I}_h(\bar{\Phi}_h^r)) + L(I_l, \bar{I}_l(\bar{\Phi}_l)) + L(I_l, \bar{I}_l(\bar{\Phi}_l^r)), \quad (3)$$

where L denotes the $L1$ loss function, $\bar{I}$ is the reprojected fringe image, $\bar{\varphi}$ is the predicted wrapped phase map, and $\bar{\Phi}$ and $\bar{\Phi}^r$ are the predicted primary and refined absolute phase maps, respectively. Since the gradient scales of the five loss terms are similar, we apply the same weight to each loss term. The FrANet is trained using unsupervised loss until convergence is achieved.

The pseudocode for training the FrANet using an unsupervised data set is given in Algorithm 1.

Algorithm 1. The pseudocode for training the FrANet using an unsupervised data set.

Input: Unsupervised data set X , max_epoch n , batch size m ;

Output: Updated parameters of the FrANet θ (weight W , and bias b);

1. Initialize parameters of the FrANet θ (weight W , and bias b) randomly;
2. **for** $epoch = 1$ **to** n **do**
3. Split unsupervised data set into N/m groups $\{X_1, X_2, \dots, X_{N/m}\}$ randomly;
4. **for** $t = 1$ **to** N/m **do**
5. Use a group of single-shot high-frequency fringe images from X_t as network input;
6. Predict the wrapped phase map $\varphi_h^{-(t)}$ using phase retrieval subnetwork;
7. Predict the absolute phase map $\Phi_h^{-(t)}$ using phase unwrapping subnetwork;
8. Refine the absolute phase map $\Phi_h^{-(t)}$ using refinement subnetwork;
9. Calculate the absolute phase map $\Phi_l^{-(t)}, \Phi_l^{-r(t)}$ for low-frequency fringe images;
10. Reproject high – frequency fringe image $I_h^{-(t)}(\varphi_h^{-(t)})$ using $\varphi_h^{-(t)}$;
11. Reproject high – frequency fringe image $I_h^{-(t)}(\Phi_h^{-(t)})$ using $\Phi_h^{-(t)}$;
12. Reproject high – frequency fringe image $I_h^{-(t)}(\Phi_h^{-(t)})$ using $\Phi_h^{-(t)}$;
13. Reproject low – frequency fringe image $I_l^{-(t)}(\Phi_l^{-(t)})$ using $\Phi_l^{-(t)}$;
14. Reproject low – frequency fringe image $I_l^{-(t)}(\Phi_l^{-(t)})$ using $\Phi_l^{-(t)}$;
15. Compute the unsupervised network loss L_{unsup} ;
16. Update θ by using unsupervised loss and **Adam** gradient descent algorithm;
17. **end(for)**
18. **end(for)**
19. **return** θ (weight W , and bias b)

2.2.2. Supervised Training

An overview of the supervised training process of the FrANet is shown in Figure 4. The supervised training stage starts after the network is pre-trained using the unsupervised data set. Supervised training enables the network to directly learn from ground truth phase maps, which are prepared using phase-shifting profilometry. Twelve-step phase shifting fringe patterns are projected onto measured objects at four frequencies, also using the real-world FPP system. These 48 fringe images are captured to obtain a ground truth phase map of one training sample. Compared to fully supervised methods, fewer training samples are needed to finetune the network with prior knowledge. Thus, less time is required to collect additional fringe images for the computation of ground truth phase maps.

Object-height-modulated phase shifting patterns can be described as:

$$I_n(x, y) = A(x, y) + B(x, y) \cos[\varphi(x, y) + \frac{2\pi n}{N}], \quad (4)$$

where n denotes the n -th phase shifting pattern, and N is the number of phase-shifting steps. The wrapped phase can be retrieved using the following equation:

$$\varphi(x, y) = \arctan \frac{\sum_{n=0}^{N-1} I_n(x, y) \sin(\frac{2\pi n}{N})}{\sum_{n=0}^{N-1} I_n(x, y) \cos(\frac{2\pi n}{N})}. \quad (5)$$

After temporal phase unwrapping, absolute phase maps are obtained. The wrapped phase maps and absolute phase maps at the highest frequency are used as the ground truth of training samples. Supervised loss is computed as:

$$L_{\text{sup}} = L(\varphi, \bar{\varphi}) + L(\Phi, \bar{\Phi}) + L(\Phi, \bar{\Phi}^r). \quad (6)$$

After two training stages, the fringe analysis network FrANet can predict absolute phase maps from single-shot fringe images and achieve 3D measurement through triangulation.

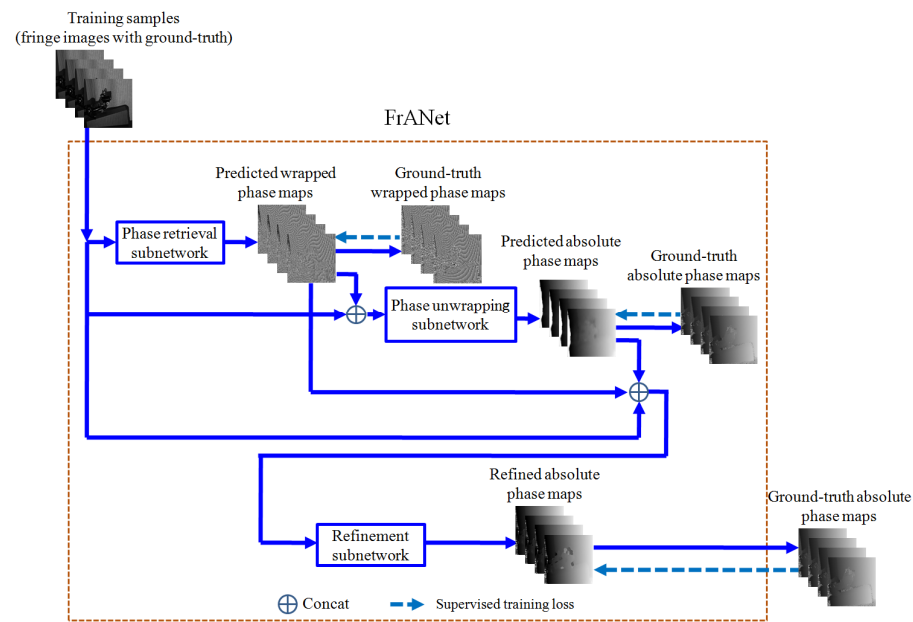


Figure 4. Overview of the supervised training process of the FrANet.

The pseudocode for training the FrANet using a supervised data set is given in Algorithm 2.

Algorithm 2. The pseudocode for training the FrANet using the supervised data set.

Input: Supervised data set X , max_epoch n , batch size m , parameters of the FrANet θ (weight W , and bias b) after unsupervised training;

Output: Updated parameters of the FrANet θ (weight W , and bias b);

1. Initialize parameters of the FrANet θ (weight W , and bias b) using unsupervised training result;
 2. Calculate the ground truth wrapped phase map φ ;
 3. Calculate the ground truth absolute phase map Φ ;
 4. **for** $epoch = 1$ to n **do**
 5. Split supervised data set to N/m groups $\{X_1, X_2, \dots, X_{N/m}\}$ randomly;
 6. **for** $t = 1$ to N/m **do**
 7. Use a group of single-shot fringe images from X_t as network input;
 8. Predict the wrapped phase map $\bar{\varphi}^{(t)}$ using phase retrieval subnetwork;
 9. Predict the absolute phase map $\bar{\Phi}^{(t)}$ using phase unwrapping subnetwork;
 10. Refine the absolute phase map $\bar{\Phi}^{r(t)}$ using refinement subnetwork;
 11. Compute the supervised network loss L_{sup} ;
 12. Update θ by using supervised loss and **Adam** gradient descent algorithm;
 13. **end(for)**
 14. **end(for)**
 15. **return** θ (weight W , and bias b)
-

3. Experiments

To verify the effectiveness of the proposed method, an FPP system consisting of a blue light projector from TengJu and digital cameras from DaHeng Mercury was built. The system is shown in Figure 5a. Only the left camera was adopted in the experiments. Figure 5b,c displays the samples for 3D measurement, including some automotive parts. The industrial part in Figure 5b belongs to the test set for evaluation, while the part in Figure 5c is used in the two-stage training. The schematic of the experimental setup is shown in Figure 6. The camera and projector are connected to the computer for coding fringe patterns and triggering data acquisition. The camera captures images of a resolution of 1296×964 pixels with three channels, and the resolution of the projector is 1280×720 pixels. During data acquisition, the projector projects vertical blue fringe patterns onto the measured objects. The air temperature during the experiments was around 20 degrees Celsius. In the unsupervised training stage, 2000 groups of fringe images are collected. Each group contains fringe images of the same scene at two frequencies. The frequency is defined as the entire number of fringes in the projected fringe pattern. The high frequency and low frequency were set to 64 and 9, respectively. In the supervised training stage, only 120 groups of twelve-step phase-shifting fringe images were captured at four frequencies of 64, 16, 4, and 1. The test set with 20 groups of fringe images was split from the supervised data set. In the following sections, the implementation details of the fringe analysis network FrANet are first provided. Then, the accuracy of the results predicted by the FrANet network is evaluated, the ablation studies on the training stages and FrANet architecture is undertaken, and lastly, the data efficiency of the proposed method is analyzed.

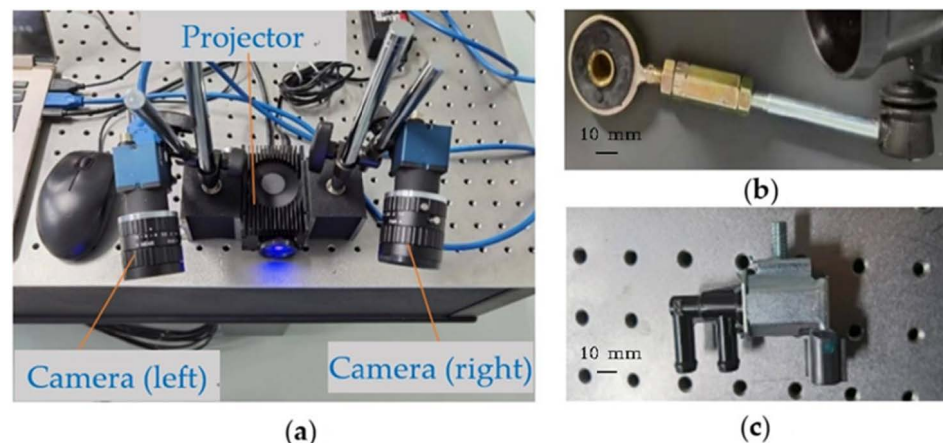


Figure 5. FPP system and industrial parts used for 3D measurement: (a) the FPP system built for the experiments; (b) an industrial part for 3D measurements in the test set; (c) an industrial part for 3D measurements in the training set.

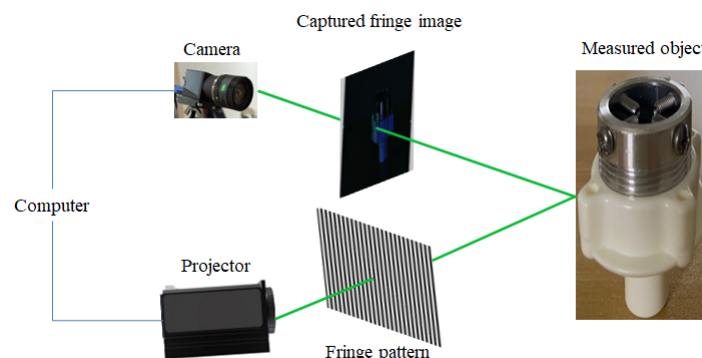
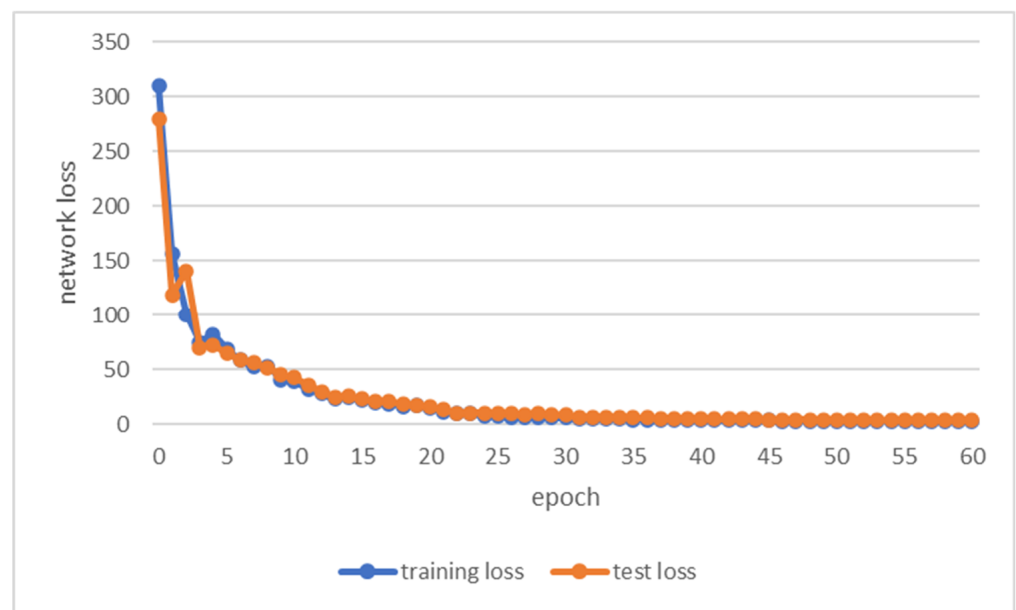


Figure 6. Schematic of the experimental setup.

3.1. Implementation Details

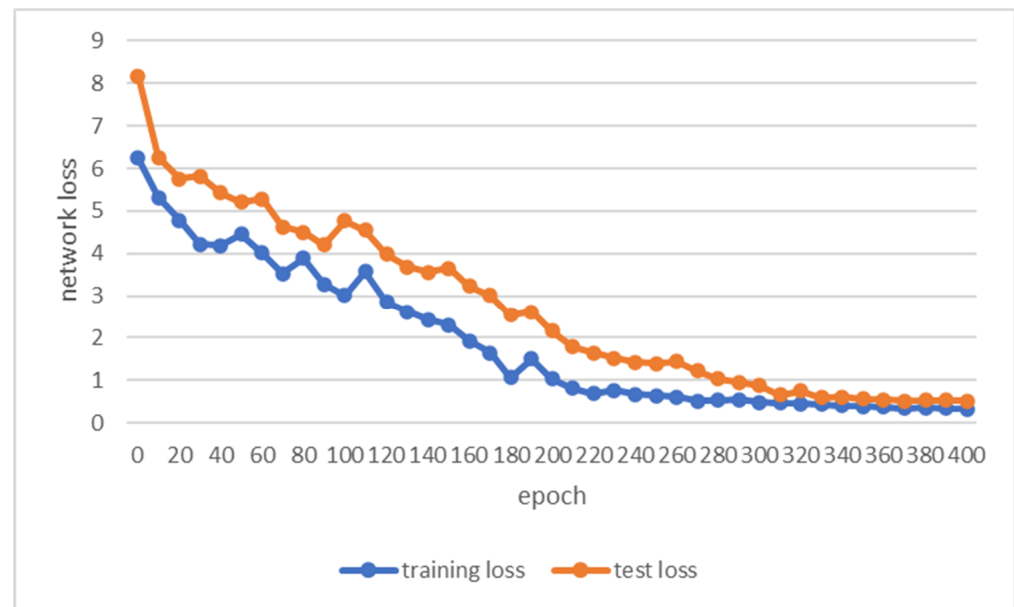
The proposed method was implemented with PyTorch, and the device used for training is one RTX 2070 SUPER GPU. The Adam optimizer was used with the parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. These parameters were set as default since default parameters are generally adopted in deep network optimization except in rare cases, for instance, Generative Adversarial Network with the parameter $\beta_1 = 0.5$. The Adam optimizer is not as strict with the initial learning rate as the SGD optimizer. The initial learning rate was selected from 1×10^{-2} , 1×10^{-3} , 1×10^{-4} , and 1×10^{-5} according to the results after one epoch of training. In addition, the learning rate was adjusted when the network loss stopped decreasing after five epochs of training. During pre-training, the initial learning rate was set to 1×10^{-3} , and was decreased to 1×10^{-4} after 30 epochs. The whole pre-training process took 60 epochs. The images were cropped to a size of 1280×768 . During finetuning, the initial learning rate was initially set to 1×10^{-4} for 200 epochs and then 5×10^{-5} for another 200 epochs. The batch size of two and multiple data augmentations were adopted. After training, the FrANet network was used for single-shot measurements with a run time of 0.87 s using the same setup. All deep networks used for comparison had the same hyperparameter for training.

The network loss of the training set and test set in the unsupervised and supervised training stage is shown in Figure 7. The network loss converged after both training stages, and the network loss in the test set was just slightly larger than the loss in the training set, which indicated that the proposed training strategy avoided the problem of overfitting.



(a)

Figure 7. Cont.



(b)

Figure 7. Network loss of the training set and test set in the training process: (a) network loss in the unsupervised training stage; (b) network loss in the supervised training stage.

3.2. Accuracy Evaluation

After the two-stage training, the accuracy of the FrANet network based on the supervised data set was evaluated. The test set, including 20 sets of fringe images, was fed into the FrANet network. The error between the predicted phase maps and ground truth obtained by using twelve-step phase-shifting techniques was calculated. The fringe images in the test set did not contain any object that appeared in the training set, thus displaying the capacity of the FrANet network for generalization.

Figure 8 displays results for one fringe image in the test set, where (a) shows the original fringe image, (b) and (c) show ground truth phase maps, and (d)–(f) show the network results, including predicted wrapped phase map, predicted absolute phase map, and refined absolute phase map. The FrANet network predicts an accurate wrapped phase map from the fringe image and performs phase unwrapping. In the first predicted absolute phase map, the shape of the measured object is less clear than in the refined phase map, which verifies the effectiveness of the refinement subnetwork. The refined absolute phase map is similar to the ground truth in the valid region, which means the FrANet network prediction effectively accounts for the shape of the measured object and provides similar results to phase-shifting profilometry for most of the parts.

To visualize the accuracy of the proposed method, error maps of the predicted wrapped phase map, predicted absolute phase map, and refined absolute phase map are shown in Figure 9. The error maps are calculated using the exact results in Figure 7 except that the background is eliminated in this experiment to focus on the measured objects. The absolute error between predicted and ground truth phase maps of each pixel is displayed in the error maps. In addition to the error maps, the mean absolute error (MAE) in the entire test set is also calculated, as shown in Table 1. Since the phase unwrapping subnetwork predicts an absolute phase map instead of fringe orders, it can refine the retrieved wrapped phase map. Unlike in classic FPP, where the maximum accuracy of absolute phase maps is determined by the accuracy of wrapped phase maps, FrANet retains the information of fringe images after wrapped phase maps are predicted and further refinement in hidden layers is conducted throughout the network. There is a relatively large error in the primary wrapped phase map, and the error is reduced after the process of phase unwrapping. The refinement subnetwork further improves the performance and eliminates fringe-shaped

errors in the phase maps. The MAE of the final output is 0.0114 rad. These results indicate that the proposed method can perform high-accuracy 3D measurements and avoids the overfitting problem.

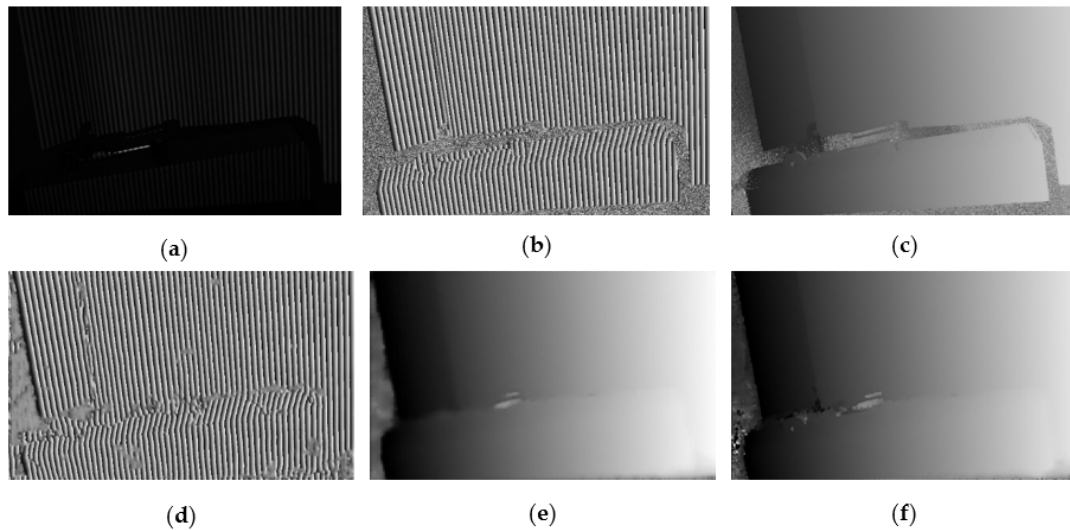


Figure 8. Results for one fringe image in the test set: (a) fringe image; (b) ground truth wrapped phase map; (c) ground truth absolute phase map; (d) predicted wrapped phase map; (e) predicted absolute phase map; (f) refined absolute phase map.

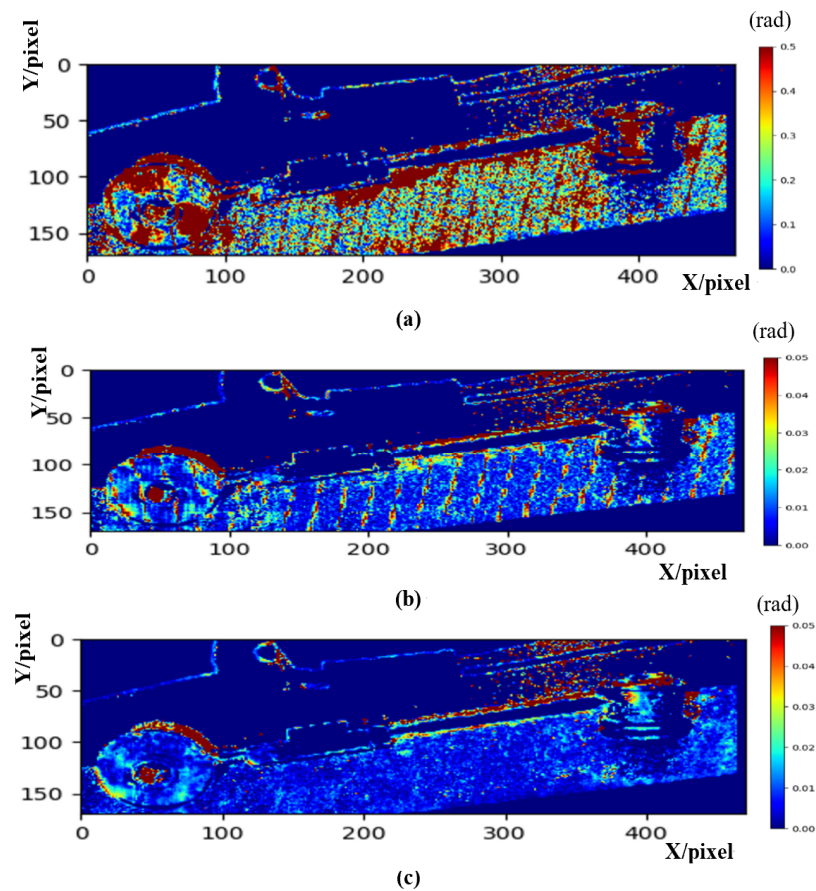


Figure 9. Error maps of predicted wrapped phase map, predicted absolute phase map, and refined absolute phase map with the corresponding MAE in the test set: (a) error maps of predicted wrapped phase map (MAE = 0.3204 rad); (b) error maps of predicted absolute phase map (MAE = 0.0200 rad); (c) error maps of refined absolute phase map (MAE = 0.0114 rad).

Table 1. MAE of predicted wrapped phase maps, predicted absolute phase maps, and refined absolute phase maps in the entire test set.

Network output	Predicted Wrapped Phase Maps	Predicted Absolute Phase Maps	Refined Absolute Phase Maps
MAE (rad)	0.3204	0.0200	0.0114

After the absolute phase maps were yielded, the 3D shapes of the measured objects were reconstructed using both the ground truth phase maps and phase maps predicted by the network. The 3D coordinates were compared to calculate the error in the 3D measurements. The results for the test set have a root mean square error (RMSE) of 0.67 mm. For a data set with various industrial parts, this accuracy is desired in single-shot measurement.

To test the generalization ability of the proposed method, the single-shot 3D shape reconstruction data set proposed by Nguyen et al. [26] was used for finetuning and evaluation. After unsupervised learning using our dataset, the pre-trained model was finetuned using the dataset in the literature. The fringe image and wrapped phase map calculated using the proposed method are shown in Figure 10. To adapt to the end-to-end format of this data set, a convolutional layer was added to transform the phase maps into height maps. Using the proposed method, the RMSE of the test set was 0.94 mm compared to 1.62 mm in the original paper [26].

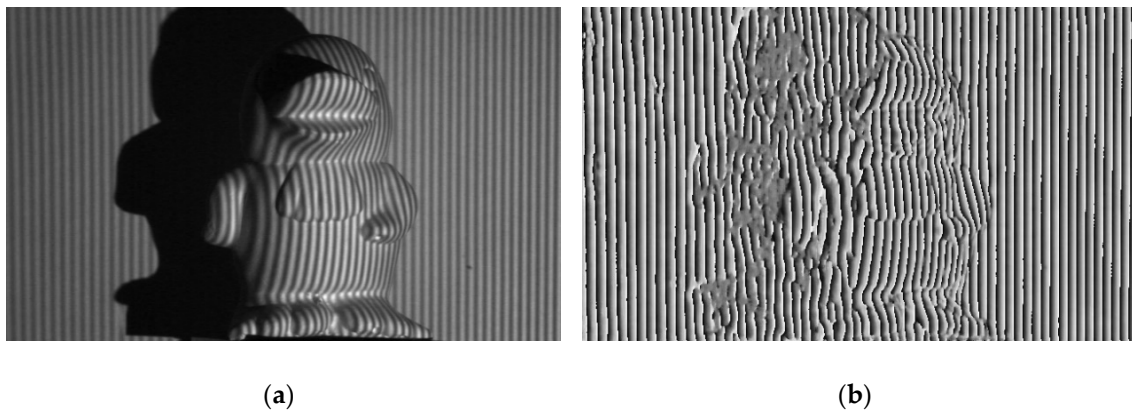


Figure 10. Fringe image and wrapped phase map calculated using the proposed method: (a) Fringe image of the single-shot 3D shape reconstruction dataset; (b) Wrapped phase map calculated using the proposed method.

3.3. Dynamic Scene

Since only one fringe image is required and used as the network input, the proposed method can be applied to dynamic measurements. For the evaluation of dynamic scenes, the object formed a pendulum, and fringe images were captured near the bottom part when the object reached the highest speed, which was estimated using a simple pendulum model. To measure the absolute error without using the phase-shifting results as a reference, standard spheres were adopted for evaluation. A diagram and dynamic scene of the standard sphere pendulum are shown in Figure 11. The diameter of the spheres is 30 mm within an error of 2 μm . The uncertainty of reference diameter is claimed by the sphere manufacturer of DS-DCB-D30L100. The length of the pendulum is $L = 69.5$ cm. Standard spheres that are released at a horizontal deviation of l will reach the estimated maximum speed of $v = \sqrt{2g(L - \sqrt{L^2 - l^2})}$, where g is the gravitational acceleration.

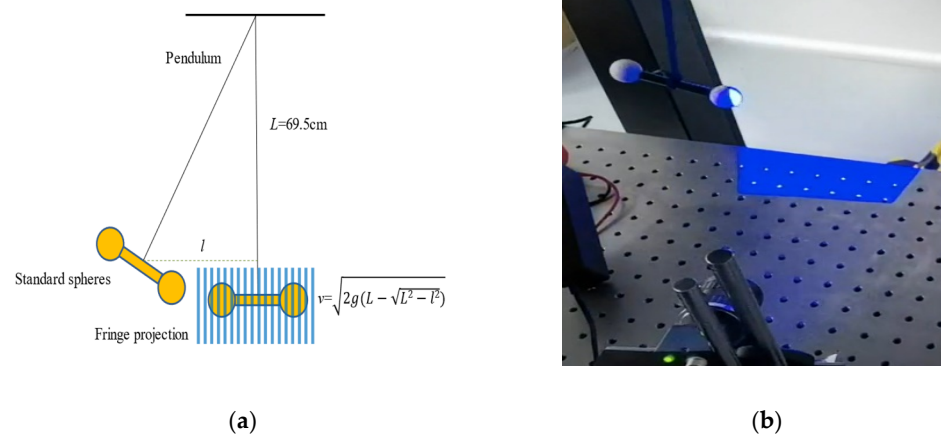


Figure 11. Diagram and dynamic scene of the standard sphere pendulum: (a) Diagram of the standard sphere pendulum for dynamic measurement; (b) Dynamic scene of the standard sphere pendulum for 3D measurement.

The results of the measurements of dynamic scenes are shown in Figure 12. The spheres were released at horizontal deviations of 10.0 cm, 15.0 cm, 20.0 cm, 25.0 cm, and 40.0 cm, corresponding to the highest speeds of 0.376 ± 0.004 m/s, 0.567 ± 0.004 m/s, 0.759 ± 0.004 m/s, 0.955 ± 0.004 m/s, and 1.580 ± 0.005 m/s. The predicted wrapped phase maps and absolute phase maps were accurate without motion blur. The measured diameters were 30.078 mm, 29.946 mm, 29.916 mm, 30.152 mm, and 30.354 mm, with errors of 78 μm , 54 μm , 84 μm , 152 μm , and 354 μm , respectively. The measured diameters were calculated using sphere fitting on the reconstruction results. The RMSEs between 3D coordinates in the reconstruction results and fitted spheres were 49 μm , 32 μm , 56 μm , 88 μm , and 197 μm , respectively. The results show that the proposed method has a high accuracy for dynamic scenes when measuring objects below the speed of 1 m/s.

3.4. Ablation Study

To provide a clear view of the contributions of each technique, we conducted ablation studies of the two training stages and three subnetworks. The results of ablation studies on the training stages are shown in Table 2. We selected the RMSE between prediction and ground truth height maps on the test set as the metric. In unsupervised learning, the pre-trained model without fine-tuning was evaluated. In supervised learning, the FrANet was trained from scratch using the supervised training set which had an insufficient number of training samples. Supervised learning yielded the worst result due to overfitting. The proposed method reached the lowest RMSE and improved the accuracy of unsupervised methods by 68%, which verifies that two-stage training is effective in achieving a high accuracy for the given data sets.

Since the FrANet network consists of multiple subnetworks, we evaluated the contribution of each subnetwork on the test set individually, as shown in Table 3. For the experiments without the phase retrieval subnetwork, WFTP was utilized to extract the wrapped phase map. Spatial phase unwrapping was adopted when testing without the phase unwrapping subnetwork. The refinement subnetwork does not affect the form of the output, and thus there is no need for replacement. The results show that all the subnetworks of the proposed network contributed, and the integration of these subnetworks yielded the best performance.

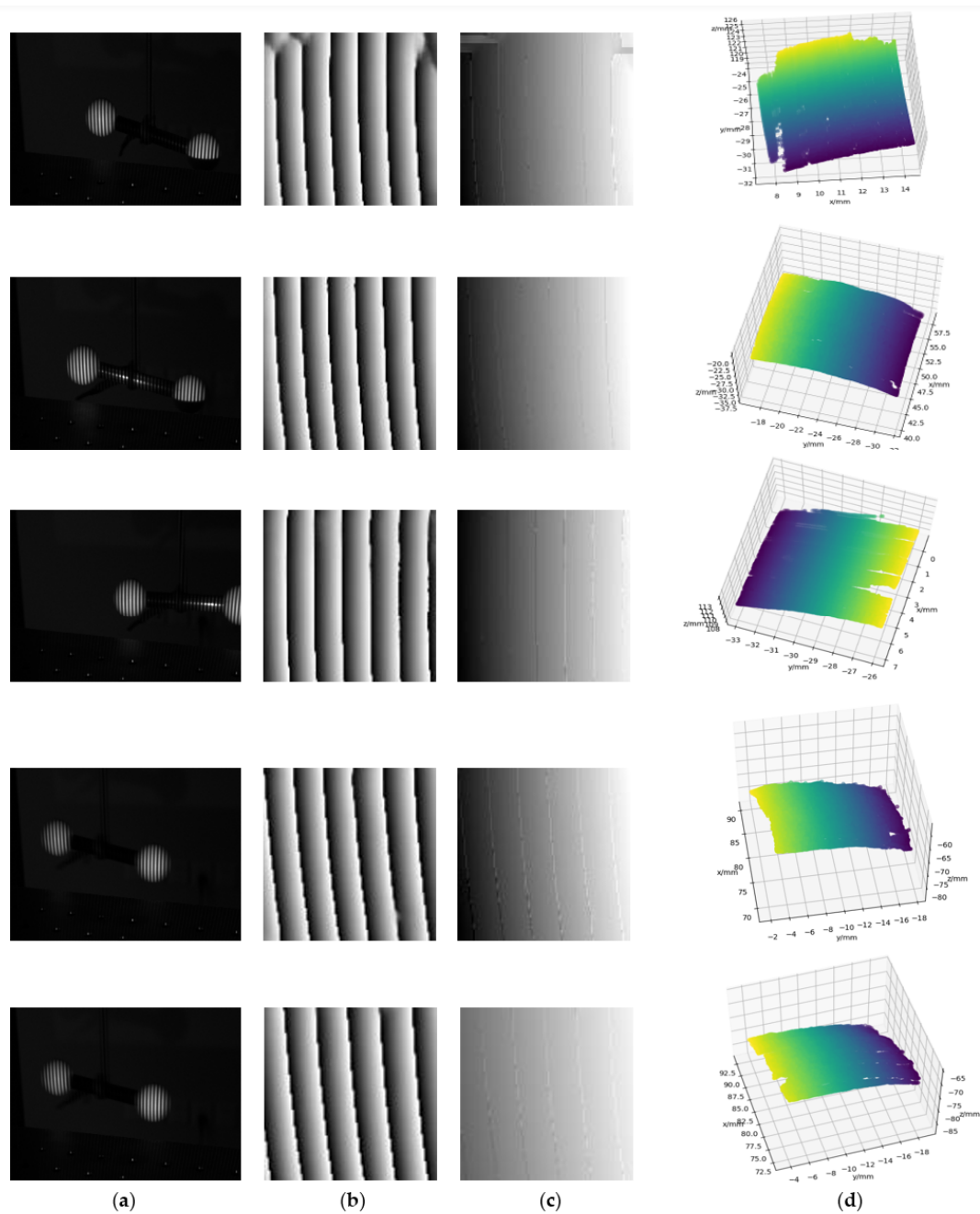


Figure 12. Results of measurements of dynamic scenes. The speeds are 0.376 m/s, 0.567 m/s, 0.759 m/s, 0.955 m/s, and 1.58 m/s from top to bottom: (a) Fringe images of standard spheres in dynamic scenes. (b) Predicted wrapped phase maps of the left sphere piece. (c) Predicted absolute phase maps of the left sphere piece. (d) 3D reconstruction of the left sphere piece.

Table 2. Results of ablation study on the training stages.

Method	Unsupervised Learning	Supervised Learning (With Insufficient Data)	The Proposed Method
RMSE (mm)	2.12	6.38	0.67
Accuracy improvement	68%	89%	/

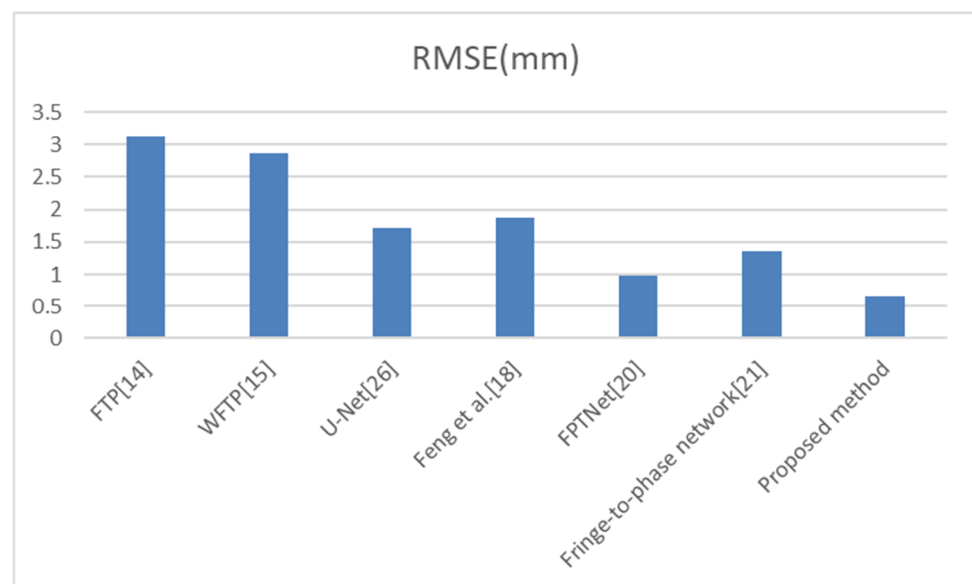
Table 3. Results of ablation studies of the subnetworks.

Subnetwork	PR	PU	RF	PR + PU	PR + RF	PU + RF	PR + PU + RF
Relative RMSE	3.89	3.37	5.90	1.15	3.78	2.95	1.00

PR: Phase retrieval; PU: Phase unwrapping; RF: Refinement.

3.5. Comparison with Other Methods

There are different methods for single-shot 3D measurement, such as traditional FTP, WFTP, and deep learning methods. In the experiments using traditional methods, fringe images were captured at four frequencies, 1, 4, 16, and 64. Temporal phase unwrapping was conducted to obtain absolute phase maps. Deep learning methods include end-to-end networks and networks used for phase retrieval. The U-Net architecture proposed by Nguyen et al. [24] as an end-to-end network was adopted for comparison. Note that we switched only the network architecture and still used two-stage training for the end-to-end network, in which fringe patterns were re-projected according to the predicted height maps in the unsupervised stage. Networks used for phase retrieval were also compared, including the convolutional network proposed by Feng et al. [18], the FPTNet [20], and the fringe-to-phase network [21]. These networks were also trained in two stages. The results of the proposed method were superior to those of traditional methods, deep networks for phase retrieval, and end-to-end deep learning methods, as shown in Figure 13. The result of each single-shot method was computed using the RMSE between 3D coordinates of reconstruction results and ground truth in the test set, where the ground truth was obtained using twelve-step phase-shifting and temporal phase unwrapping. The FPTNet yielded the second-best performance, but the input of the FPTNet had an additional low-frequency fringe image for phase unwrapping. The proposed method achieved the highest accuracy when processing one single-shot fringe image.

**Figure 13.** Comparison with different single-shot methods.

The 3D reconstruction results using different single-shot methods are shown in Figure 14. Methods with the highest accuracy, namely, WFTP [15] in traditional methods and FPTNet [20] in deep-learning-based methods, were adopted for qualitative comparison. The result of WFTP contained rough edges and notable shape distortion, which indicated a large phase error. FPTNet recovered the shape of the object well but lost a small portion of pixels for reconstruction due to phase unwrapping errors. The proposed method yielded the most accurate 3D reconstruction result.

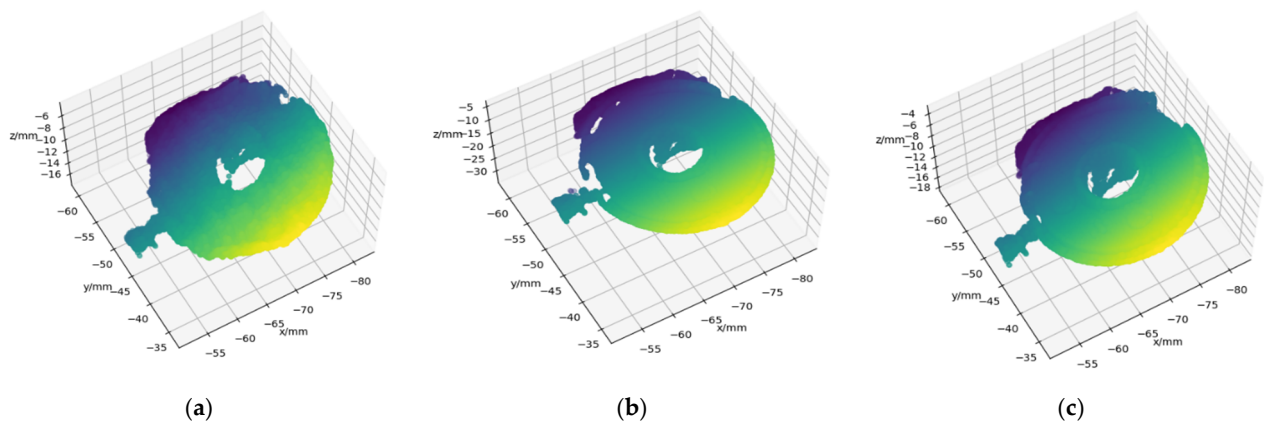


Figure 14. 3D reconstruction results using different single-shot methods: (a) result of WFTP [15]; (b) result of the FPTNet [20]; (c) result of the proposed method.

3.6. Data Acquisition Efficiency

We compared the efficiency of the proposed method with unsupervised and supervised methods. To ensure a fair comparison, the other methods were assumed to have the same number of training samples as the unsupervised data set used in this work. Unsupervised methods require one to collect 4000 fringe images in data acquisition, while supervised methods require 96,000 fringe images, and the proposed method requires 8800 fringe images. The results suggest that the proposed method has a comparable data efficiency to the unsupervised methods and is superior to the supervised methods.

4. Discussion

Single-shot methods extract phase maps from one fringe image and are robust against movement. Traditional Fourier transform, wavelet transform, and other frequency domains transform single-shot methods are sensitive to noise, which results in low accuracy, especially when measuring objects with strong textures. The RMSEs of 3D reconstruction results using traditional FTP [14] and WFTP [15] methods are 3.12 mm and 2.87 mm, respectively, which are significantly larger than that of the proposed method. Single-shot methods based on deep learning are more accurate and robust than traditional methods. However, these methods mostly adopt network architectures without subnetworks that focus on either phase unwrapping or height map prediction. Fringe-to-phase networks mostly use generic CNNs or U-Net for phase unwrapping, while end-to-end networks adopt variations of U-Net architecture to directly predict the height map. These tasks are ill-posed tasks that require long-range information to remove phase ambiguity, and a single convolutional network cannot process long-range information due to the small receptive field of the convolutions. The FrANet retains long-range information in the predicted phase map by using the combination of three subnetworks. The proposed method achieved high accuracy with an RMSE of 0.67 mm in 3D reconstruction, which exceeds other methods based on deep learning, including U-Net [26], Feng et al. [18], the FPTNet [20], and the Fringe-to-phase network [21]. Despite the high accuracy of deep learning methods, the time-consuming nature of data acquisition is a significant limitation for practical applications. Methods using computer graphics to tackle this problem have not been thoroughly evaluated in real-world setups. The difference between synthetic and real-world domains remains a problem. The proposed method utilizes two-stage training to bridge the gap between the domains and achieve high accuracy while preserving high data efficiency using a real-world FPP system. The proposed method only requires 8800 fringe images to facilitate training, while supervised methods require fringe images. Other two-stage training methods could also be explored in further research. For instance, pre-training using a synthetic data set and fine-tuning using a real-world data set might also offer solutions. Variations of the

proposed network, in which the U-Net architecture of subnetworks can be adjusted, could also be further explored.

5. Conclusions

Multi-shot methods that mostly adopt phase-shifting techniques provide highly accurate measurement data for static objects by capturing multiple fringe pattern images. However, the performance may be degraded due to disturbance from vibration. Moreover, these methods are time consuming and inaccurate when applied to dynamic scenes. In this work, a single-shot 3D measurement method using a deep neural network named FrANet was proposed. The FrANet consists of three subnetworks for fringe analysis: a phase retrieval subnetwork, a phase unwrapping subnetwork, and a refinement subnetwork. This system renders long-range information accessible, which is necessary for accurate phase unwrapping or height map prediction. Generic convolutional networks have small receptive fields and cannot extract long-range information. The phase unwrapping subnetwork in the FrANet acquires long-range information using additional down-sampling layers, and the refinement subnetwork conducts further refinement. All the subnetworks in the FrANet adopt the improved U-Net architecture, which is efficient in processing high-resolution images. To reduce the number of supervised training samples required, in this work, a two-stage training strategy was designed. The two stages included pre-training using unsupervised learning and fine-tuning using supervised learning. Re-projected fringe images were obtained using network predictions to construct unsupervised loss. Twelve-step phase-shifting techniques were adopted to acquire ground truth for supervised learning. We also explored the network architecture for efficient fringe analysis. The proposed network consists of a phase retrieval subnetwork, phase unwrapping subnetwork, and refinement subnetwork. The experimental results obtained using a real-world FPP system indicate that the proposed method achieves accurate single-shot 3D measurements with an RMSE of 0.67 mm on the test set. The measurements of the moving standard spheres verify the effectiveness of the method for dynamic scenes. The measurement errors of sphere diameter were 78 μm , 54 μm , 84 μm , 152 μm , and 354 μm corresponding to speeds of 0.376 ± 0.004 m/s, 0.567 ± 0.004 m/s, 0.759 ± 0.004 m/s, 0.955 ± 0.004 m/s, and 1.580 ± 0.005 m/s. Standard sphere diameters at most speeds were measured with an error of around 100 μm , which is considered high accuracy in single-shot 3D measurement. Two-stage training with 8800 fringe images saves time during data acquisition compared to supervised methods. The ablation studies verify the effectiveness of using two training stages and three network subnetworks, as two-stage training reduces the error in unsupervised training by 68%, and the network comprising all the subnetworks achieves the highest accuracy. The results of the proposed method are superior to those of FTP, WFTP, and end-to-end networks. The proposed method achieves a high data efficiency comparable to that of unsupervised methods and high accuracy in real-world setups. Future work includes reducing the complexity of the training strategy and using light-weighted networks to further accelerate the measurement process. Variations of the network architecture utilizing multiple subnetworks could be explored and other two-stage training strategies could also be attempted.

Author Contributions: M.W. is responsible for Conceptualization, Methodology, Software, Validation, Investigation, and Writing—original draft preparation; L.K. is responsible for Conceptualization, Methodology, Investigation, Supervision, Funding acquisition, Writing—review, and editing; X.P. is responsible for Validation, Formal analysis, Writing—review, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 52075100).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zuo, C.; Feng, S.; Huang, L.; Tao, T.; Yin, W.; Chen, Q. Phase shifting algorithms for fringe projection profilometry: A review. *Opt. Lasers Eng.* **2018**, *109*, 23–59. [\[CrossRef\]](#)
2. Zuo, C.; Huang, L.; Zhang, M.; Chen, Q.; Asundi, A. Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review. *Opt. Lasers Eng.* **2016**, *85*, 84–103. [\[CrossRef\]](#)
3. Yu, S.; Gong, T.; Wu, H.; Sun, X.; Zhao, Y.; Wu, S.; Yu, X. 3D shape measurement based on the unequal-period combination of shifting Gray code and dual-frequency phase-shifting fringes. *Opt. Commun.* **2022**, *516*, 128236. [\[CrossRef\]](#)
4. Peng, R.; Tian, M.; Xu, L.; Yang, L.; Yue, H. A novel method of generating phase-shifting sinusoidal fringes for 3D shape measurement. *Opt. Lasers Eng.* **2021**, *137*, 106401. [\[CrossRef\]](#)
5. Hu, Y.; Chen, Q.; Liang, Y.; Feng, S.; Tao, T.; Zuo, C. Microscopic 3D measurement of shiny surfaces based on a multi-frequency phase-shifting scheme. *Opt. Lasers Eng.* **2019**, *122*, 1–7. [\[CrossRef\]](#)
6. Wu, H.; Cao, Y.; An, H.; Li, Y.; Li, H.; Xu, C.; Yang, N. A novel phase-shifting profilometry to realize temporal phase unwrapping simultaneously with the least fringe patterns. *Opt. Lasers Eng.* **2022**, *153*, 107004. [\[CrossRef\]](#)
7. Li, H.; Cao, Y.; Wan, Y.; Xu, C.; Zhang, H.; An, H.; Wu, H. An improved temporal phase unwrapping based on super-grayscale multi-frequency grating projection. *Opt. Lasers Eng.* **2022**, *153*, 106990. [\[CrossRef\]](#)
8. Li, L.; Zheng, Y.; Yang, K.; Su, X.; Wang, Y.; Chen, X.; Wang, Y.; Li, B. Modified three-wavelength phase unwrapping algorithm for dynamic three-dimensional shape measurement. *Opt. Commun.* **2021**, *480*, 126409. [\[CrossRef\]](#)
9. Li, J.; Guan, J.; Du, H.; Xi, J. Error self-correction method for phase jump in multi-frequency phase-shifting structured light. *Appl. Opt.* **2021**, *60*, 949–958. [\[CrossRef\]](#)
10. Pistellato, M.; Bergamasco, F.; Albarelli, A.; Cosmo, L.; Gasparetto, A.; Torsello, A. Robust phase unwrapping by probabilistic consensus. *Opt. Lasers Eng.* **2019**, *121*, 428–440. [\[CrossRef\]](#)
11. Lilenblum, E.; Michaelis, B. Optical 3D Surface Reconstruction by a Multi-Period Phase Shift Method. *J. Comput.* **2007**, *2*, 73–83. [\[CrossRef\]](#)
12. Qi, X.; Zhou, C.; Ding, Y.; Wang, Y.; Si, S.; Li, H. Novel absolute phase measurement method with few-patterns. *Opt. Lasers Eng.* **2022**, *154*, 107031. [\[CrossRef\]](#)
13. Yao, P.; Gai, S.; Da, F. Coding-Net: A multi-purpose neural network for Fringe Projection Profilometry. *Opt. Commun.* **2021**, *489*, 126887. [\[CrossRef\]](#)
14. Takeda, M.; Mutoh, K. Fourier transform profilometry for the automatic measurement of 3-D object shapes. *Appl. Opt.* **1983**, *22*, 3977–3982. [\[CrossRef\]](#)
15. Kemao, Q. Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations. *Opt. Lasers Eng.* **2007**, *45*, 304–317. [\[CrossRef\]](#)
16. Zhong, J.; Weng, J. Spatial carrier-fringe pattern analysis by means of wavelet transform: Wavelet transform profilometry. *Appl. Opt.* **2004**, *43*, 4993–4998. [\[CrossRef\]](#)
17. Kawasaki, H.; Furukawa, R.; Sagawa, R.; Yagi, Y. Dynamic scene shape reconstruction using a single structured light pattern. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AL, USA, 24–26 June 2008.
18. Feng, S.; Chen, Q.; Gu, G.; Tao, T.; Zhang, L.; Hu, Y.; Yin, W.; Zuo, C. Fringe pattern analysis using deep learning. *Adv. Photonics* **2019**, *1*, 025001. [\[CrossRef\]](#)
19. Qiao, G.; Huang, Y.; Song, Y.; Yue, H.; Liu, Y. A single-shot phase retrieval method for phase measuring deflectometry based on deep learning. *Opt. Commun.* **2020**, *476*, 126303. [\[CrossRef\]](#)
20. Yu, H.; Chen, X.; Zhang, Z.; Zuo, C.; Zhang, Y.; Zheng, D. Dynamic 3-d measurement based on fringe-to-fringe transformation using deep learning. *Opt. Express* **2020**, *28*, 9405–9418. [\[CrossRef\]](#)
21. Nguyen, H.; Novak, E.; Wang, Z. Accurate 3D reconstruction via fringe-to-phase network. *Measurement* **2022**, *190*, 110663. [\[CrossRef\]](#)
22. Zhang, L.; Chen, Q.; Zuo, C.; Feng, S. High-speed high dynamic range 3D shape measurement based on deep learning. *Opt. Lasers Eng.* **2020**, *134*, 106245. [\[CrossRef\]](#)
23. Qian, J.; Feng, S.; Li, Y.; Tao, T.; Han, J.; Chen, Q.; Zuo, C. Single-shot absolute 3d shape measurement with deep-learning-based color fringe projection profilometry. *Opt. Lett.* **2020**, *45*, 1842–1845. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Qian, J.; Feng, S.; Tao, T.; Hu, Y.; Li, Y.; Chen, Q.; Zuo, C. Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement. *Appl. Photonics* **2020**, *5*, 046105. [\[CrossRef\]](#)
25. Van der Jeught, S.; Dirckx, J.J. Deep neural networks for single shot structured light profilometry. *Opt. Express* **2019**, *27*, 17091–17101. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Nguyen, H.; Wang, Y.; Wang, Z. Single-shot 3d shape reconstruction using structured light and deep convolutional neural networks. *Sensors* **2020**, *20*, 3718. [\[CrossRef\]](#) [\[PubMed\]](#)

27. Machineni, R.C.; Spoorthi, G.E.; Vengala, K.S.; Gorthi, S.; Gorthi, R.K.S.S. End-to-end deep learning-based fringe projection framework for 3D profiling of objects. *Comput. Vis. Image Und.* **2020**, *199*, 103023. [[CrossRef](#)]
28. Nguyen, H.; Tan, T.; Wang, Y.; Wang, Z. Three-dimensional shape reconstruction from single-shot speckle image using deep convolutional neural networks. *Opt. Lasers Eng.* **2021**, *143*, 106639. [[CrossRef](#)]
29. Zheng, Y.; Wang, S.-D.; Li, Q.; Li, B.-W. Fringe projection profilometry by conducting deep learning from its digital twin. *Opt. Express* **2020**, *28*, 21692–21703. [[CrossRef](#)]
30. Wang, F.; Wang, C.; Guan, Q. Single-shot fringe projection profilometry based on deep learning and computer graphics. *Opt. Express* **2021**, *29*, 8024–8040. [[CrossRef](#)]
31. Fan, S.; Liu, S.; Zhang, X.; Huang, H.; Liu, W.; Jin, P. Unsupervised deep learning for 3D reconstruction with dual-frequency fringe projection profilometry. *Opt. Express* **2021**, *29*, 32547–32566. [[CrossRef](#)]
32. Bolon-Canedo, V.; Remeseiro, B. Feature selection in image analysis: A survey. *Artif. Intell. Rev.* **2020**, *53*, 2905–2931. [[CrossRef](#)]
33. Kabir, H.; Garg, N. Machine learning enabled orthogonal camera goniometry for accurate and robust contact angle measurements. *Sci. Rep.* **2023**, *13*, 1497. [[CrossRef](#)] [[PubMed](#)]
34. Jia, D.; Wei, D.; Socher, R.; Li, L.-J.; Kai, L.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
35. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019.
36. Mittal, S.; Tatarchenko, M.; Brox, T. Semi-supervised semantic segmentation with high- and low-level consistency. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1369–1379. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.