sensors

MDPI

Article

Generalized Fringe-To-Phase Framework for Single-Shot 3D Reconstruction Integrating Structured Light with Deep Learning

Andrew-Hieu Nguyen [1,2], Khanh L. Ly [3], Van Khanh Lam [4] and Zhaoyang Wang [1,*]

[1] Department of Mechanical Engineering, The Catholic University of America, Washington, DC 20064, USA; hieu.nguyen@nih.gov
[2] Neuroimaging Research Branch, National Institute on Drug Abuse, National Institutes of Health, Baltimore, MD 21224, USA
[3] Department of Biomedical Engineering, The Catholic University of America, Washington, DC 20064, USA
[4] Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital, Washington, DC 20012, USA
* Correspondence: wangz@cua.edu

Abstract: Three-dimensional (3D) shape acquisition of objects from a single-shot image has been highly demanded by numerous applications in many fields, such as medical imaging, robotic navigation, virtual reality, and product in-line inspection. This paper presents a robust 3D shape reconstruction approach integrating a structured-light technique with a deep learning-based artificial neural network. The proposed approach employs a single-input dual-output network capable of transforming a single structured-light image into two intermediate outputs of multiple phase-shifted fringe patterns and a coarse phase map, through which the unwrapped true phase distributions containing the depth information of the imaging target can be accurately determined for subsequent 3D reconstruction process. A conventional fringe projection technique is employed to prepare the ground-truth training labels, and part of its classic algorithm is adopted to preserve the accuracy of the 3D reconstruction. Numerous experiments have been conducted to assess the proposed technique, and its robustness makes it a promising and much-needed tool for scientific research and engineering applications.

Keywords: three-dimensional image acquisition; three-dimensional sensing; single-shot imaging; fringe-to-phase transformation; convolutional neural network; deep learning

Citation: Nguyen, A.-H.; Ly, K.L.; Lam, V.K.; Wang, Z. Generalized Fringe-To-Phase Framework for Single-Shot 3D Reconstruction Integrating Structured Light with Deep Learning. *Sensors* **2023**, *23*, 4209. https://doi.org/10.3390/s23094209

Academic Editor: Shijie Feng

Received: 15 March 2023
Revised: 16 April 2023
Accepted: 18 April 2023
Published: 23 April 2023

1. Introduction

In recent years, three-dimensional (3D) reconstruction has been widely adopted in numerous fields such as medical imaging, robotic navigation, palletization, virtual reality, 3D animation modeling, and product in-line inspection [1–4]. Generally speaking, 3D reconstruction is a process of creating the 3D geometric shape and appearance of a target, typically from single-view or multiple-view two-dimensional (2D) images of the target. It can be classified into passive and active 3D reconstructions [5–7]. The passive approach does not interfere with the target, instead recording the radiance reflected or emitted by its surface. The passive technique requires only imaging components, so it is easy to implement with respect to hardware. Much work has been accomplished in the past many years to increase the speed of the passive 3D reconstruction; nevertheless, the acquired 3D representation tends to have low accuracy if the target lacks sufficient texture variations on its surface [8]. On the other hand, active 3D reconstruction involves using radiation (e.g., laser or structured light) to interfere with the target. It uses relatively more complex hardware components but is often capable of providing 3D shape results with higher accuracy than the passive counterpart, particularly for regions without texture. This paper focuses on the study of the active approach.

Commonly used sensing methods in active 3D reconstruction include laser scanning, time-of-flight (ToF), and structured-light methods [9–11]. The laser scanning and ToF techniques are similar, and they measure the 3D profiles of a target by calculating distances traveled by the light. The two methods have gained considerable attention in recent years because of their capabilities of measuring long distances and generating 3D representations at real-time speed (e.g., 30Hz), which satisfy the sensing requirements in autonomous vehicles and virtual/augmented reality [12–14]. A drawback of the techniques is their relatively low accuracy. In addition, the ToF techniques usually use infrared light, which is often interfered with by the background lights and reflections. Such interference hinders its popularity in more sophisticated applications [15]. The structured-light system typically uses a projector to project structured patterns onto a target. The patterns are distorted by the geometric shape of the target surface, and they are then captured for 3D shape reconstruction. The structured-light techniques fall into two categories according to the input: single-shot and multi-shot approaches [16]. Compared with the multi-shot methods, the single-shot ones are faster and can be used for real-time 3D reconstruction. With recent developments, the accuracy of the single-shot techniques has been notably improved [17]. For instance, Fernandez developed a one-shot dense point reconstruction technique integrated with an absolute coding phase-unwrapping algorithm, which provides a dynamic and highly accurate phase map for depth reconstruction [18]. Later, Moreno et al. proposed an approach to estimate the coordinates of the calibration points in the projector image plane using local homographs, through which the acquisition time can be substantially reduced while enhancing the resolution of the acquired 3D shapes [19]. Jensen and colleagues used a differentiable rendering technique to directly compute the surface mesh without an intermediate point cloud, which shortens the computation time and improves the accuracy, especially for objects with sharp features [20]. Recently, Tran et al. built a structured-light RGB-D camera system with a gray-code coding scheme to produce high-quality 3D reconstruction in a real-world environment [21].

Over the past few years, deep learning has gained significant interest in computer vision thanks to its superior modeling and computational capabilities. Specifically, deep learning has been applied to many complicated computer vision tasks such as image classification, object detection, face recognition, and human pose estimation [22–25]. With unique feature learning capability and extreme computational power, convolutional neural network (CNN) can be utilized for various tasks based on given training datasets. In contrast, unsupervised learning features of deep belief networks, deep Boltzmann machines, and stacked autoencoders can remove the need for labeled datasets [26]. Additionally, deep learning can improve the accuracy of challenging tasks in computer vision while requiring less expert analysis and fine-tuning using trained neural networks compared with the traditional computer vision techniques [27]. With no exception, in 3D reconstruction, deep learning has also been widely employed in numerous studies [28,29]. Recently, Zhang et al. proposed a RealPoint3D network comprising an encoder, a 2D–3D fusion module, and a decoder to reconstruct fine-grained 3D representations from a single image. The 2D–3D fusion module helps generate detailed results from images with either solid-color or complex backgrounds [30]. Jeught and colleagues constructed a CNN on a large set of simulated height maps with associated deformed fringe patterns, which was then able to predict the 3D height map from unseen input fringe patterns with high accuracy [31].

Deep learning-based approaches have been adopted in optical measurement and experimental mechanics to accomplish several classic tasks such as phase extraction, fringe analysis, interferogram denoising, and deformation determination [32–37]. Numerous devoted topics in 3D shape measurement involving deep learning have been introduced in the last few years. Notably, a well-known structured-light technique, fringe projection profilometry (FPP), has been integrated frequently with deep learning methods to prepare accurate 3D shapes of objects as high-quality ground-truth labels. A typical FPP-based 3D imaging technique contains several key steps, such as phase extraction, phase unwrapping, fringe order determination, and depth estimation. Although the deep learning models can

be applied at various stages of the FPP technique, the objective of producing a high-accuracy 3D shape remains the same.

The 3D shape reconstruction techniques integrating the FPP-based method with deep learning typically fall into two groups: fringe-to-depth and fringe-to-phase approaches. The former group intends to perform an image-to-image transformation using a neural network model where the input–output pair is a single structured-light image and a corresponding depth map [38–43]. Compared with the typical deep learning-based depth estimation techniques from a single image in the computer vision field, the major differences include a structured-light input and a high-quality depth map output [44–46]. First, the structured-light illumination applies desired feature patterns for accurate geometric information extraction, particularly in the textureless regions. Second, the ground-truth depth maps produced by the FPP-based 3D imaging technique provide higher accuracy (originated from full-field sub-pixel image matching) than the ones obtained from the RGB-D sensors.

The latter fringe-to-phase group aims to transform fringe pattern(s) into several potential intermediate outputs before determining the unwrapped phase and 3D shape by the conventional technique [47–50]. Here, *unwrapped* refers to the demodulation of the wrapped phase because conventional algorithms generally yield phase data wrapped in a small value range. Figure 1 demonstrates the pipeline of recent fringe-to-phase approaches. Researchers in [51–54] developed the pattern-to-pattern schemes where fringe pattern(s) can be converted into multiple phase-shifted fringe patterns. In order to simplify the output and reduce the storage space, the numerator and denominator terms of the arctangent function have been selected as the training output in some neural network-based approaches [55–57]. Furthermore, the fringe-pattern input can be transformed directly into the wrapped phase map with a phase range of $[-\pi, \pi)$ [58–60]. While either phase-shifted fringe patterns, the numerator and denominator, or the wrapped phase map can be obtained via deep learning networks, a map of integer fringe orders is still necessary for the succeeding phase-unwrapping process. The integer fringe orders can be determined using deep learning via an approach such as linear prediction of coarse phase map [61,62] or fringe-order segmentation [63–66].

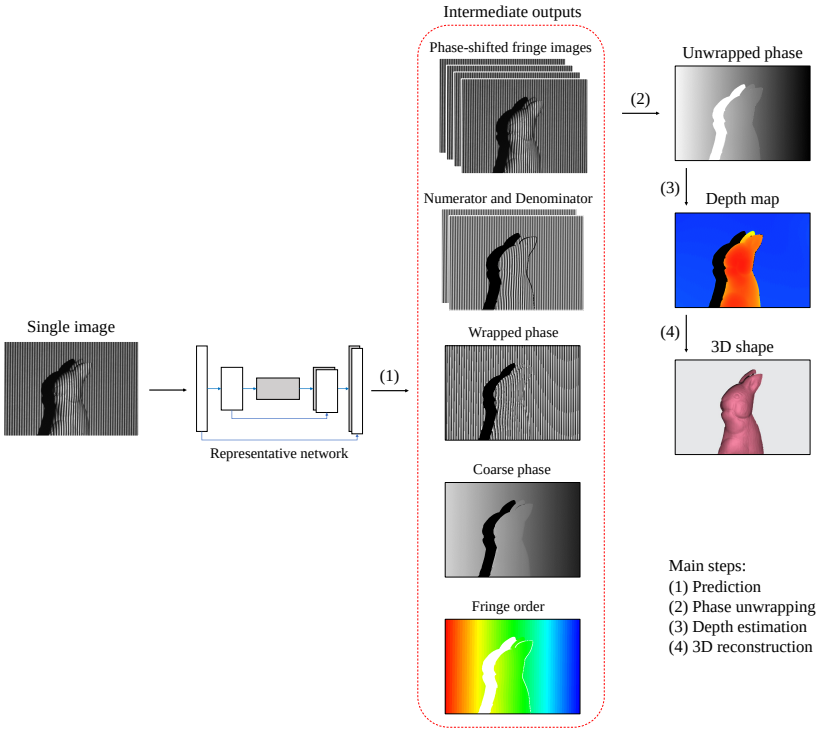

Figure 1. Recent fringe-to-phase approaches integrating structured light with deep learning.

Even though the recently developed fringe-to-phase approaches can help achieve high-accuracy 3D shape measurements, there are still many aspects to improve such as eliminating the usage of multiple sub-networks, multiple inputs, reference image, color composite image, and so on. Inspired by the recent fringe-to-phase methods' advantages, weaknesses, and limitations, this paper presents a novel 3D shape reconstruction technique that transforms a single fringe pattern into two intermediate outputs, phase-shifted fringe patterns and a coarse unwrapped phase, via a single deep learning-based network. Upon successful completion of training, the network can predict outputs of phase-shifted fringe patterns and coarse unwrapped phase to obtain the wrapped phase and integer fringe orders, which further yield a true unwrapped phase for subsequent 3D reconstruction. It is important to note that the full FPP-based technique is only employed for preparing the training dataset, and the 3D reconstruction after prediction just uses part of the FPP-based algorithm. In comparison to the previous fringe-to-depth and fringe-to-phase methods, significant contributions of the proposed technique lie in the following:

1. It requires a single image and a single network. A single network is proposed to transform a single image into four phase-shifted fringe images and an unwrapped coarse phase map;
2. It preserves the accuracy advantage of the classic FPP-based method while eliminating the disadvantage of slow speed originated from capturing multiple phase-shifted fringe patterns;
3. It uses a concise network. Only a single network is used for phase determination instead of multiple sub-networks;
4. It takes a simple image. A single grayscale image is utilized for the training network rather than using a color composite image or an additional reference image;
5. It yields higher accuracy than the image-to-depth approaches.

The rest of the paper is organized as follows. Section 2 depicts the details of the FPP-based technique and the proposed network for phase measurement. Several experiments and relevant assessments are conducted and described in Section 3 to validate the proposed approach. Section 4 includes discussions, and Section 5 gives a brief summary.

2. Methodology

The proposed approach uses a supervised deep-learning network to reconstruct the 3D object shape from a single structured-light image. The main task of the network is transforming a single fringe pattern into two intermediate outputs, i.e., phase-shifted sinusoidal fringe patterns and coarse phase distributions, from which the 3D shape can then be reconstructed using a conventional algorithm. In particular, an FPP technique is adopted to prepare ground-truth training labels and part of its algorithm is utilized to accomplish the subsequent 3D reconstruction task after prediction. The employed conventional FPP technique and the proposed network are described in the following subsections.

2.1. Fringe Projection Profilometry Technique for 3D Shape Reconstruction

The FPP-based 3D imaging system mainly consists of a camera and a projector. During imaging, the projector emits a series of phase-shifted fringe patterns onto the object's surface, and the synchronous camera captures the distorted fringe patterns in turns. The distorted patterns, in which height or depth information of the target is encoded, are then analyzed to acquire the phase distributions and 3D shapes according to the geometric triangulation. Figure 2 illustrates the pipeline of the classic FPP-based 3D shape reconstruction technique.

The initial fringe patterns are generated following a sinusoidal waveform [67]:

$$I_n^{p,(m)}(u,v) = I_0\left[1+\cos\left(\phi^{(m)}(u,v)+\delta_n\right)\right] \quad (1)$$

where I^p is the intensity value of the fringe image at pixel coordinate (u, v); the superscript (m) indicates the mth frequency with $m = \{1, 2, 3\}$; the subscript n denotes the nth phase-shifted image with $n = \{1, 2, 3, 4\}$; I_0 is the constant amplitude of the fringes and is normally set to $\frac{255}{2}$; ϕ is the fringe phase defined as $\phi^{(m)}(u, v) = 2\pi f^{(m)} \frac{u}{W}$, with f and W being the fringe frequency (i.e., the number of fringes in the whole pattern) and the width of the generated fringe pattern, respectively; δ is the phase-shifting amount with $\delta_n = \frac{(n-1)\pi}{2}$.

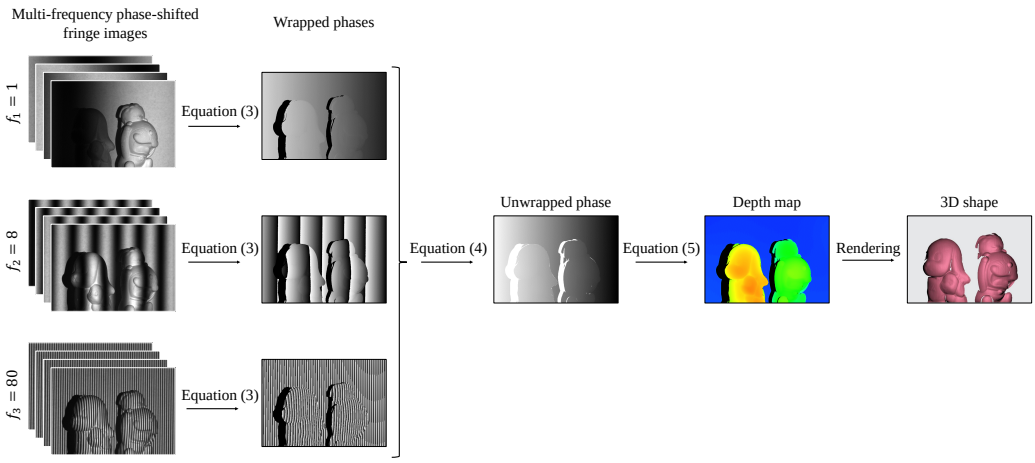

Figure 2. Flowchart of the conventional FPP technique for 3D reconstruction.

After being projected onto an object's surface, the initially evenly spaced fringes follow the surface height or depth profiles and are distorted when seen from the camera view. The distorted fringes have surface profile information encoded into them. Mathematically, the captured fringe patterns can be described as:

$$I_n^{(m)}(u, v) = I_a^{(m)}(u, v) + I_b^{(m)}(u, v) \cos\left[\phi^{(m)}(u, v) + \delta_n\right] \tag{2}$$

where I, I_a, and I_b are the captured fringe intensity, the background intensity, and the amplitude of the intensity modulation at (u, v), respectively. For simplicity, the pixel coordinate (u, v) will be left out in the following equations.

The phase ϕ at each frequency can be retrieved by an inverse trigonometric function as:

$$\phi_w^{(m)} = \text{atan2} \frac{I_4^{(m)} - I_2^{(m)}}{I_1^{(m)} - I_3^{(m)}} \tag{3}$$

In the equation, atan2 denotes the two-argument four-quadrant inverse tangent function; the subscript w implies *wrapped* because the output of the arctangent function is in a range of $[-\pi, \pi)$, whereas the true or unwrapped phase values have a much broader range. In order to resolve phase ambiguities and retrieve the true phase distributions, we use a multi-frequency phase-shifting (MFPS) scheme that is commonly adopted by the FPP methods. With the MFPS scheme, the true phase distributions are determined by the patterns with the highest frequency, and the wrapped phase values of the lower frequency patterns only serve to obtain the integer fringe orders. Specifically, the phase distributions can be determined by the MFPS scheme as [68,69]:

$$\phi^{(m)} = \phi_w^{(m)} + \text{INT}\left(\frac{\phi^{(m-1)} \frac{f^{(m)}}{f^{(m-1)}} - \phi_w^{(m)}}{2\pi}\right) 2\pi \tag{4}$$

where ϕ and ϕ_w imply the unwrapped and wrapped phase distributions, respectively; INT denotes a function of rounding to the nearest integer; and again, $f^{(m)}$ represents the mth

fringe frequency. In this work, the frequencies satisfy $f^{(3)} > f^{(2)} > f^{(1)}$ with $f^{(3)} = 80$, $f^{(2)} = 8$, and $f^{(1)} = 1$, which performs well in practice on dealing with multiple objects and objects with complex shapes. The phase distributions are calculated in the recursive order of $\phi^{(1)}$, $\phi^{(2)}$, and $\phi^{(3)}$, where $\phi^{(1)} = \phi_w^{(1)}$ is automatically fulfilled for $f^{(1)} = 1$. The desired phase distributions to use for the 3D reconstruction are $\phi = \phi^{(3)}$.

The out-of-plane height and depth information of the target being imaged is determined by the following model:

$$\begin{aligned} z &= \frac{\langle \mathbf{C}, \mathbf{P} \rangle_{\mathrm{F}}}{\langle \mathbf{D}, \mathbf{P} \rangle_{\mathrm{F}}} \\ \mathbf{C} &= \begin{bmatrix} 1 & c_1 & c_2 & c_3 & \cdots & c_{27} & c_{28} & c_{29} \end{bmatrix} \\ \mathbf{D} &= \begin{bmatrix} d_0 & d_1 & d_2 & d_3 & \cdots & d_{27} & d_{28} & d_{29} \end{bmatrix} \\ \mathbf{P} &= \begin{bmatrix} 1 & u & v & u^2 & uv & v^2 & u^3 & u^2v & uv^2 & v^3 & u^4 & u^3v & u^2v^2 & uv^3 & v^4 \end{bmatrix} \otimes \begin{bmatrix} 1 & \phi \end{bmatrix} \end{aligned} \tag{5}$$

where z is the physical out-of-plane height or depth of the measurement target, $\langle\ \rangle_{\mathrm{F}}$ and $\otimes$ denote the Frobenius inner product and Kronecker product, respectively; $c_1 - c_{29}$ and $d_0 - d_{29}$ are 59 parameters. These parameters, together with the camera intrinsic and extrinsic parameters, can be acquired by using a flexible calibration process [70,71].

A dataset of 2048 samples was prepared using the described FPP imaging system, and the sample objects include dozens of sculptures and a number of lab tools. In the data acquisition, the projector sequentially projects 14 pre-generated images, and the camera synchronously captures 14 corresponding images. Specifically, the first 12 images are required by the conventional FPP method, as described previously; they are the four-step phase-shifted fringe images with three frequencies of 1, 8, and 80 fringes per image. Among the corresponding 12 captured fringe images, the four images with the highest fringe frequency serve as the first output labels of the network model. Meanwhile, the unwrapped phase map determined from the 12 capture images by using Equations (3) and (4) serves as the second output label. The input, on the other hand, is one of the aforementioned four images (e.g., the fourth one) with the highest fringe frequency. Therefore, the input to the network is a high-frequency fringe image, and the output includes the input image itself and three associated phase-shifted fringe images as well as a corresponding unwrapped phase map. Furthermore, since the numerator and denominator (ND) terms of the arctangent variable or the wrapped phase (WP) shown in Equation (3) may substitute the phase-shifted fringe images as the network's first output, they are obtained as well during the FPP processing for comparison purposes. Consequently, three input–output pairs are generated, where the datasets are the same except for the first output. Figure 3a demonstrates a few representative input–output pairs.

Along with the fringe-to-phase datasets, we also prepared a few image-to-depth datasets to compare the proposed method with other relevant deep learning-based techniques. For this reason, two additional projection images, i.e., the 13th and 14th images, are included in the dataset preparation. They include a random speckle pattern image and a uniform white image. In the image-to-depth datasets, the depth map is generated by using the first 12 phase-shifted fringe patterns and Equations (3)–(5), and each of the last three of the 14 captured images, i.e., the high-frequency fringe image, the speckle pattern, and the plain image, acts as an input for an image-to-depth network. Figure 3b displays some exemplars of the image input and depth-map output pairs. The datasets have been made temporarily accessible at [72].

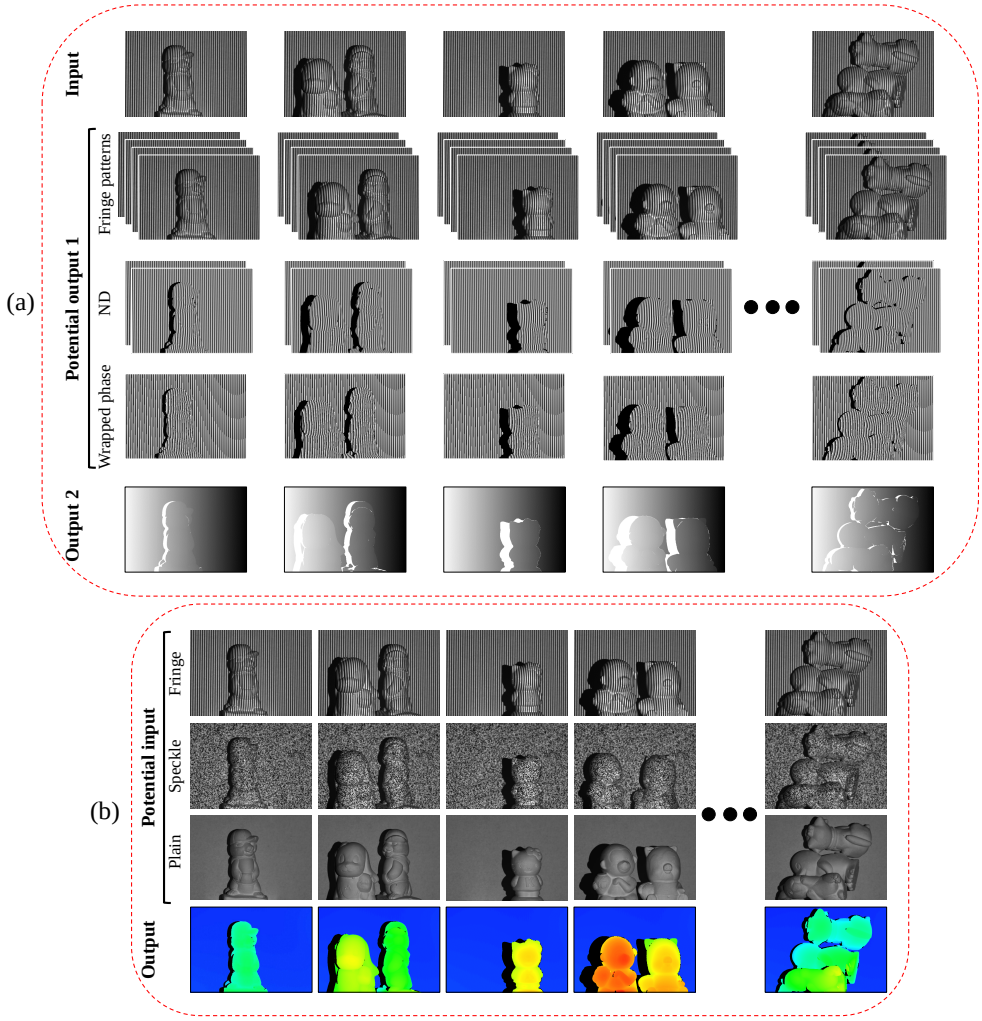

Figure 3. Demonstration of the input–output pairs for (**a**) fringe-to-phase and (**b**) fringe-to-depth approaches.

2.2. Single-Input Dual-Output Network for Fringe Image Transformation and Phase Retrieval

Determining the unwrapped phase map in the fringe-to-phase approach is critical despite different intermediate output selections. In the proposed approach, the unwrapped phase map can be determined by using $\phi = \phi_w + 2\pi k$, which is equivalent to Equation (4). Here, the wrapped phase ϕ_w is determined from the predicted first output of phase-shifted fringe patterns with Equation (3); the integer fringe order k is obtained by the second output (i.e., the predicted unwrapped phase ϕ') following $k = \text{INT}\left(\frac{\phi' - \phi_w}{2\pi}\right)$. It is noteworthy that the predicted ϕ' is not directly used as the true unwrapped phase for subsequent depth calculation because it is relatively noisy. Figure 4 demonstrates the pipeline of the proposed single-shot 3D reconstruction technique where the first part describes the employment of a deep learning network to predict two intermediate outputs, and the second part explains the subsequent 3D reconstruction with part of a classic FPP algorithm. Since the proposed fringe-to-phase approach transforms a single fringe-pattern image into two outputs, it is called a single-input dual-output (SIDO) network hereafter.

The SIDO network is adapted from the well-known autoencoder-based network, UNet [73]. The SIDO network preserves the prominent concatenation of the UNet but contains two decoder paths instead of one. The left portion of Figure 4 displays the network architecture. In the network, the encoder path and the first decoder path perform a pattern-to-pattern transformation where a single fringe pattern can be converted to four phase-shifted fringe patterns of the same frequency with an even phase-shifting increment

of $\pi/2$. The second learning path, including the same encoder and a different decoder, is trained to obtain an unwrapped phase map. The encoder path extracts local features from the input with ten convolution layers (a kernel size of 3×3) and four max-pooling layers (a window size of 2×2). After each max-pooling layer, the resolution of the feature maps is reduced by half, but the filter depth after each pair of convolution layers is doubled. In contrast, each decoder path consists of eight convolution layers and four transposed convolution layers. The input feature maps from the encoder path are enriched to higher resolution while decreasing the filter depths. The sequence of the filter depths in the encoder path is 32, 64, 128, 256, and 512, while the ones in both decoder paths are 256, 128, 64, and 32. In addition, symmetric concatenations between the encoder and decoder paths are employed to maintain the precise feature transformation from the input to the output. In particular, a 1×1 convolution layer with a filter size of 4 is attached to the end of the first decoder path to lead the internal feature maps to the corresponding four phase-shifted fringe images. Similarly, a 1×1 convolution layer (a filter size of 1) is appended after the second decoder path for the unwrapped phase prediction. Since both outputs contain continuous variables, a linear activation function and a common regression loss function, mean-squared error (MSE), are implemented for the training of the proposed fringe-to-phase framework. Importantly, a leaky rectified linear unit (LeakyReLU) function with a negative coefficient of 0.1 is applied over the convolution layers to avoid the zero-gradient problem. Moreover, a dropout function is added between the encoder path and the two decoder paths.

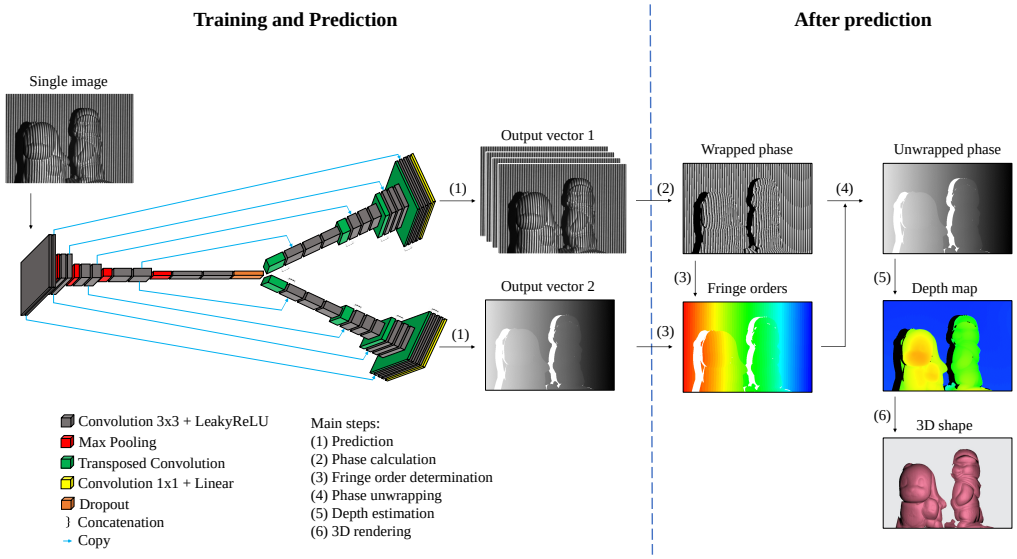

Figure 4. Pipeline of the proposed SIDO network and subsequent 3D reconstruction steps.

The multidimensional data format of the input, output, and internal hidden layers is a four-dimensional tensor of shape $[s, h, w, c]$ where s denotes the number of data samples; h and w represent the height and width of the input, output, or feature maps at the sub-scale resolution layer, respectively; c is the channel or filter depth. In this work, c is set to 1 for the input of a single grayscale image, and c for the two outputs are set to 4 and 1, corresponding to the four phase-shifted fringe images and the unwrapped phase map, respectively.

Using the backpropagation process, we trained the network parameters through 400 epochs using a mini-batch size of 1 or an equivalent stochastic gradient descent scheme. Adam optimizer with an initial learning rate of 0.0001 is set for the first 300 epochs. After that, a step decay schedule [74] is adopted to gradually reduce the learning rate for further convergence. Several data augmentation functions (e.g., ZCA whitening, brightness and contrast augmentation) are also adopted to prevent unwanted overfitting. Finally, some typical measures (e.g., MSE, RMSE) with Keras callbacks (e.g., History, ModelCheckpoint) are taken to monitor the history of the training process and save the best convergent model.

3. Experiments and Assessments

A series of experiments and analyses have been performed to validate the robustness and practicality of the proposed fringe-to-phase approach. An RVBUST RVC-X mini 3D camera, which offers a desired camera-projector-target triangulation setup, is employed to capture the datasets with a resolution of 640 × 448 pixels. The reduced resolution comes from directly cropping the full-size images (1920 × 1200 pixels). The use of reduced resolution is to accommodate the assigned computation resource, which includes multiple GPU nodes in the Biowulf cluster of the High Performance Computing group at the National Institutes of Health to accelerate the training process. The main graphics processing units (GPUs) include 4 × NVIDIA A100 GPUs 80GB VRAM, 4 × NVIDIA V100-SXM2 GPUs 32GB VRAM, and 4 × NVIDIA V100 GPUs 16GB VRAM. Furthermore, Nvidia CUDA Toolkit 11.0.3 and cuDNN v8.0.3 were installed to enhance the performance of the aforementioned units. TensorFlow and Keras, two popular open-source and user-friendly deep learning frameworks and Python libraries, were adopted to construct the network architecture and to use in data predictions and analysis.

3.1. Can the Predicted Unwrapped Phase Map Be Used Directly for 3D Reconstruction?

After a successful training process, the achieved network model is ready for an application test. It takes in a fringe-pattern image of the test target as a single input and produces two intermediate outputs: four phase-shifted fringe-pattern images and a coarse unwrapped phase map. As previously described, the obtained fringe patterns serve to calculate the accurate wrapped phase map, which is then applied together with the predicted coarse unwrapped phase to get the integer fringe orders. After that, the refined unwrapped phase distribution can be determined. This procedure is established as an equation as follows:

$$\begin{aligned} \phi_w &= \text{atan2}\frac{I_4^{(p)} - I_2^{(p)}}{I_1^{(p)} - I_3^{(p)}} \\ \phi &= \phi_w + 2\pi \cdot \text{INT}\left(\phi^{(p)} - \phi_w\right) \end{aligned} \tag{6}$$

where $I_1^{(p)}$–$I_4^{(p)}$ indicate the intensities of the network-predicted images and $\phi^{(p)}$ denotes the network-predicted coarse unwrapped phase.

The whole phase determination process using a single grayscale image and a single network is illustrated in Figure 5a. Two types of evaluation metrics, MSE and Structural Similarity Index Measure (SSIM), are employed to assess the performance of the proposed network in the fringe prediction task. Following the quantitative assessment shown in Figure 5a, the metrics reveal that the four predicted fringe patterns are very similar to the ground-truth fringe patterns; particularly, the last one has the smallest MSE error and the highest SSIM. This is expected since the last image ideally should be identical to the input.

As the second output is the predicted unwrapped phase, one question which can arise is the following: can the predicted unwrapped phase be employed directly for the 3D shape reconstruction following Equation (5)? In the extended investigation shown in Figure 5b,c, the predicted unwrapped phase $\phi^{(p)}$ and the refined phase ϕ are compared. It is clear that the errors in the refined phase map occur in small and limited areas, whereas the errors originating from the predicted unwrapped phase spread out to larger regions. The depth maps and the 3D shapes associated with both evaluated unwrapped phases are reconstructed to further demonstrate their performance. It is evident from the results that the predicted phase map results in an inconsistent depth map and noisy 3D shape, while the refined phase map from the proposed method can produce a high-accuracy 3D shape similar to the ground-truth label.

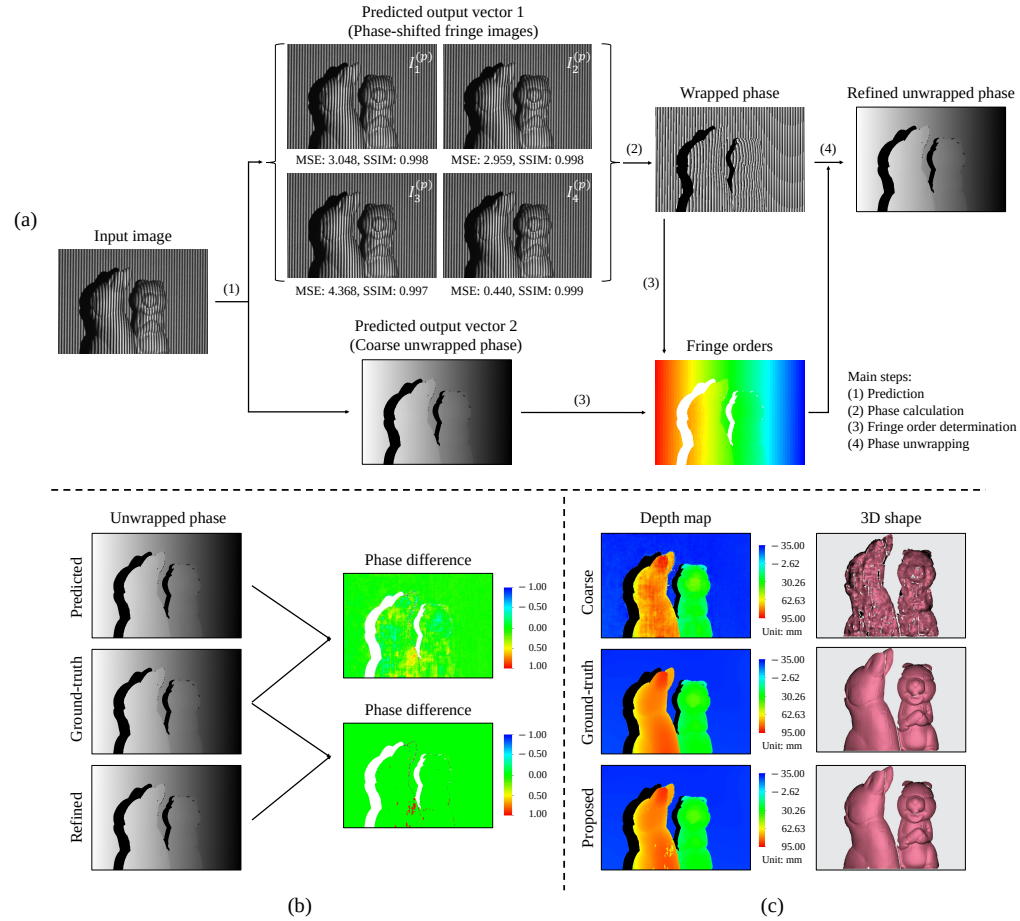

Figure 5. Qualitative and quantitative 3D measurement of a test sample. (**a**) Output prediction and phase determination process; (**b**) phase comparisons; and (**c**) 3D visualizations.

3.2. Full-Surface 360° 3D Reconstruction of a Single Object

Three-dimensional full-surface reconstruction is a classic yet challenging task in computer vision and relevant fields because it demands high-accuracy 3D point clouds. An acquisition of multi-view 3D surface profiles has been conducted to demonstrate the capability of our proposed technique. In this experiment, an untrained object (i.e., never used in the network training) has been rotated and positioned randomly for capturing. Figure 6 demonstrates eight single-view 3D shape reconstructions of the test object. The first two columns are the plain images and the input images, followed by the phase differences between the ground-truth phase and the refined unwrapped phase distributions. The last three columns display the ground-truth 3D shape, the direct 3D reconstruction of the proposed technique, and the 3D shape after post-processing, respectively. In general, major errors in phase difference are observed along the edges where the fringes could not be illuminated effectively. This results in some holes on the surfaces of the 3D shapes that can be easily erased or fixed in a post-processing step thanks to the accurately reconstructed points surrounding them. The final 3D shapes obtained from our proposed 3D reconstruction are nearly identical to the ground-truth 3D acquired by the conventional temporal-based FPP technique. Finally, a 360° full-surface 3D image is built after data alignment and stitching (see Supplementary Video S1).

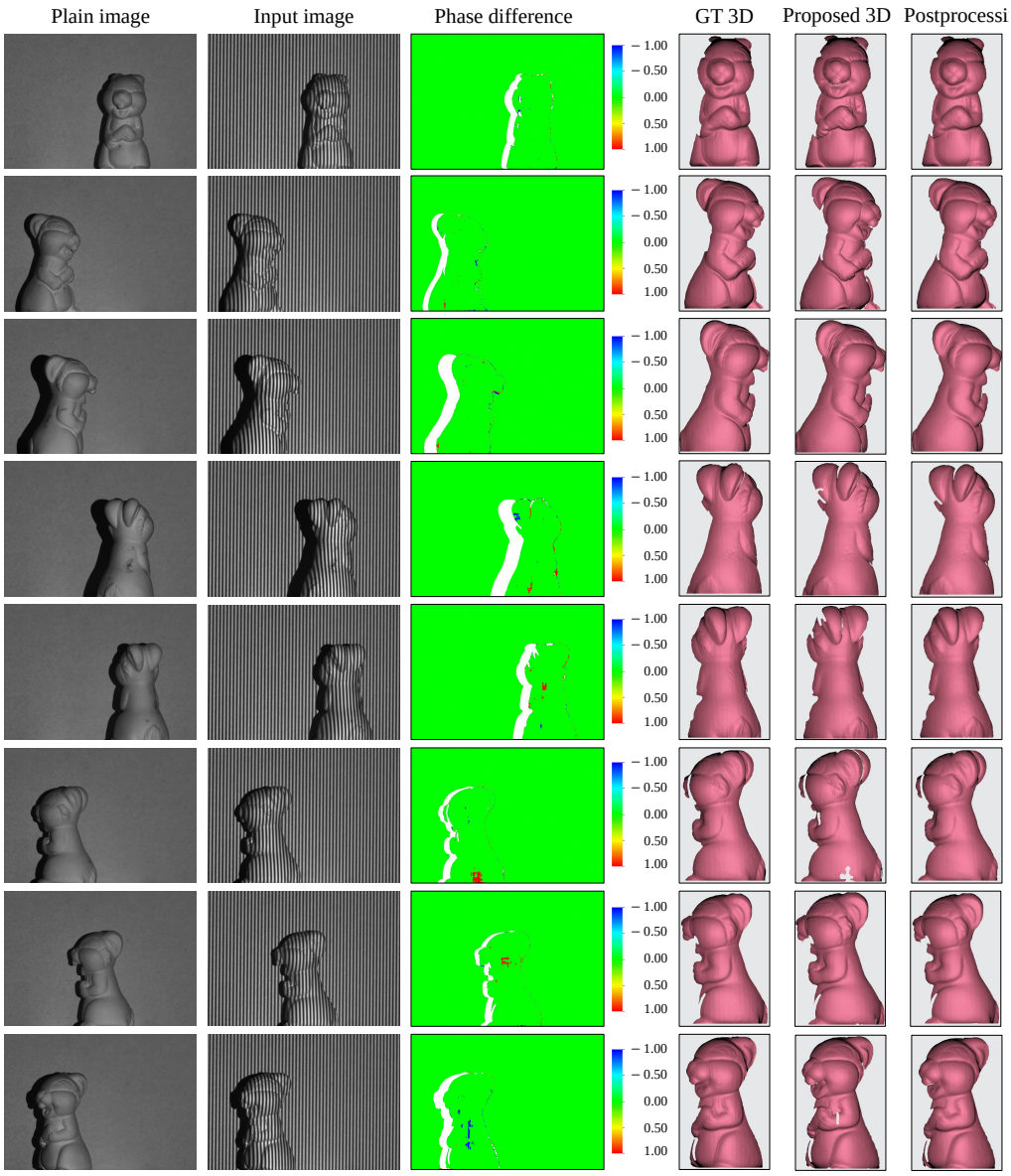

Figure 6. Full-surface 360° 3D shape reconstruction of an object.

3.3. Three-Dimensional Reconstruction of Multiple Objects

In real-world applications, the target of interest may include multiple separated objects, which typically bring an issue of geometric discontinuity. In this experiment, several untrained objects were grouped and randomly located in the scene to verify the capability of the proposed technique. Figure 7 shows the 3D reconstruction results. It is evident that the SIDO network can generally handle multiple objects well, and the reconstructed 3D surface profiles are close to the ground-truth 3D shapes with sub-millimeter errors. In the meantime, it can be seen that there are more errors than in the case of a single object. This is reasonable since there are more complex edges with multiple separated objects. These errors present as holes or stay close to the edges, and they can be easily fixed or eliminated by post-processing.

3.4. Three-Dimensional Reconstruction of Objects with Varied Colors

To further explore the applicability of the proposed method, we proceed with the testings using a few complex targets with various colors and contrasts. Figure 8 presents the objects and their 3D shape reconstruction results. The first and second columns display the plain images and the input images, respectively, followed by the determined depth

maps, the ground-truth 3D shapes, and the reconstructed 3D shapes. Our method can clearly produce high-quality 3D shape reconstructions for complex objects, as shown by a comparison with the ground-truth shapes in the first, second, and fourth rows. It is noted that evident errors emerge in the third object because that object contains intricate textures and more shaded regions, making the 3D reconstruction more challenging even for the conventional technique.

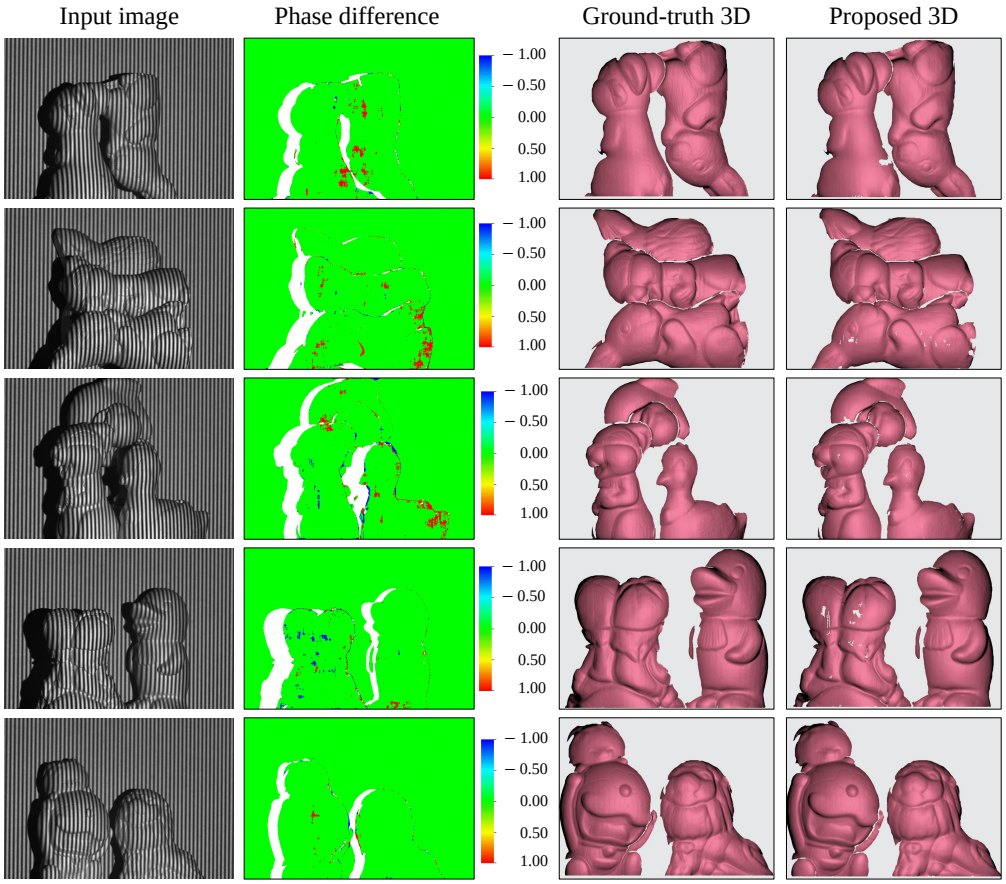

Figure 7. Three-dimensional shape reconstruction of multiple separated objects.

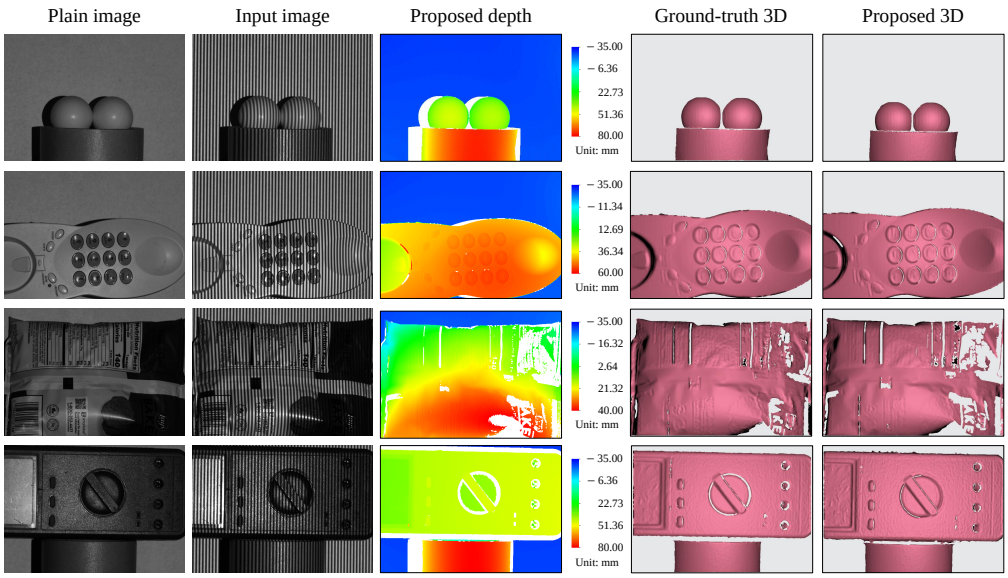

Figure 8. Three-dimensional shape reconstruction of different objects with a variety of colors and contrasts.

3.5. Accuracy Comparison of the Proposed 3D Reconstruction Technique with Other Fringe-to-Phase and Fringe-to-Depth Methods

In recent years, the integration of structured light with deep learning has received ever-increasing attention. To demonstrate the efficacy of the proposed techniques, we attempt to compare the performance of our technique with five other existing state-of-the-art techniques, classified into fringe-to-phase and image-to-depth groups. In the fringe-to-phase group, the second output of unwrapped phase remains unchanged while the first output is replaced with either phase-shifted fringe images (proposed method), numerator and denominator (fringe-to-ND), or wrapped phase map (fringe-to-WP). For the image-to-depth technique, since it is a direct transformation from a grayscale image into a depth map, the depth map is fixed and only the grayscale input image is replaced among the fringe (fringe-to-depth), speckle (speckle-to-depth), and plain images (plain-to-depth). These image-to-depth techniques are trained based on the adopted autoencoder-based network with multi-scale feature fusion [75]. It is important to note that since the primary comparison in this experiment only takes a single grayscale image as input, other techniques that require a composite red-green-blue image [53,54,57,58], multiple images [11,41,51,52,62,64], or a reference image [34,40,55] are not considered. Several error and accuracy metrics commonly used for the evaluation of monocular depth reconstruction are adopted here. The quantitative assessments contain the following four error metrics and three accuracy metrics:

- Absolute relative error (rel): $\frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{z}_i - z_i|}{\hat{z}_i}$;
- Root-mean-square error (rms): $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{z}_i - z_i)^2}$;
- Average log_{10} error (log): $\frac{1}{n}\sum_{i=1}^{n}\left|log_{10}(\hat{z}_i) - log_{10}(z_i)\right|$;
- Root-mean-square log error (rms log): $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|log_{10}(\hat{z}_i) - log_{10}(z_i)\right|}$;
- Threshold accuracy: $\delta = (\frac{\hat{z}_i}{z_i}, \frac{z_i}{\hat{z}_i}) < thr$; $thr \in \left\{1.25, 1.25^2, 1.25^3\right\}$.

Figure 9 shows the corresponding 3D reconstructions of the proposed, fringe-to-phase, and image-to-depth techniques. It is clear that the three fringe-to-phase techniques outperform the image-to-depth counterparts because the intermediate results help retain more feature information. It can also be seen that several holes and incorrect shape regions appear in the fringe-to-ND and fringe-to-WP results, which is probably due to the misalignment between predicted outputs and the obtained fringe orders. It is also shown that the speckle pattern is not a desired option for the image-to-depth network [36] since surface textures can be concealed by random pattern (unlike regular fringe pattern). Furthermore, the plain-to-depth technique obtains noisy 3D results despite being capable of yielding discernible textures. Table 1 summarizes the numerical errors and accuracy metrics of the proposed and other techniques. The first three rows from Table 1 reveal that the accuracy performances of the proposed and the fringe-to-ND techniques are similar, while the fringe-to-WP approach offers a slightly lower performance.

Table 1. Quantitative performance comparison of different fringe-to-phase and image-to-depth techniques.

Method	Error (Lower Is Better)				Accuracy (Higher Is Better)		
	rel	**rms**	**log**	**rms log**	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Proposed method	0.004	0.385	0.003	0.050	99.5%	99.8%	99.9%
Fringe-to-ND	0.005	0.392	0.003	0.049	99.5%	99.8%	99.9%
Fringe-to-WP	0.007	0.610	0.004	0.061	99.2%	99.7%	99.8%
Fringe-to-depth	0.023	0.916	0.015	0.119	97.7%	99.1%	99.5%
Speckle-to-depth	0.027	0.923	0.016	0.120	98.0%	99.2%	99.6%
Plain-to-depth	0.031	1.104	0.020	0.134	96.8%	98.9%	99.4%

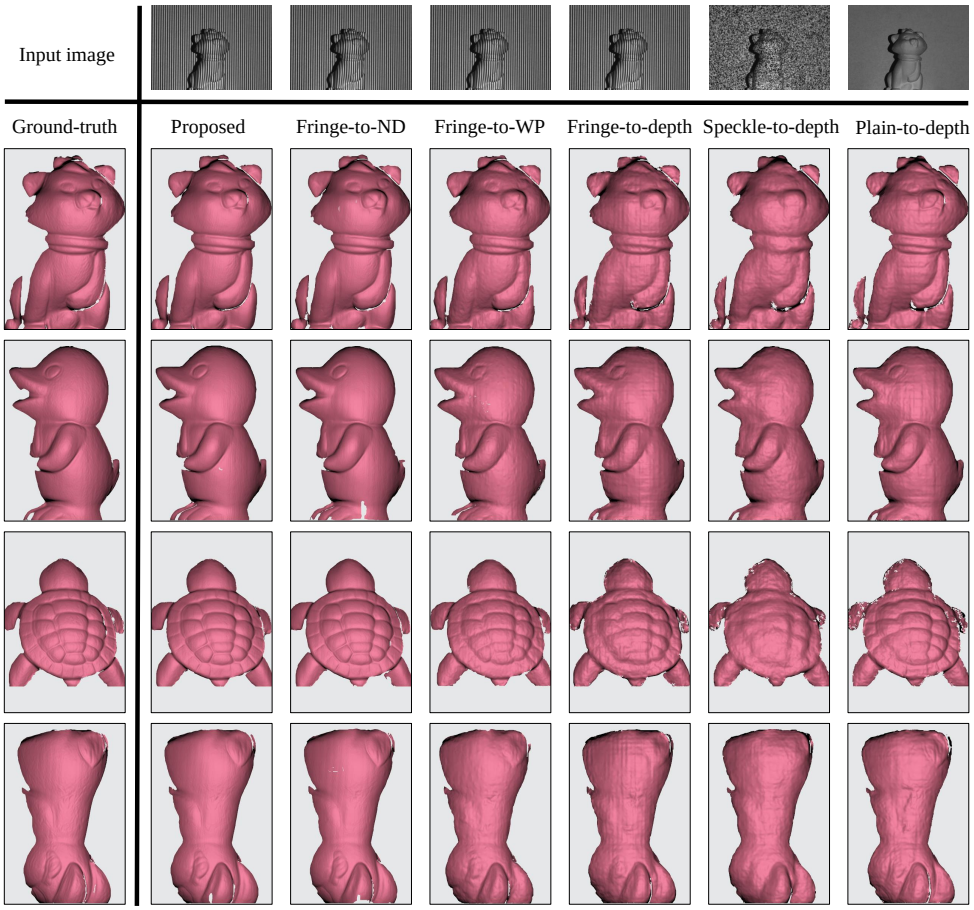

Figure 9. Qualitative comparisons between the proposed and other deep learning-based structured-light techniques.

4. Discussions

This work explores a supervised learning-based approach to intervening in the classic FPP-based technique for single-shot 3D shape reconstruction. The proposed technique uses a single fringe image and a SIDO network to acquire two intermediates for subsequent accurate 3D shape reconstruction. The well-known conventional FPP technique is introduced and implemented to generate high-quality datasets for training the deep-learning network. Compared with many other techniques in the same category, the proposed technique uses only a single grayscale fringe image as input instead of requiring multiple images, a composite RGB image, or a reference image. This makes the proposed technique very appealing for numerous engineering applications where multi-shot capturing is undesired. Such applications typically involves dynamic motions, such as robotic navigation, automatic palletization and material handling in warehouse, in-line inspection of products during manufacturing, virtual/augmented reality, and 3D human body scanning.

Even though the proposed technique closely resembles the well-known classic FPP technique for accurate 3D shape reconstruction, the results can be partially sketchy because of the dependence on the supervised-learning datasets. As the phase distributions depend on not only the actual object profiles but also the geometric setup of the imaging system, the trained model is valid only for the specific system configuration used in dataset generation. If the relative geometric configuration between the camera and projector changes, recapturing new datasets for training is imperative. For this reason, generating synthetic datasets based on the calibration parameters of the actual system for network training can be a favorable solution [38,45]. Nevertheless, further investigation is necessary to make such a network model reliable for real-world applications.

The misalignment between two outputs can create errors in the phase map that can lead to incorrect or invalid height/depth. In addition, the unbalanced intensities of the predicted phase-shifted fringe patterns near the edges can construct observable errors there. Despite that, a post-processing step can easily eliminate the wrong points and readily fill the holes without noticeable issues thanks to the accurate reconstruction of the neighboring points.

The series of experiments conducted have verified the validity and robustness of the proposed method for 3D shape reconstructions. The experiments also demonstrate that the fringe-to-phase methods can yield better 3D shapes than the image-to-depth ones, thanks to the feature-preserving characteristics of the intermediate outputs. Although the fringe-to-phase approach requires extra computation time (typically a few to tens of milliseconds) to calculate the 3D point clouds from the predicted outputs, it is worth the wait since 3D shapes with high-quality texture details can be achieved. In addition, the extra calculation can be accelerated by parallel processing with the TensorFlow framework. It is noticed that using the numerator and denominator (fringe-to-ND) instead of using four phase-shifted fringe images as the first output is feasible to save storage space without perceptibly affecting the final 3D results. Lastly, it is noteworthy that the predicted unwrapped phase can be directly used to reconstruct the 3D shape with evident errors; however, we believe that the results can be considerably improved upon exploring a more suitable network with meticulously tuned learning hyperparameters and network parameters.

5. Conclusions

In summary, the paper presents an innovative 3D shape reconstruction technique integrating the classic FPP technique with deep learning. The technique requires only a single concise network, and it takes a single fringe image as the input. An advanced autoencoder-based network inspired by the UNet has been implemented to convert the single input image into two immediate outputs of four phase-shifted fringe patterns and a coarse unwrapped phase map. These outputs can yield a refined accurate unwrapped phase map, which can then be used to determine the depth map and reconstruct the 3D shapes. The key advantage of the proposed technique lies in using a single image for 3D imaging and shape reconstruction, thus, the capturing speed can be maximized; At the same time, using network-predicted multiple images for subsequent calculations helps preserve the high-accuracy nature of the result. Therefore, both fast speed and high accuracy can be achieved for the 3D shape reconstruction, which provides a promising tool in numerous scientific and engineering applications.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s23094209/s1, Video S1: 360° full-surface 3D image of an object.

Author Contributions: Conceptualization, A.-H.N., V.K.L. and Z.W.; methodology, A.-H.N. and Z.W.; software, A.-H.N. and Z.W.; validation, A.-H.N., K.L.L. and V.K.L.; formal analysis, A.-H.N. and V.K.L.; investigation, A.-H.N. and K.L.L.; resources, A.-H.N. and K.L.L.; data curation, A.-H.N. and Z.W.; writing—original draft preparation, A.-H.N., K.L.L. and V.K.L.; writing—review and editing, A.-H.N. and Z.W.; visualization, A.-H.N. and V.K.L.; supervision, Z.W.; project administration, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov, accessed on 14 March 2023).

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

References

1. Su, X.; Zhang, Q. Dynamic 3-D shape measurement method: A review. *Opt. Lasers Eng.* **2010**, *48*, 191–204. [CrossRef]
2. Bruno, F.; Bruno, S.; Sensi, G.; Luchi, M.; Mancuso, S.; Muzzupappa, M. From 3D reconstruction to virtual reality: A complete methodology for digital archaeological exhibition. *J. Cult. Herit.* **2010**, *11*, 42–49. [CrossRef]
3. Huang, S.; Xu, K.; Li, M.; Wu, M. Improved Visual Inspection through 3D Image Reconstruction of Defects Based on the Photometric Stereo Technique. *Sensors* **2019**, *19*, 4970. [CrossRef] [PubMed]
4. Bennani, H.; McCane, B.; Corwall, J. Three-dimensional reconstruction of In Vivo human lumbar spine from biplanar radiographs. *Comput. Med. Imaging Graph.* **2022**, *96*, 102011. [CrossRef] [PubMed]
5. Do, P.; Nguyen, Q. A Review of Stereo-Photogrammetry Method for 3-D Reconstruction in Computer Vision. In Proceedings of the 19th International Symposium on Communications and Information Technologies, Ho Chi Minh City, Vietnam, 25–27 September 2019; pp. 138–143. [CrossRef]
6. Blais, F. Review of 20 years of range sensor development. *J. Electron. Imaging* **2004**, *13*, 231–243. [CrossRef]
7. Chen, F.; Brown, G.; Song, M. Overview of threedimensional shape measurement using optical methods. *Opt. Eng.* **2000**, *39*, 10–22. [CrossRef]
8. Bianco, G.; Gallo, A.; Bruno, F.; Muzzuppa, M. A Comparative Analysis between Active and Passive Techniques for Underwater 3D Reconstruction of Close-Range Objects. *Sensors* **2013**, *13*, 11007–11031. [CrossRef]
9. Khilar, R.; Chitrakala, S.; Selvamparvathy, S. 3D image reconstruction: Techniques, applications and challenges. In Proceedings of the 2013 International Conference on Optical Imaging Sensor and Security, Coimbatore, India, 2–3 July 2013; pp. 1–6. [CrossRef]
10. Zhang, S. High-speed 3D shape measurement with structured light methods: A review. *Opt. Lasers Eng.* **2018**, *106*, 119–131. [CrossRef]
11. Nguyen, H.; Ly, K.; Nguyen, T.; Wang, Y.; Wang, Z. MIMONet: Structured-light 3D shape reconstruction by a multi-input multi-output network. *Appl. Opt.* **2021**, *60*, 5134–5144. [CrossRef]
12. Prusak, A.; Melnychuk, O.; Roth, H.; Schiller, I.; Koch, R. Pose estimation and map building with a Time-Of-Flight-camera for robot navigation. *Int. J. Intell. Syst. Technol. Appl.* **2008**, *5*, 355–364. [CrossRef]
13. Kolb, A.; Barth, E.; Koch, R.; Larsen, R. Time-of-Flight Sensors in Computer Graphics. In Proceedings of the Eurographics 2009—State of the Art Reports, Munich, Germany, 30 March–3 April 2009. [CrossRef]
14. Kahn, S.; Wuest, H.; Fellner, D. Time-of-flight based Scene Reconstruction with a Mesh Processing Tool for Model based Camera Tracking. In Proceedings of the International Conference on Computer Vision Theory and Applications—Volume 1: VISAPP, Angers, France, 17–21 May 2010; pp. 302–309. [CrossRef]
15. Kim, D.; Lee, S. Advances in 3D Camera: Time-of-Flight vs. Active Triangulation. In Proceedings of the Intelligent Autonomous Systems 12. Advances in Intelligent Systems and Computing, Jeju Island, Republic of Korea, 26–29 June 2012; Volume 193, pp. 301–309. [CrossRef]
16. Geng, J. Structured-light 3D surface imaging: A tutorial. *Adv. Opt. Photonics* **2011**, *3*, 128–160. [CrossRef]
17. Jeught, S.; Dirckx, J. Real-time structured light profilometry: A review. *Sensors* **2016**, *87*, 18–31. [CrossRef]
18. Fernandez, S.; Salvi, J.; Pribanic, T. Absolute phase mapping for one-shot dense pattern projection. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010; pp. 128–144. [CrossRef]
19. Moreno, D.; Taubin, G. Simple, Accurate, and Robust Projector-Camera Calibration. In Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, Zurich, Switzerland, 13–15 October 2012; pp. 464–471. [CrossRef]
20. Jensen, J.; Hannemose, M.; Bærentzen, A.; Wilm, J.; Frisvad, J.; Dahl, A. Surface Reconstruction from Structured Light Images Using Differentiable Rendering. *Sensors* **2021**, *21*, 1068. [CrossRef] [PubMed]
21. Tran, V.; Lin, H.Y. A Structured Light RGB-D Camera System for Accurate Depth Measurement. *Int. J. Opt.* **2018**, *2018*, 8659847. [CrossRef]
22. Diba, A.; Sharma, V.; Pazandeh, A.; Pirsiavash, H.; Gool, L. Weakly Supervised Cascaded Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5131–5139. [CrossRef]
23. Doulamis, N. Adaptable deep learning structures for object labeling/tracking under dynamic visual environments. *Multimed. Tools. Appl.* **2018**, *77*, 9651–9689. [CrossRef]
24. Toshev, A.; Szegedy, C. DeepPose: Human Pose Estimation via Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1653–1660. [CrossRef]
25. Lin, L.; Wang, K.; Zuo, W.; Wang, M.; Luo, J.; Zhang, L. A Deep Structured Model with Radius–Margin Bound for 3D Human Activity Recognition. *Int. J. Comput. Vis.* **2016**, *118*, 256–273. [CrossRef]
26. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 13. [CrossRef]
27. Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Velasco-Hernandez, G.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep Learning vs. Traditional Computer Vision. In Proceedings of the 2019 Computer Vision Conference (CVC), Las Vegas, NV, USA, 2–3 May 2019; pp. 128–144. [CrossRef]

28. Han, X.F.; Laga, H.; Bennamoun, M. Image-Based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1578–1604. [CrossRef]
29. Fu, K.; Peng, J.; He, Q.; Zhang, H. Single image 3D object reconstruction based on deep learning: A review. *Multimed. Tools Appl.* **2020**, *80*, 463–498. [CrossRef]
30. Zhang, Y.; Liu, Z.; Liu, T.; Peng, B.; Li, X. RealPoint3D: An Efficient Generation Network for 3D Object Reconstruction From a Single Image. *IEEE Access* **2019**, *7*, 57539–75749. [CrossRef]
31. Jeught, S.; Dirckx, J. Deep neural networks for single shot structured light profilometry. *Opt. Express* **2019**, *27*, 17091–17101. [CrossRef] [PubMed]
32. Yang, G.; Wang, Y. Three-dimensional measurement of precise shaft parts based on line structured light and deep learning. *Measurement* **2022**, *191*, 110837. [CrossRef]
33. Guan, J.; Li, J.; Yang, X.; Chen, X.; Xi, J. Defect detection method for specular surfaces based on deflectometry and deep learning. *Opt. Eng.* **2022**, *61*, 061407. [CrossRef]
34. Li, J.; Zhang, Q.; Zhong, L.; Lu, X. Hybrid-net: A two-to-one deep learning framework for three-wavelength phase-shifting interferometry. *Opt. Express* **2021**, *29*, 34656–34670. [CrossRef]
35. Yan, K.; Yu, Y.; Huang, C.; Sui, L.; Qian, K.; Asundi, A. Fringe pattern denoising based on deep learning. *Opt. Commun.* **2019**, *437*, 148–152. [CrossRef]
36. Nguyen, H.; Tran, 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]
37. Zhu, X.; Han, Z.; Song, L.; Wang, H.; Wu, Z. Wavelet based deep learning for depth estimation from single fringe pattern of fringe projection profilometry. *Optoelectron. Lett.* **2022**, *18*, 699–704. [CrossRef]
38. 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]
39. Jia, T.; Liu, Y.; Yuan, X.; Li, W.; Chen, D.; Zhang, Y. Depth measurement based on a convolutional neural network and structured light. *Meas. Sci. Technol.* **2022**, *33*, 025202. [CrossRef]
40. Machineni, R.; Spoorthi, G.; Vengala, K.; Gorthi, S.; Gorthi, R. End-to-end deep learning-based fringe projection framework for 3D profiling of objects. *Comput. Vis. Image Underst.* **2020**, *199*, 103023. [CrossRef]
41. 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*, 32547–32567. [CrossRef] [PubMed]
42. Wang, L.; Lu, D.; Qiu, R.; Tao, J. 3D reconstruction from structured-light profilometry with dual-path hybrid network. *EURASIP J. Adv. Signal Process.* **2022**, *2022*, 14. [CrossRef]
43. Nguyen, A.; Sun, B.; Li, C.; Wang, Z. Different structured-light patterns in single-shot 2D-to-3D image conversion using deep learning. *Appl. Opt.* **2022**, *61*, 10105–10115. [CrossRef] [PubMed]
44. 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]
45. Zheng, Y.; Wang, S.; Li, Q.; Li, B. Fringe projection profilometry by conducting deep learning from its digital twin. *Opt. Express* **2020**, *28*, 36568–36583. [CrossRef]
46. Wang, L.; Lu, D.; Tao, J.; Qiu, R. Single-shot structured light projection profilometry with SwinConvUNet. *Opt. Eng.* **2022**, *61*, 114101. [CrossRef]
47. Nguyen, H.; Nicole, D.; Li, H.; Wang, Y.; Wang, Z. Real-time 3D shape measurement using 3LCD projection and deep machine learning. *Apt. Opt.* **2019**, *58*, 7100–7109. [CrossRef] [PubMed]
48. Yu, H.; Chen, X.; Huang, R.; Bai, L.; Zheng, D.; Han, J. Untrained deep learning-based phase retrieval for fringe projection profilometry. *Opt. Lasers Eng.* **2023**, *164*, 107483. [CrossRef]
49. Wang, J.; Li, Y.; Ji, Y.; Qian, J.; Che, Y.; Zuo, C.; Chen, Q.; Feng, S. Deep Learning-Based 3D Measurements with Near-Infrared Fringe Projection. *Sensors* **2022**, *22*, 6469. [CrossRef]
50. Xu, M.; Zhang, Y.; Wang, N.; Luo, L.; Peng, J. Single-shot 3D shape reconstruction for complex surface objects with colour texture based on deep learning. *J. Mod. Opt.* **2022**, *69*, 941–956. [CrossRef]
51. Yu, H.; Chen, X.; Zhang, Z.; Zuo, C.; Zhang, Y.; Zheng, D.; Han, J. Dynamic 3-D measurement based on fringe-to-fringe transformation using deep learning. *Opt. Express* **2020**, *28*, 9405–9418. [CrossRef]
52. Yang, Y.; Hou, Q.; Li, Y.; Cai, Z.; Liu, X.; Xi, J.; Peng, X. Phase error compensation based on Tree-Net using deep learning. *Opt. Lasers Eng.* **2021**, *143*, 106628. [CrossRef]
53. Nguyen, H.; Wang, Z. Accurate 3D Shape Reconstruction from Single Structured-Light Image via Fringe-to-Fringe Network. *Photonics* **2021**, *8*, 459. [CrossRef]
54. Nguyen, A.; Ly, K.; Li, C.; Wang, Z. Single-shot 3D shape acquisition using a learning-based structured-light technique. *Appl. Opt.* **2022**, *61*, 8589–8599. [CrossRef]
55. 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]
56. 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* **2022**, *30*, 3424–3442. [CrossRef] [PubMed]

57. Zhang, B.; Lin, S.; Lin, J.; Jiang, K. Single-shot high-precision 3D reconstruction with color fringe projection profilometry based BP neural network. *Opt. Commun.* **2022**, *517*, 128323. [CrossRef]
58. Nguyen, H.; Novak, E.; Wang, Z. Accurate 3D reconstruction via fringe-to-phase network. *Measurement* **2022**, *190*, 110663. [CrossRef]
59. Liang, J.; Zhang, J.; Shao, J.; Song, B.; Yao, B.; Liang, R. Deep Convolutional Neural Network Phase Unwrapping for Fringe Projection 3D Imaging. *Sensors* **2020**, *20*, 3691. [CrossRef] [PubMed]
60. Shi, J.; Zhu, X.; Wang, H.; Song, L.; Guo, Q. Label enhanced and patch based deep learning for phase retrieval from single frame fringe pattern in fringe projection 3D measurement. *Opt. Express* **2019**, *27*, 28929–28943. [CrossRef]
61. Yin, W.; Chen, Q.; Feng, S.; Tao, T.; Huang, L.; Trusiak, M.; Asundi, A.; Zuo, C. Temporal phase unwrapping using deep learning. *Sci. Rep.* **2019**, *9*, 20175. [CrossRef]
62. Qian, J.; Feng, S.; 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]
63. Yu, H.; Han, B.; Bai, L.; Zheng, D.; Han, J. Untrained deep learning-based fringe projection profilometry. *APL Photonics* **2022**, *7*, 016102. [CrossRef]
64. Yao, P.; Gai, S.; Chen, Y.; Chen, W.; Da, F. A multi-code 3D measurement technique based on deep learning. *Opt. Lasers Eng.* **2021**, *143*, 106623. [CrossRef]
65. Li, W.; Yu, J.; Gai, S.; Da, F. Absolute phase retrieval for a single-shot fringe projection profilometry based on deep learning. *Opt. Eng.* **2021**, *60*, 064104. [CrossRef]
66. Nguyen, A.; Rees, O.; Wang, Z. Learning-based 3D imaging from single structured-light image. *Graph. Models* **2023**, *126*, 101171. [CrossRef]
67. Nguyen, H.; Nguyen, D.; Wang, Z.; Kieu, H.; Le, M. Real-time, high-accuracy 3D imaging and shape measurement. *Appl. Opt.* **2015**, *54*, A9–A17. [CrossRef]
68. Nguyen, H.; Liang, J.; Wang, Y.; Wang, Z. Accuracy assessment of fringe projection profilometry and digital image correlation techniques for three-dimensional shape measurements. *J. Phys. Photonics* **2021**, *3*, 014004. [CrossRef]
69. Le, H.; Nguyen, H.; Wang, Z.; Opfermann, J.; Leonard, S.; Krieger, A.; Kang, J. Demonstration of a laparoscopic structured-illumination three-dimensional imaging system for guiding reconstructive bowel anastomosis. *J. Biomed. Opt.* **2018**, *23*, 056009. [CrossRef]
70. Du, H.; Wang, Z. Three-dimensional shape measurement with an arbitrarily arranged fringe projection profilometry system. *Opt. Lett.* **2007**, *32*, 2438–2440. [CrossRef]
71. Vo, M.; Wang, Z.; Pan, B.; Pan, T. Hyper-accurate flexible calibration technique for fringe-projection-based three-dimensional imaging. *Opt. Express* **2012**, *20*, 16926–16941. [CrossRef]
72. Single-Input Dual-Output 3D Shape Reconstruction. Available online: https://figshare.com/s/c09f17ba357d040331e4 (accessed on 13 April 2023).
73. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
74. Keras. ExponentialDecay. Available online: https://keras.io/api/optimizers/learning_rate_schedules/ (accessed on 13 April 2023).
75. Nguyen, H.; Ly, K.L.; Tran, T.; Wang, Y.; Wang, Z. hNet: Single-shot 3D shape reconstruction using structured light and h-shaped global guidance network. *Results Opt.* **2021**, *4*, 100104. [CrossRef]