Sim-to-Real Metrology: Calibrated Digital Twins for Fringe Projection Profilometry

Noble Austine[1], Eshwar Sai Vuppu[2], Vaishnavi Ravi[3],
Madhu S. Nair[1], Rama Krishna Sai Subrahmanyam Gorthi[2,*]

[1]Artificial Intelligence & Computer Vision Lab, Department of Computer Science,
Cochin University of Science and Technology, Kochi, India.
[2]Computer Vision Lab, Department of Electrical Engineering, Indian Institute of Technology Tirupati, India.
[3]Medical Imaging Group, Indian Institute of Science Bangalore, India.
{nobleaustine, msn}@cusat.ac.in {ee24m308, rkg}@iittp.ac.in vaishnavi1712@gmail.com

Abstract

While Fringe Projection Profilometry (FPP) is renowned for high-precision 3D measurements, acquiring dense absolute depth maps from a single deformed fringe image remains a fundamentally ill-posed problem. Deep learning offers a transformative solution for single-shot reconstruction; however, its potential is severely bottlenecked by the lack of large-scale, physically accurate training datasets with pixel-perfect ground truth. To bridge this critical sim-to-real gap, this paper introduces a highly calibrated digital twin methodology for synthetic data generation. By injecting authentic sensor noise and real-world radiometric characteristics into a precise virtual replica of a physical FPP setup, we construct ***FPP-Depth150****, a comprehensive hybrid dataset comprising over 21,600 diverse fringe-height pairs. Furthermore, we propose a novel baseline that bridges the paradigms of Monocular Depth Estimation (MDE) and active-light FPP. By leveraging the robust DINOv2 Vision Transformer (ViT) backbone from the Depth Anything model, our architecture harmonizes the strengths of both domains: the MDE foundation provides a deep understanding of global shape and complex geometric features, while the active FPP data grounds these features with the precise, high-frequency scale required for absolute metrology. Our evaluations demonstrate that models trained exclusively on our digital twin dataset successfully generalize to challenging real-world data, establishing a robust new benchmark FPP dataset and algorithm development for bridging foundation models and optical metrology.*

*Corresponding author

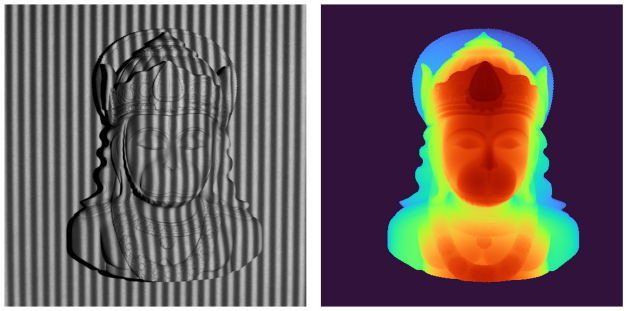

Figure 1. A sample data pair from the proposed FPP-Depth150 dataset: a single-shot deformed fringe pattern (left) and its corresponding high-precision absolute depth map (right).

1. Introduction

Fringe Projection Profilometry (FPP) is a non-destructive optical method enabling high-resolution 3D surface acquisition for industrial quality control, cultural preservation, and medical diagnostics [33]. While multi-shot Phase-Shifting Profilometry (PSP) yields high precision, its vulnerability to dynamic motion artifacts has driven demand for real-time, single-shot FPP. However, analytically extracting absolute depth from a single deformed fringe is a challenging problem. Classical single-shot pipelines, like Fourier Transform Profilometry (FTP), degrade rapidly under deep shadows or high-dynamic-range (HDR) reflections and require computationally expensive, noise-sensitive phase unwrapping.

Although deep learning offers a transformative, end-to-end alternative [14, 23, 24], progress is critically constrained by a severe lack of standardized training data. Relying on localized, proprietary physical captures prevents fair baseline comparisons and decelerates the field.

Table 1. Comparison of publicly available FPP datasets and the proposed benchmark.

Dataset	Size / Objects	Data Type	Target Task	Acquisition Source	Primary Limitation
SynthFringe [11]	18,000 / 500	Fringe & Wrapped Phase	Wrapped Phase	Synthetic (Blender)	Pure mathematical projection; lacks real-world sensor noise
FPP-ML-Bench [12]	15,600 / 50	Fringe & Depth Map	Absolute Depth	Synthetic (Isaac Sim)	Limited geometric diversity; low object count
HDR-NET [3]	720 / 14	Fringe Images	Phase Retrieval	Real Capture	Extremely small scale; strictly HDR-specific focus
HiPhase [1]	16,884 / -	Wrapped Phase Map	Phase Unwrapping	Real Capture	Restricted scope; no end-to-end depth regression
FPP-Depth150 (Ours)	**21,600 / 150**	**Fringe & Depth Map**	**Absolute Depth**	**Hybrid Digital Twin**	**Currently for limited dynamic range**

To resolve this critical gap, we propose **FPP-Depth150**, a large-scale, highly calibrated synthetic benchmark for optical metrology (see Fig. 1). Our core contributions are:

- A novel sim-to-real digital twin pipeline that injects authentic physical sensor noise and projector non-linearities into procedurally rendered 3D environments.
- The release of **FPP-Depth150**, providing 21,600 paired deformed fringes and absolute depth maps that capture challenging HDR reflectances and severe self-occlusions.
- A robust baseline bridging Monocular Depth Estimation (MDE) and active-light FPP. By adapting the DINOv2 backbone from Depth Anything, we leverage strong MDE shape priors and ground them with precise, absolute metrological scale.

2. Related Work

2.1. Deep Learning for FPP

Deep learning integration in FPP has evolved rapidly from modular pipeline enhancements to end-to-end architectures. Early works successfully replaced individual processing stages, utilizing CNNs and U-Nets for direct phase extraction [5, 20], phase-shifted sequence mapping [29], and wrapped phase unwrapping [8]. Recent state-of-the-art architectures bypass the intermediate phase domain entirely, employing encoder-decoder networks to regress absolute 3D depth directly from single deformed fringes. To handle complex object geometries, recent models incorporate multi-frequency multiplexing [11], hybrid temporal phase unwrapping [13], and transformer architectures designed specifically for shadow mitigation [19].

Despite these algorithmic strides, a critical bottleneck persists: acquiring precise ground truth for supervised learning still requires traditional multi-shot techniques. Consequently, modern models are predominantly trained and evaluated on proprietary physical datasets, making independent verification fundamentally difficult.

2.2. Synthetic Data Generation in FPP

To circumvent the grueling process of physical data collection, researchers have increasingly turned to synthetic generation [25, 31]. Early generation methods relied on pure mathematical projections, frequently utilizing Gaussian Mixture Models (GMMs) to procedurally generate surface heights and analytically phase-modulate a sinusoidal carrier frequency. While these methods produce massive volumes of data rapidly, they fail to capture the radiometric realities of an optical lab, completely lacking camera sensor noise, projector non-linearities, and global illumination artifacts like deep shadows.

Subsequent attempts transitioned to basic 3D rendering engines to simulate realistic geometries. However, these pipelines typically projected mathematically perfect textures without rigorously calibrating the virtual camera-projector baseline, leaving a substantial sim-to-real domain gap. As detailed in Table 1, existing public datasets reflect these historical limitations: SynthFringe [11] lacks real-world sensor noise, FPP-ML-Bench [12] offers limited geometric diversity (50 models), and physical datasets like HDR-NET [3] suffer from severe scale constraints. FPP-Depth150 systematically resolves these limitations through a rigorously calibrated, noise-injected digital twin.

2.3. Monocular Depth Estimation (MDE)

Recent breakthroughs in Monocular Depth Estimation (MDE) have demonstrated exceptional zero-shot generalization across diverse natural scenes. Modern foundation models—such as Depth Anything [26, 27], Marigold [9], and MiDaS [17, 18]—leverage either diffusion-based priors or large-scale Vision Transformers (e.g., DINOv2 [16]) trained on massive datasets to extract robust geometric and structural features from single RGB images.

However, despite these architectural advances, standard MDE frameworks are inherently designed to predict relative or affine-invariant depth [2]. They fundamentally lack the physical scale and sub-millimeter precision strictly required for active-light optical metrology. While MDE ex-

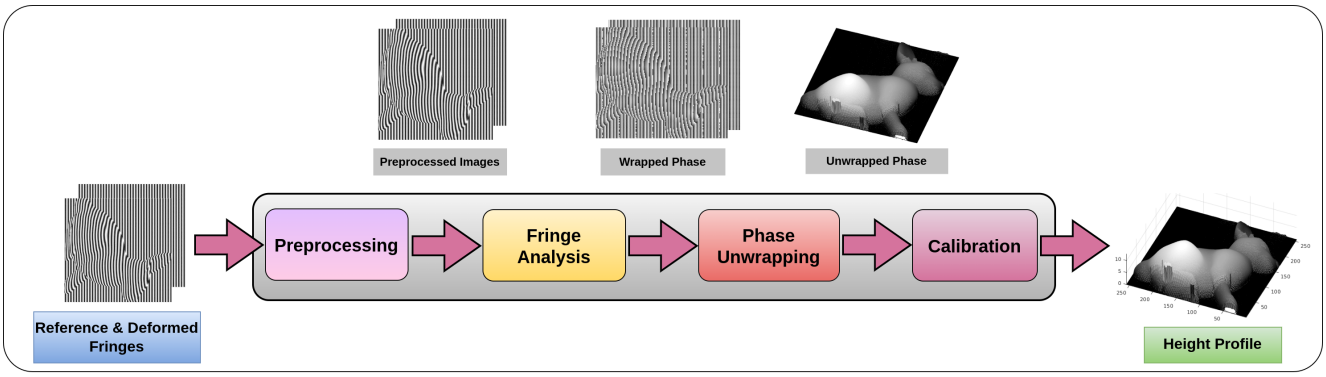

Figure 2. The conventional 3D surface profile estimation pipeline, detailing the sequential processing steps from initial fringe capture to the final 3D depth map.

cels at comprehending global shape and complex topologies, it cannot independently resolve the absolute physical measurements provided by active structured-light systems. In this work, we bridge these two worlds by utilizing our synthetic digital twin data to fine-tune MDE foundation models, effectively grounding their advanced structural priors with precise, absolute metrological scale.

3. Background

In a physical FPP system, a sinusoidal pattern along the x-direction is projected onto a target object. The intensity variation of the captured deformed fringe image $I(x, y)$ can be mathematically formulated as:

$$I(x, y) = A(x, y) + B(x, y)\cos(2\pi f_c x + \phi(x, y)) \quad (1)$$

where $A(x, y)$ represents the background ambient illumination and $B(x, y)$ is the fringe amplitude. The term $2\pi f_c x$ represents the actual phase of the projected fringe with a spatial frequency of f_c. The phase distortion induced by the topological height of the target object is given by $\phi(x, y)$. The primary objective of any FPP algorithm is to accurately extract this phase shift $\phi(x, y)$, that captures the object height variations.

To convert the extracted phase into actual physical depth, the system calculates the phase difference relative to a flat reference plane, denoted as $\phi_0(x, y)$, resulting in $\Delta\phi(x, y) = \phi(x, y) - \phi_0(x, y)$. Based on the calibrated optical system, this phase difference is mapped to actual depth $Z(x, y)$ given by:

$$Z(x, y) = \frac{L_0 \Delta\phi(x, y)}{\Delta\phi(x, y) + 2\pi f_c d} \quad (2)$$

where L_0 is the perpendicular working distance from the camera to the reference plane, and d is the baseline distance between the camera and the projector [6].

To execute this complete 3D surface profile estimation pipeline (illustrated in Fig. 2), conventional techniques sequentially deploy a rigid set of image processing algorithms. For initial fringe quality enhancement, Block Matching and 3D Filtering (BM3D) is frequently applied to suppress inherent camera sensor noise [4]. To extract the wrapped phase from multiple images, Phase Shifting Profilometry (PSP) utilizes phase-shifted patterns to retrieve highly accurate phase maps [21]. For single-shot phase extraction, Fourier Transform Profilometry (FTP) isolates the fundamental carrier frequency in the spectral domain [22]. To address noise sensitivity in complex surfaces, Windowed Fourier Transform Profilometry (WFTP) applies localized spatial windowing [10]. Additionally, the 1D Continuous Wavelet Transform (CWT) leverages time-frequency analysis to accurately demodulate non-stationary fringe signals [7].

Once the wrapped phase is obtained, Multi-Frequency Temporal Phase Unwrapping (MFTPU) resolves phase discontinuities across the full field by cross-referencing sequential patterns to eliminate fringe ambiguities [28]. For spatial unwrapping from a single map, Quality Guided Phase Unwrapping (QGPU) prioritizes continuous phase recovery along high-confidence paths, effectively isolating noisy shadow regions to prevent severe error propagation [30]. Finally, to map the unwrapped absolute phase to physical depth, phase-height models establish geometric calculations relative to a fixed reference plane [32], while alternative triangular stereo methods calibrate both the camera and the projector using pinhole models to implicitly handle system non-linearities [6].

4. Methodology

4.1. Physical Setup and Calibration

Our primary physical FPP setup comprises a Digital Light Processing (DLP) analog projector, an industrial Imaging Source DFK 27BUJ003 camera equipped with a 10.7-megapixel sensor (1.67 μm pixel size), and a flat reference

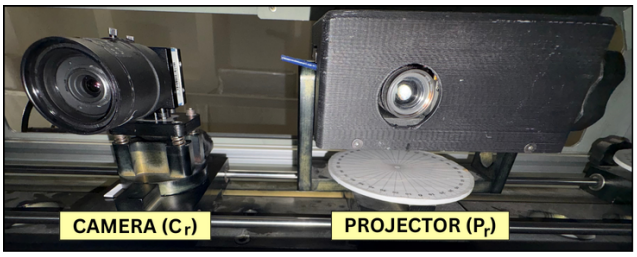

Figure 3. Close-up of the physical FPP hardware: the industrial camera C_r (left) and the DLP projector P_r (right).

plane. This reference plane is a 62 cm $\times$ 39 cm surface designated as S_r for the real setup. The system geometry was meticulously calibrated and fixed to ensure that the projected fringe patterns maintain a consistent fringe order throughout the x-direction.

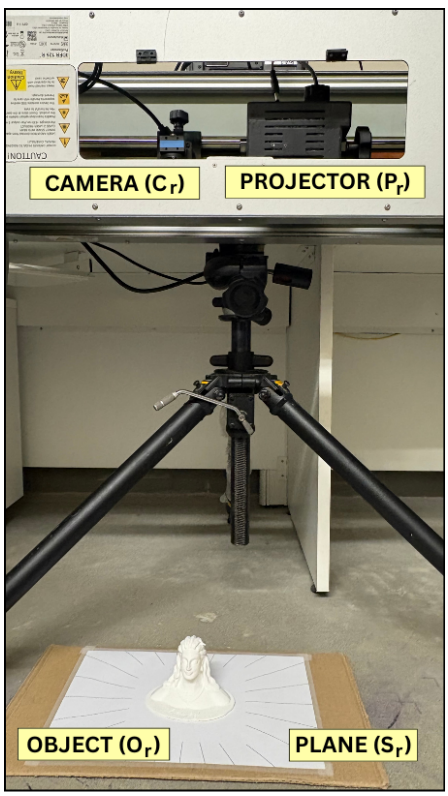

Figure 4. Full physical experimental setup illustrating the camera (top left) and projector (top right) positioned above the target object (bottom left) resting on the reference plane (bottom right).

Specifically, the real camera (C_r) and projector (P_r) were positioned with a baseline distance of 15 cm to simultaneously minimize severe self-shadowing and maximize the observable surface area. Both components were mounted at a perpendicular working distance of 74 cm from the reference plane. This height was strategically selected so that the system easily accommodates our current 14 cm target objects while preserving sufficient optical clearance to scale up to 30 cm objects in future iterations. The close-up hardware configuration of C_r and P_r is detailed in Fig. 3, while the complete environmental arrangement—comprising the camera, projector, reference plane (S_r), and target object (O_r)—is illustrated in Fig. 4.

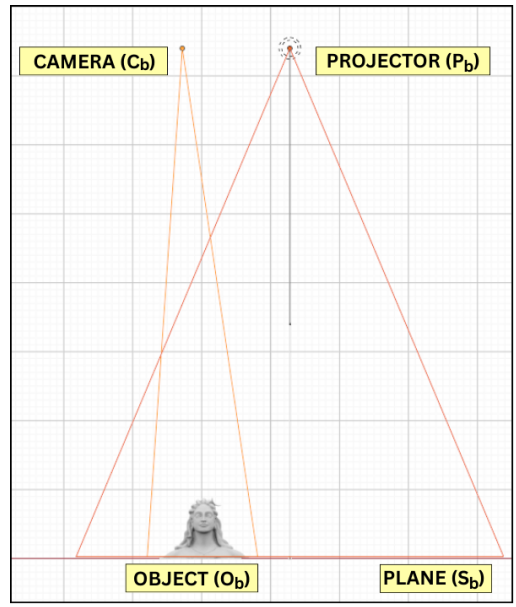

Figure 5. Digital twin environment within Blender, replicating the physical geometry with the virtual camera C_b (top left) and virtual projector P_b (top right) directed at the 3D target object O_b and reference plane S_b.

For active illumination, we generated a sinusoidal fringe pattern ($f_c = 32$) via MATLAB. To standardize inputs for neural network processing, IC Capture software was used to extract a 640 $\times$ 640 pixel Region of Interest (ROI) from the full field of view, calibrated to a 16 cm $\times$ 16 cm physical area on the reference plane. To mitigate radial lens distortion that inherently bends fringe lines, this ROI was strictly centered along the camera's principal optical axis. This strategic placement minimizes phase distortions, ensuring the captured optical depth can be reliably and linearly converted into absolute physical height.

4.2. Digital Twin Construction

To scale data generation, a highly accurate digital twin of the physical system was constructed within the Blender 3D environment (illustrated in Fig. 5). To ensure photorealistic light transport, accurate self-occlusions, and authentic global illumination artifacts, we strictly utilized Blender's physically-based path-tracing engine, Cycles. The exact spatial measurements from the laboratory were rigidly replicated in the virtual space. The virtual camera (C_b) and virtual projector (P_b) were configured to mimic the 15 cm

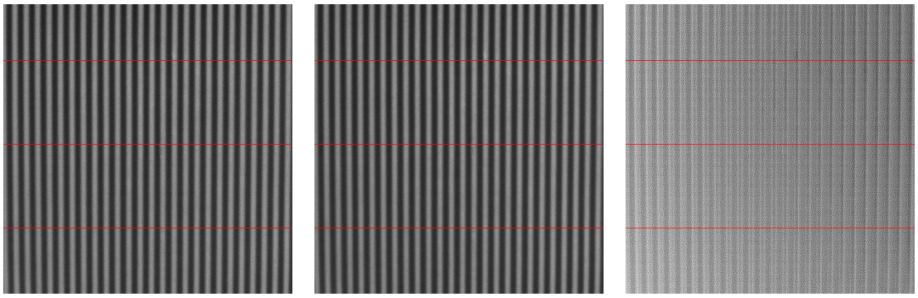

Figure 6. Qualitative comparison of the reference fringe patterns used for digital twin validation. Left: The real physical laboratory capture. Middle: The Blender-generated synthetic reference fringe. Right: The difference map between the two, demonstrating high-fidelity pixel-level spatial alignment.

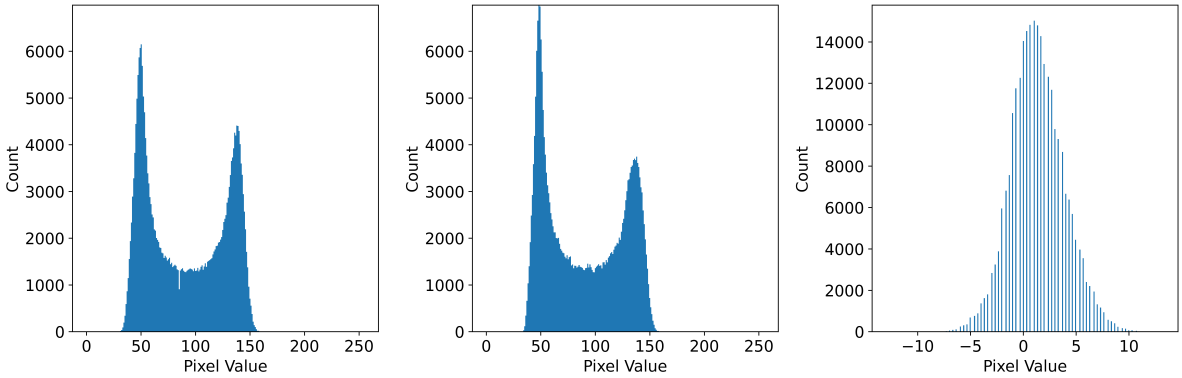

Figure 7. Global intensity distribution histograms corresponding to the reference fringes in Fig. 6. Left: The physical sensor capture. Middle: The synthetic render, successfully replicating the bimodal radiometric profile. Right: Histogram of the difference, exhibiting a Gaussian-like error distribution that characterizes the empirical sensor noise.

baseline and 74 cm perpendicular working distance. Furthermore, the intrinsic parameters of C_b, including focal length and sensor dimensions, were meticulously matched to the physical industrial camera to eliminate field-of-view discrepancies and preserve spatial resolution.

Correspondingly, the virtual reference plane (S_b) precisely matches the physical 62 cm $\times$ 39 cm dimensions to support the virtual target object (O_b). To accurately model the complex radiometric interactions observed in the physical laboratory, the surface properties of both S_b and O_b were defined using Blender's Principled BSDF shader. This enabled procedural control over the bidirectional reflectance distribution function (BRDF), allowing us to dynamically vary surface roughness and metallic parameters to simulate the challenging HDR specular highlights for robust model training. Finally, global ambient environment lighting was disabled to ensure that the active structured light emitted from P_b remained the sole dominant illumination source, effectively replicating a tightly controlled optical darkroom.

The main virtual projector (P_b) was instantiated as a modified spotlight utilizing the open-source Ocupe Projectors add-on for Blender [15]. Furthermore, high-precision virtual ground truth absolute depth maps (H_{gt}) are continuously extracted using Blender's internal Z-buffer to serve as exact, noise-free regression targets.

4.3. Sim-to-Real Validation and Calibration

To ensure the synthetic data perfectly mimics real experimental acquisitions and minimizes the domain gap for downstream neural network training, we executed a rigorous, multi-step calibration protocol, the progression of which is visualized in Fig. 8 and quantified in Fig. 9:

- **Spatial Alignment (Setup 1 $\rightarrow$ Setup 2):** Initially, the virtual camera and projector were placed at the exact measured physical baselines (15 cm baseline, 74 cm height), designated as **Setup 1**. However, internal projector optics and sensor ROI cropping introduced a subtle spatial shift, causing the synthetic shadows and phase to misalign with reality (orange line, Fig. 9). To correct this without altering the fixed geometric constraints, we introduced an auxiliary virtual projector at the exact spatial coordinates of C_b. By projecting a real-world captured image of a CAD model through this auxiliary projector, we iteratively fine-tuned the $X - Y$ projection shift of P_b and the sensor

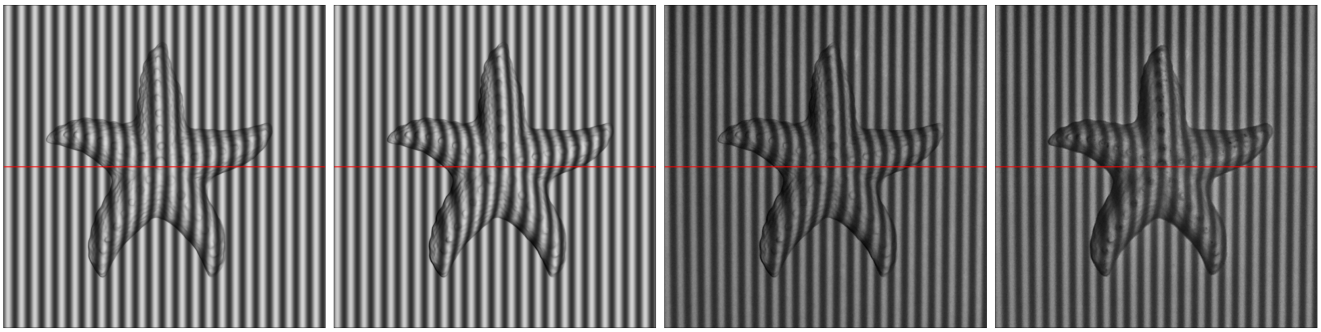

Figure 8. Ablation of the digital twin calibration process on the starfish model. From left to right: **Setup 1** (pure geometric baseline with mathematical MATLAB projection), **Setup 2** (incorporating spatial X-Y shifts for accurate spatial alignment), **Our Final Setup** (spatial shifts combined with cropped physical radiometric projection mapping), and the **Ground Truth** physical laboratory capture. The red line indicates the cross-section used for 1D scanline evaluation.

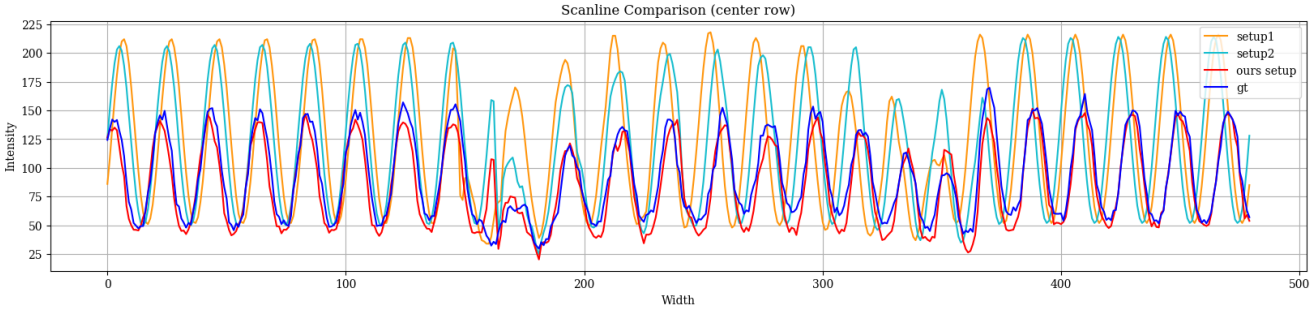

Figure 9. 1D cross-sectional scanline analysis corresponding to the red line in Fig. 8. Setup 1 (orange) exhibits severe phase misalignment. Setup 2 (cyan) corrects the spatial phase shift but fails to match the radiometric amplitude. Our final setup (red) successfully matches both the spatial frequency and the complex fringe modulation amplitude of the physical ground truth (blue).

capture shift of C_b. This alignment, designated as **Setup 2**, ensured that angles perfectly mimicked the real setup (cyan line, Fig. 9).

- **Radiometric Projection Mapping (Setup 2 $\rightarrow$ Ours):** While Setup 2 achieved spatial phase alignment, it still utilized a mathematically perfect MATLAB sinusoidal pattern. This resulted in fringe amplitudes that lacked authentic physical characteristics and exhibited artificially high contrast. To resolve this, we transitioned to **Our Setup** by capturing a physical reference fringe using the real camera. From this full capture, we specifically cropped the 16 cm $\times$ 16 cm Region of Interest (ROI) and explicitly mapped this cropped patch as the source texture for the virtual projector. This projection was precisely positioned 1 cm from the right edge and 8 cm from the top edge of the virtual reference plane. As shown in Fig. 8, this perfectly replicates the physical sensor's radiometric profile.
- **Ground Truth Synthesis for Physical Captures:** To rigorously evaluate our trained models on real-world data, we required highly accurate ground truth depth maps for physical test objects. To achieve this, we physically 3D-printed a selected subset of our complex CAD models (see Fig. 10) at exact, known scales. Because the physical objects and virtual CAD files shared identical reference dimensions, we could project the physical camera capture into the digital twin using the auxiliary projector. The corresponding virtual CAD model was then manually translated and rotated until its rendered silhouette and geometric topology perfectly overlapped the physical capture. Once registered, the absolute ground truth depth map was extracted flawlessly using Blender's internal Z-buffer.
- **Sensor Noise Modeling:** To accurately model temporal sensor noise, we captured 500 static reference images from the physical setup. By computing the residual spatial differences against the corresponding base Blender renders (visualized in Fig. 6), we derived an empirical Gaussian noise distribution. As demonstrated by the global intensity histograms in Fig. 7, our synthetic rendering pipeline successfully replicates the bimodal radiometric profile of the real sensor, with the residual error tightly matching this physical noise distribution. During training, noise sampled directly from this empirical distri-

bution is added to the synthetic data to guarantee robust real-world generalization.

- **Signal Fidelity Validation:** Finally, the 1D cross-sectional scanline analysis (Fig. 9) serves as the ultimate validation of our pipeline. It definitively confirms that our calibration strategy successfully bridges the domain gap, with our final synthetic prediction (red) mirroring the spatial frequency, fringe modulation amplitude, and edge-induced phase disruptions of the physical ground truth (blue) with sub-pixel fidelity.

Figure 10. A representative subset of the physically 3D-printed CAD models utilized for data generation, chosen for their complex topological features: Adiyogi (top left), Starfish (top right), Hanuman (bottom left), and Leo (bottom right).

5. Proposed Baseline: Depth Anything for FPP

5.1. Repurposing MDE Foundation Models

Monocular Depth Estimation (MDE) models, such as Depth Anything, leverage the DINOv2 Vision Transformer (ViT) to extract exceptionally robust global geometric features from single images. However, these foundation models are intrinsically trained on natural scenes to predict relative or affine-invariant depth maps. They fundamentally lack the sub-millimeter precision and absolute physical scale required for active-light metrology.

To bridge this paradigm gap, we propose a direct adaptation framework that repurposes MDE architectures for single-shot FPP. Instead of relying on classical phase-unwrapping or multi-image inputs, we feed the single deformed fringe image I_{def} directly into the network. To fully leverage the pre-trained DINOv2 encoder without destroying its learned structural priors, the single-channel grayscale FPP image is duplicated across the three standard input channels. We explore three distinct encoder scales provided by the Depth Anything architecture: ViT-S (Small), ViT-B (Base), and ViT-L (Large). The ViT encoder thus extracts a pyramid of multi-scale features rich in topological understanding and global object shape.

The critical transition from relative depth to absolute metrological scale occurs within the Dense Prediction Transformer (DPT) decoder. To formulate this as a stable regression task, the raw physical depth maps (Z, which maintain a fixed background working distance of 74 cm) are converted into normalized height maps (H_{norm}). Specifically, the depth map is subtracted from the 74 cm reference background to yield absolute height, which is then divided by the maximum structural height range of 14 cm:

$$H_{norm} = \frac{74 - Z}{14} \tag{3}$$

To perfectly align the network's output space with this $[0, 1]$ normalized target distribution, we append a Sigmoid activation function to the final layer of the DPT decoder. This architectural modification ensures the network inherently bounds its absolute height predictions.

By explicitly training the DPT decoder on our highly calibrated FPP-Depth150 dataset, the network is forced to implicitly demodulate the physical phase shift embedded within the structured light pattern. The training objective is formulated to ensure both pixel-wise accuracy and structural fidelity. We utilize Mean Absolute Error ($\mathcal{L}_{\text{MAE}}$) to aggressively penalize absolute height deviations, and we define a structural loss component as $\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(\hat{H}, H_{\text{norm}})$ to strictly enforce the sharp preservation of geometric boundaries and sinusoidal phase lines. The final simplified total loss function is thus a weighted combination of these objectives:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{mae}} \mathcal{L}_{\text{MAE}} + \lambda_{\text{ssim}} \mathcal{L}_{\text{SSIM}} \tag{4}$$

Through empirical evaluation, the hyperparameter weights were optimally set to $\lambda_{\text{mae}} = 1.0$ and $\lambda_{\text{ssim}} = 0.5$ to ensure the network accurately resolves absolute metrological scale without inducing high-frequency structural artifacts.

To retain the powerful zero-shot relative depth priors of the pre-trained DINOv2 encoder while allowing it to gently adapt to the active-light domain, we employ a differential learning rate strategy. The DINOv2 encoder is fine-tuned with a conservative learning rate of 5×10^{-6}. Conversely, the DPT decoder—which must learn the novel mapping to absolute metrological scale from scratch—is trained with a learning rate of 5×10^{-5}, exactly ten times larger. To ensure stable convergence across the massive dataset, the proposed framework was implemented in PyTorch and trained using a distributed data-parallel architecture across four NVIDIA A100 Tensor Core GPUs (80 GB each) with a global batch size of 64.

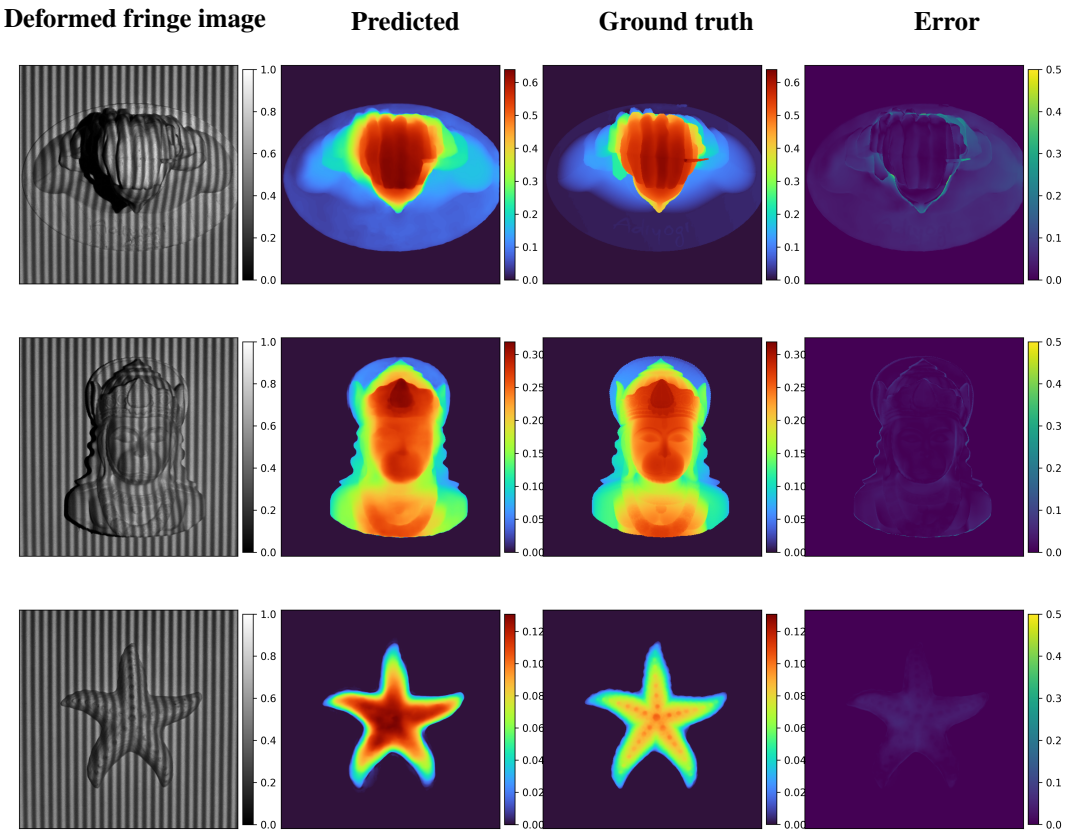

Figure 11. Visual result of testing on real data. From left to right: deformed fringe image, predicted depth map, ground truth depth map, and error map.

6. Results and Discussion

6.1. Quantitative Baseline Results

To benchmark the repurposed MDE foundation models on the FPP-Depth150 hidden test set, we evaluate three encoder variants (ViT-S, ViT-B, ViT-L) using four standard metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for absolute height deviations, Threshold Accuracy ($\delta_1 < 1.1$) for relative reliability, and the Structural Similarity Index (SSIM) for spatial consistency and edge preservation. As detailed in Table 2, scaling the DINOv2 encoder's parameter count yields a monotonic performance improvement. This confirms that larger foundation models provide significantly stronger geometric priors for resolving absolute phase-to-depth mappings in optical metrology.

Table 2. Quantitative baseline results on the FPP-Depth150 hidden test set evaluating the repurposed Depth Anything architecture across three DINOv2 encoder scales (ViT-S, ViT-B, ViT-L).

Encoder Backbone	**MAE** ↓	**RMSE** ↓	$\delta_1 < 1.1$ ↑	**SSIM** ↑
Object Region				
Depth Anything (ViT-S)	0.0333	0.0533	0.6094	0.9289
Depth Anything (ViT-B)	0.0221	0.0387	0.7494	0.9554
Depth Anything (ViT-L)	0.0186	0.0350	0.7980	0.9621
Full Image				
Depth Anything (ViT-S)	0.0075	0.0262	—	—
Depth Anything (ViT-B)	0.0051	0.0202	—	—
Depth Anything (ViT-L)	0.0043	0.0186	—	—

7. Conclusion

By abstracting the complex physical calibration processes of FPP into a standardized digital twin pipeline, we introduce FPP-Depth150 to lower the barrier to entry for computer vision researchers tackling optical metrology. Our proposed baseline successfully bridges MDE foundation models and structured-light physical priors, proving that highly calibrated synthetic data—combined with robust Vision Transformer backbones—can achieve accurate and reliable sim-to-real generalization for single-shot 3D surface profiling.

References

[1] Songlin Bai, Xiaolong Luo, Kun Xiao, Chunqian Tan, and Wanzhong Song. Deep absolute phase recovery from single-frequency phase map for handheld 3d measurement. *Optics Communications*, 512:128008, 2022. 2

[2] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, 2023. 2

[3] Jianbin Cao, Xu Zhang, Guangya Zhou, and Dawei Tu. Hdrnet: A hdr dataset of fringe projection based on deep network. 2024. 2

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 3

[5] Shijie Feng, Qian Chen, Guohua Gu, Tianyang Tao, Liang Zhang, Yan Hu, Wei Yin, and Chao Zuo. Fringe pattern analysis using deep learning. *Advanced photonics*, 1(2):025001–025001, 2019. 2

[6] Shijie Feng, Chao Zuo, Liang Zhang, Tianyang Tao, Yan Hu, Wei Yin, Jiaming Qian, and Qian Chen. Calibration of fringe projection profilometry: A comparative review. *Optics and lasers in engineering*, 143:106622, 2021. 3

[7] Munther A Gdeisat, David R Burton, and Michael J Lalor. Spatial carrier fringe pattern demodulation by use of a two-dimensional continuous wavelet transform. *Applied optics*, 45(34):8722–8732, 2006. 3

[8] Wenxin Hu, Hong Miao, Keyu Yan, and Yu Fu. A fringe phase extraction method based on neural network. *Sensors*, 21(5):1664, 2021. 2

[9] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9492–9502, 2024. 2

[10] Qian Kemao. Windowed fourier transform for fringe pattern analysis. *Applied Optics*, 43(13):2695–2702, 2004. 3

[11] Won-Hoe Kim, Bongjoong Kim, Hyung-Gun Chi, and Jae-Sang Hyun. Enhanced fringe-to-phase framework using deep learning. *arXiv preprint arXiv:2402.00977*, 2024. 2

[12] Anush Lakshman, Adam Haroon, and Beiwen Li. Comprehensive machine learning benchmarking for fringe projection profilometry with photorealistic synthetic data. In *Photonic Instrumentation Engineering XIII*, page 1390402. SPIE, 2026. 2

[13] Ze Li, Jianhua Wang, Yixin Ji, Suzhen Wang, Wen Zhang, Shuo Shan, and Yanxi Yang. Fringe projection profilometry based on deep learning phase demodulation combined with temporal phase unwrapping. *Applied Physics B*, 130(12): 213, 2024. 2

[14] Rakesh Chowdary Machineni, GE Spoorthi, Krishna Sumanth Vengala, Subrahmanyam Gorthi, and Rama Krishna Sai S Gorthi. End-to-end deep learning-based fringe projection framework for 3d profiling of objects. *Computer vision and image understanding*, 199:103023, 2020. 1

[15] Ocupe. Projectors. https://github.com/Ocupe/Projectors, 2023. GitHub repository. 5

[16] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 2

[17] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence*, 44(3):1623–1637, 2020. 2

[18] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12179–12188, 2021. 2

[19] Vaishnavi Ravi, Siddharth Parlapalli, Sameer Ranjan, and Rama Krishna Sai Subrahmanyam Gorthi. Transformer-based fringe restoration for shadow mitigation in fringe projection profilometry. In *27th International Conference on Pattern Recognition (ICPR)*, pages 339–354. Springer, 2024. 2

[20] GE Spoorthi, Subrahmanyam Gorthi, and Rama Krishna Sai Gorthi. A deep learning-based model for phase unwrapping. In *Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing*, pages 1–8, 2018. 2

[21] Venugopal Srinivasan, Hsin-Chu Liu, and Maurice Halioua. Automated phase-measuring profilometry of 3-d diffuse objects. *Applied optics*, 23(18):3105–3108, 1984. 3

[22] Mitsuo Takeda and Kazuhiro Mutoh. Fourier transform profilometry for the automatic measurement of 3-d object shapes. *Applied optics*, 22(24):3977–3982, 1983. 3

[23] Maciej Trusiak and Małgorzata Kujawińska. Deep learning enabled single-shot absolute phase recovery in high-speed composite fringe pattern profilometry of separated objects. *Opto-Electronic Advances*, 6(12), 2023. 1

[24] Bowen Wang, Wenwu Chen, Jiaming Qian, Shijie Feng, Qian Chen, and Chao Zuo. Single-shot super-resolved fringe projection profilometry (sssr-fpp): 100,000 frames-per-second 3d imaging with deep learning. *Light: Science & Applications*, 14(1):70, 2025. 1

[25] Fanzhou Wang, Chenxing Wang, and Qingze Guan. Single-shot fringe projection profilometry based on deep learning and computer graphics. *Optics Express*, 29(6):8024–8040, 2021. 2

[26] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10371–10381, 2024. 2

[27] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *Advances in Neural Information Processing Systems*, 37:21875–21911, 2024. 2

[28] Wei Yin, Qian Chen, Shijie Feng, Tianyang Tao, Lei Huang, Maciej Trusiak, Anand Asundi, and Chao Zuo. Temporal

phase unwrapping using deep learning. *Scientific reports*, 9 (1):20175, 2019. 3

[29] Haotian Yu, Xiaoyu Chen, Zhao Zhang, Chao Zuo, Yi Zhang, Dongliang Zheng, and Jing Han. Dynamic 3-d measurement based on fringe-to-fringe transformation using deep learning. *Optics express*, 28(7):9405–9418, 2020. 2

[30] Ming Zhao, Lei Huang, Qican Zhang, Xianyu Su, Anand Asundi, and Qian Kemao. Quality-guided phase unwrapping technique: comparison of quality maps and guiding strategies. *Applied optics*, 50(33):6214–6224, 2011. 3

[31] Yi Zheng, Shaodong Wang, Qing Li, and Beiwen Li. Fringe projection profilometry by conducting deep learning from its digital twin. *Optics express*, 28(24):36568–36583, 2020. 2

[32] W-S Zhou and X-Y Su. A direct mapping algorithm for phase-measuring profilometry. *Journal of modern optics*, 41 (1):89–94, 1994. 3

[33] Chao Zuo, Shijie Feng, Lei Huang, Tianyang Tao, Wei Yin, and Qian Chen. Phase shifting algorithms for fringe projection profilometry: A review. *Optics and lasers in engineering*, 109:23–59, 2018. 1