







Article

From Single Shot to Structure: End-to-End Network-Based Deflectometry for Specular Free-Form Surface Reconstruction

M.Hadi Sepanj ¹, Saed Moradi ¹, Amir Nazemi ¹, Claire Preston ², Anthony M. D. Lee ²
and Paul Fieguth ^{1,*}

¹ Vision and Image Processing Laboratory, Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada; mhsepanj@uwaterloo.ca (M.S.); saed.moradi@uwaterloo.ca (S.M.); amir.nazemi@uwaterloo.ca (A.N.)

² General Fusion Inc., 6020 Russ Baker Way, Richmond, BC V7B 1B4, Canada

* Correspondence: paul.fieguth@uwaterloo.ca

Abstract: Deflectometry is a key component in the precise measurement of specular (mirrored) surfaces; however, traditional methods often lack an end-to-end approach that performs 3D reconstruction in a single shot with high accuracy and generalizes across different free-form surfaces. This paper introduces a novel deep neural network (DNN)-based approach for end-to-end 3D reconstruction of free-form specular surfaces using single-shot deflectometry. Our proposed network, VUDNet, innovatively combines discriminative and generative components to accurately interpret orthogonal fringe patterns and generate high-fidelity 3D surface reconstructions. By leveraging a hybrid architecture integrating a Variational Autoencoder (VAE) and a modified U-Net, VUDNet excels in both depth estimation and detail refinement, achieving superior performance in challenging environments. Extensive data simulation using Blender leading to a dataset which we will make available, ensures robust training and enables the network to generalize across diverse scenarios. Experimental results demonstrate the strong performance of VUDNet, setting a new standard for 3D surface reconstruction.

Keywords: deflectometry; specular surface reconstruction; deep neural networks; single-shot measurement; data simulation; variational autoencoder (VAE)



Citation: Sepanj, M.H.; Moradi, S.; Nazemi, A.; Preston, C.; Lee, A.M.D.; Fieguth, P. From Single Shot to Structure: End-to-End Network-Based Deflectometry for Specular Free-Form Surface Reconstruction. *Appl. Sci.* **2024**, *14*, 10824. <https://doi.org/10.3390/app142310824>

Academic Editor: Xi Zhao

Received: 19 September 2024

Revised: 1 November 2024

Accepted: 14 November 2024

Published: 22 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, rapid advancements in optical-based industries have significantly influenced sectors such as microelectronics [1–3] and automotive design [4–6], leading to a strong demand for highly accurate methods for measuring specular surfaces [7,8]. Deflectometry [9–11] is an optical measurement technique which relies on the reflection of structured light patterns from an unknown surface to infer its shape. Deflectometry is highly effective for measuring specular surfaces due to its accuracy, speed, and non-contact approach, making it particularly suitable for capturing detailed surface variations [12]. Additionally, free-form surface reconstruction [13,14] is crucial for modeling complex, irregular shapes found in advanced optical systems [15]. Together, these methods enhance precision and innovation in optical manufacturing.

The more specific context of single-shot deflectometry [16–18], performing deflectometry based on only a single image, is crucial for several reasons, particularly for high-speed imaging of fast dynamically changing surfaces. This includes applications such as fast-moving or morphing specular surfaces like liquids (water, metals), where capturing multiple images is impractical due to the rapidly changing nature of the surface. Additionally, in industrial scanning, single-shot deflectometry enhances efficiency by eliminating the need for multiple images, important for quality control of reflective objects [9]. For example, in General Fusion's Magnetized Target Fusion technology scheme, measurement

of the 3D shape of an imploding specular liquid metal vortex to compress plasma to generate nuclear fusion with microsecond precision is required [19]. Traditional methods [10] require multiple captures and complex setups to obtain surface-gradient information in perpendicular directions, which can be time-consuming and problematically inefficient for dynamic applications.

To further enhance the efficiency and accuracy of single-shot deflectometry, recent research has explored the use of composite fringe patterns in deep learning-based 3D surface measurement techniques [17,20]. These composite patterns are particularly effective in addressing the challenges associated with single-shot deflectometry, such as encoding complex surface details in a single capture. Historically, in the use of fringe patterns for surface measurement, vertical and horizontal patterns are often combined into a single composite pattern. The separation of these orthogonal fringe patterns into their individual components is crucial for accurate analysis and has traditionally been achieved through methods like the 2D Fourier algorithm [21] which, despite its utility, often results in phase errors and spectrum aliasing [22].

In contrast, in this paper, we propose a novel deep neural network (DNN)-based approach designed for the end-to-end 3D reconstruction of free-form surfaces. Our network, VUDNet, processes 2D images captured from a surface onto which an orthogonal fringe pattern is projected (Figure 1). This discriminative–generative hybrid network excels in depth estimation. To the best of our knowledge, VUDNet represents the first DNN-based approach in deflectometry, pioneering the use of generative methods for depth estimation.

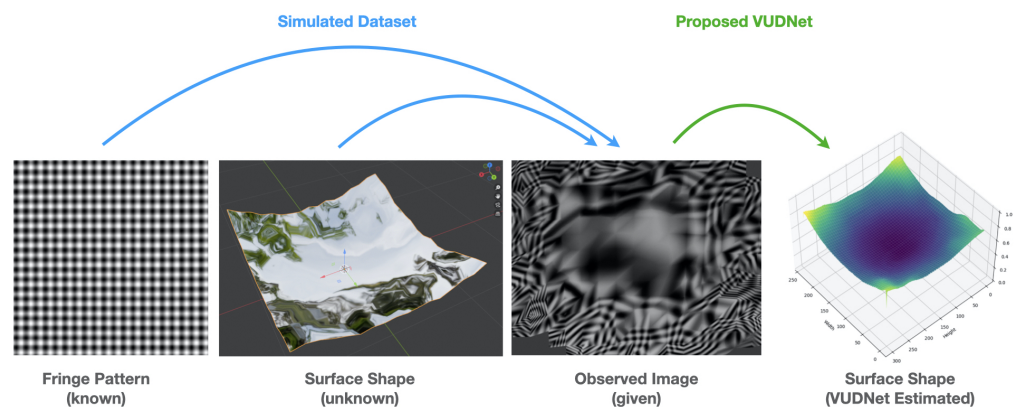


Figure 1. Overview of the context faced by this paper: A known fringe pattern is reflected by some shape of interest, and the resulting reflection is captured by a camera. Our proposed method, VUDNet, reconstructs the estimated shape based on the observed image, trained on a dataset of simulated reflections.

The integration of both generative and discriminative components in our network is pivotal for achieving high-fidelity 3D surface reconstruction. The discriminative part of the network is essential for accurately identifying and interpreting the orthogonal fringe patterns projected onto the specular surface. This involves extracting meaningful features from an image, which are crucial for determining the gradient information necessary for depth estimation. The discriminative model’s ability to parse complex visual patterns with precision directly impacts the accuracy of the subsequent 3D reconstruction.

On the other hand, the generative component plays a crucial role in translating the extracted gradient information into a coherent 3D representation. The generative model excels in learning and representing the underlying data distribution through embedding representations [23,24]. These embeddings capture the intricate details and variations of the surface, allowing the reconstruction of a highly detailed and accurate 3D shape. The learned embeddings are vital because they encapsulate the spatial relationships and depth cues that are not readily apparent in the raw 2D images [24]. This deep representation allows the generative model to fill in gaps and correct any inconsistencies in the depth

information, resulting in a precise 3D reconstruction. This synergy between the two models not only enhances the overall performance of the network but also ensures that the 3D reconstructions are both geometrically precise and visually coherent.

Paper Contributions

1. **Novel DNN-Based Approach for Deflectometry:** We introduce VUDNet, the first deep neural network (DNN) designed specifically for end-to-end 3D reconstruction of free-form specular surfaces using single-shot deflectometry. VUDNet leverages a hybrid architecture that combines the strengths of both generative and discriminative models, ensuring high accuracy and generalization.
2. **Dataset Simulation:** To train and evaluate our model, we simulated an extensive dataset. This dataset, which includes a variety of deformed specular surfaces and their corresponding depth maps, will be made publicly available to support future research in this field.
3. **Robust Performance in Challenging Environments:** Experimental results demonstrate that VUDNet significantly outperforms existing methods in reconstructing 3D surfaces from single-shot 2D images, particularly in challenging environments. Our network demonstrates the ability to generalize across diverse scenarios.

2. Related Work

Deflectometry has high accuracy, speed, and is non-contact, making it suitable for a variety of industrial applications [9]. However, traditional deflectometry methods, such as phase-shifting techniques, require multiple images to be captured in order to obtain surface gradients in two perpendicular directions [9,17]. This process can be time-consuming and inefficient, particularly for dynamic or real-time applications.

Single-shot deflectometry was developed to overcome these limitations [9,17,18,25,26]. Seo et al. developed a single-shot freeform surface profiler based on spatially phase-shifted lateral shearing interferometry [17]. This approach, while simplifying the measurement process, still struggles with complex surface geometries. Phase measuring profilometry has also seen advancements, with deep learning methods enhancing measurement capabilities [27,28]. Working from a single image significantly reduces the time required for surface measurement; however, there are resulting challenges in interpreting the complex patterns reflected by specular surfaces, especially those with low reflectivity or intricate geometries.

The application of deep learning in optical metrology, including deflectometry, has shown significant promise [12,17,26,29]. Deep learning models can learn complex patterns and provide robust phase retrieval from noisy or low-quality data. DYNet++ [17] utilizes deep learning for single-shot deflectometry, capable of retrieving phase information from composite patterns even in challenging conditions with both closed- and open-loop fringe patterns.

Wang et al. proposed a deep learning-based deflectometric method for freeform surface measurement [26]. Their approach involves devising a deep neural network specifically for freeform surface reconstruction. This method offers a solution for the general measurement of freeform surfaces while minimizing measurement errors caused by noise and system geometry calibration.

3. Method

The architecture of our proposed network represents a sophisticated blend of specialized components designed to tackle the unique challenges of reconstructing 3D specular surfaces from single-shot 2D images. Central to this architecture are two main components: the discriminative and generative branches which form a novel hybrid depth estimator network. Our proposed model is an ensemble of a Variational Auto Encoder (VAE) [30–32] and a modified U-Net [33] architecture, adapted here to enhance depth map refinement. This section details each component's role and functionality within our network, elucidat-

ing how they collectively contribute to 3D surface reconstruction. Figure 2 illustrates the comprehensive architecture of our network.

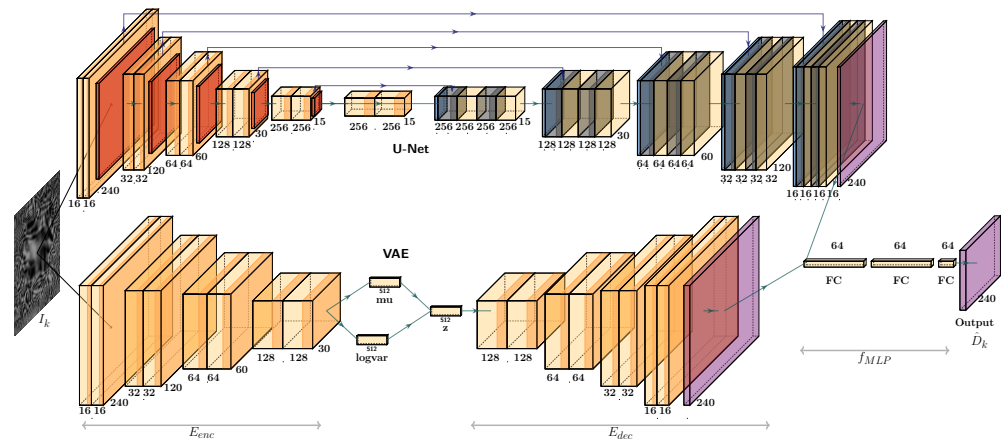


Figure 2. General architecture of the proposed VUDNet for end-to-end 3D reconstruction of specular free-form surfaces. The network integrates a Variational Autoencoder (VAE, **bottom**) for coarse depth estimation and a modified U-Net (**top**) for detail refinement. The ensemble approach leverages both generative and discriminative components, combining their outputs (**right**) to produce accurate depth maps from single-shot 2D images.

3.1. Problem Formulation

In this work, we address the challenge of reconstructing a 3D surface from a single-shot 2D image of an orthogonal sinusoidal fringe pattern reflected on a specular surface. To formulate the problem, we introduce the following notation:

- $I = \{I_k \mid k = 1, \dots, n\}$: The 2D reflected image of a pattern \mathcal{P} reflected by surface \mathcal{S} .
- $\mathcal{D} = \{D_k \mid k = 1, \dots, n\}$: The ground truth depth map corresponding to surface \mathcal{S} .
- $\hat{\mathcal{D}}$: The estimated depth map produced by our network.
- $f_{MLP} : \mathbb{R}^{H \times W} \times \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{H \times W}$: The MLP function that combines the outputs of the VAE and U-Net to produce $\hat{\mathcal{D}}$.
- \mathcal{C} : camera parameters.
- $\mathcal{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$: The projection to render a 2D image I from surface \mathcal{S} .
- $\psi : \mathbb{R}^3 \rightarrow \mathbb{R}$: The function mapping a 3D point to depth.

The objective is to develop a hybrid depth estimator network, VUDNet, that integrates both generative and discriminative modeling approaches. To train and evaluate our model, we simulate a comprehensive dataset that includes 2D images I and corresponding ground truth depth maps \mathcal{D} . The generated dataset captures the complex interactions of light with specular surfaces, providing a robust foundation for training our hybrid depth estimator network. This approach ensures that our model can generalize well to real-world scenarios involving specular reflections.

3.2. Architecture Overview

The VUDNet architecture integrates a Variational Auto Encoder (VAE) and a modified U-Net, both of which are provided with an input 2D image I , the reflection of pattern \mathcal{P} on some unknown surface \mathcal{S} . The VAE is a generative model that aims to learn the underlying distribution of the data, providing a coarse estimation of the depth map. In contrast, the U-Net is a discriminative model designed to refine this estimation, yielding a finer and more precise depth map. The hybrid nature of our architecture leverages the strengths of both generative and discriminative approaches, creating a comprehensive depth estimation model.

The VAE component of our architecture consists of an encoder function $E_{enc} : \mathbb{R}^{H \times W \times c} \rightarrow \mathbb{R}^Z$ that compresses the input image into a latent space representation $z = E_{enc}(I)$,

followed by a decoder function $E_{\text{dec}} : \mathbb{R}^Z \rightarrow \mathbb{R}^{H \times W}$ that reconstructs the depth map from this latent representation, producing depth estimate $\hat{D}_k^V = E_{\text{dec}}(z)$. This process allows the VAE to capture and model the complex variations in specular surfaces. On the other hand, the modified U-Net architecture, tailored for depth estimation, employs a series of downsampling and upsampling layers specifically structured to refine the depth map, producing estimate \hat{D}_k^U . Unlike the traditional U-Net, our modified version diminishes the effect of skip connections to prevent the direct transfer of the texture information present in the patterning of input image I , thereby ensuring that the output is a detailed depth map rather than a combination of depth and texture features.

The output depth maps from both the VAE and the modified U-Net are then fed into a fully connected Multi-Layer Perceptron (MLP) that serves as an ensemble method. We let $f_{\text{MLP}} : \mathbb{R}^{H \times W} \times \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{H \times W}$ denote the MLP function that combines the depth maps from the two networks:

$$\hat{D}_k = f_{\text{MLP}}(\hat{D}_k^V, \hat{D}_k^U) \approx \psi(\mathcal{F}^{-1}(I)) \quad (1)$$

Such an ensemble approach is beneficial for several reasons:

- **Reduced Overfitting:** Ensemble methods inherently reduce the risk of overfitting [34]. They are known for their ability to improve generalization by leveraging the strengths of multiple models and mitigating individual model biases [34]. In our work, this is especially true as we combine one discriminative model and one generative model. Each model captures different aspects of the data distribution, and their combination leads to a more generalizable solution.
- **Complementary Strengths:** The VAE's ability to model complex data distributions complements the U-Net's strength in preserving spatial details, making the ensemble approach particularly effective for depth estimation tasks.
- **Improved Performance:** Hybrid models that incorporate both generative and discriminative components have been shown to outperform single-method models in various tasks [35–37]. This combination harnesses the power of both methodologies, leading to improved performance in reconstructing 3D surfaces from 2D images.

3.3. Loss Function

The loss function used in our network is a composite of reconstruction loss and regularization terms. This combination is designed to ensure that the network not only accurately reconstructs the desired output but also maintains generalization capabilities and prevents overfitting [38]. The reconstruction loss directly measures the difference between the predicted and true values, ensuring precise output generation. Meanwhile, the regularization terms add constraints that promote smoother and more robust model predictions, enhancing the overall performance and stability of the network [39]. The VAE component of the loss

$$\mathcal{L}_V = \|\hat{D}_k^V - D_k\|_2^2 + \beta \text{KL}(q(z|I) \| p(z)) \quad (2)$$

consists of a Mean Squared Error (MSE) reconstruction loss (left) and a Kullback–Leibler (KL) divergence for regularization (right). Here, β is a weighting factor, $q(z|I)$ is the posterior distribution, and $p(z)$ is the prior distribution.

For the modified U-Net component,

$$\mathcal{L}_U = \|\hat{D}_k^U - D_k\|_2^2 + \lambda_{\text{TV}} \sum_{i,j} \sqrt{(\hat{D}_{i+1,j}^U - \hat{D}_{i,j}^U)^2 + (\hat{D}_{i,j+1}^U - \hat{D}_{i,j}^U)^2} \quad (3)$$

there is again an MSE term (left), but additionally a Total Variation (TV) regularization [40] term (right), controlled by weighting factor λ_{TV} which encourages spatial smoothness in the predicted depth map by penalizing large gradients [41] and reducing noise [42], leading to improved generalization and robustness [43].

The final output from the MLP ensemble is optimized using a combined loss function from both VAE and U-Net losses:

$$\mathcal{L}_{\text{final}} = \alpha\mathcal{L}_{\text{V}} + \gamma\mathcal{L}_{\text{U}} \quad (4)$$

Statistically, the integration of these loss terms ensures that the model benefits from both global structural understanding (via VAE) and local detail refinement (via U-Net). The MLP ensemble leverages this dual optimization, effectively balancing bias and variance trade-offs [44]. This approach mitigates the risk of overfitting through regularization while maintaining high reconstruction accuracy, resulting in a model that performs well across varied and unseen data distributions [45].

3.4. Vudnet Discussion

Ensemble methods combine multiple models to improve overall performance by leveraging the strengths of each model [46]. In the context of this work, our results showcase a compelling strategy by ensembling a discriminative model (U-Net) with a generative model (VAE) through an MLP, harnessing the complementary advantages of both types of models. Drawing on the theoretical foundations provided by Bishop and Lasserre's work [47], we can explore the benefits of such an ensemble.

Ensembling helps to mitigate overfitting by combining models that have different error patterns [48]. We let ϵ_{U} and ϵ_{V} be the errors of the U-Net and VAE, respectively. The combined error $\epsilon_{\text{ensemble}}$ can be expressed as

$$\epsilon_{\text{ensemble}} = \beta\epsilon_{\text{U}} + (1 - \beta)\epsilon_{\text{V}} \quad (5)$$

where $0 \leq \beta \leq 1$ is a weight learned by the MLP. By appropriately weighting the errors of U-Net and VAE, the ensemble can reduce the overall error variance. If the errors of U-Net and VAE are uncorrelated, the covariance term drops out, leading to a reduced overall variance. In our work, this is the case because VAE and U-Net approach the problem from different angles: VAE focuses on capturing the global structure and underlying distribution of the data, while U-Net emphasizes local details and spatial relationships. Consequently, the sources of their errors are fundamentally different and likely unrelated. Since ϵ_{V} and ϵ_{U} are not strictly independent, we assume

$$\mathbb{E}[\epsilon_{\text{V}}\epsilon_{\text{U}}] = \delta, \quad \text{where } \delta \text{ is small but not necessarily zero} \quad (6)$$

If δ approaches zero, the covariance term becomes negligible, and the ensemble error approaches the minimum possible value based on the variances of the individual models. The optimal weight β balances the contributions of the VAE and U-Net, leveraging their complementary strengths to reduce the overall error. Even with a non-zero δ , the error $\epsilon_{\text{ensemble}}$ of the ensemble model still has the potential to be lower than the individual errors:

$$\mathbb{E}[\epsilon_{\text{ensemble}}^2] \leq \min(\mathbb{E}[\epsilon_{\text{V}}^2], \mathbb{E}[\epsilon_{\text{U}}^2]) \quad (7)$$

Given an adequate learning process to learn the optimal weights, the ensemble model combines the strengths of both VAE and U-Net, reducing the overall error through weighted averaging.

4. Dataset Development

To the best of our knowledge, there is no public dataset available for specular free-form surface reconstruction; therefore, we generate our own synthetic dataset.

By creating a virtual environment that approximates key optical aspects of real-world conditions for a reflective textured surface according to physical laws of reflection embedded in the Blender Cycles ray-tracing engine, we can systematically explore the nuances of specular surface reconstruction. This section details the precise setup and execution of our simulation environment, ensuring that every aspect, from lighting and camera positioning

to surface properties, is meticulously configured to generate reliable and varied datasets necessary for robust training.

4.1. Setting up the Simulation Environment

The simulation environment was created using Blender, the Cycles ray-tracing engine, for its ability to produce realistic, physically accurate simulations. This level of realism ensures that the synthetic data closely mimic real-world conditions, thereby improving model robustness and reliability. Simple reflection models often fail to account for optical phenomena such as diffraction, scattering, and the intricate interplay of light with micro-surface textures. In contrast, Blender's comprehensive simulation environment allows for the accurate replication of these effects, providing a more robust and realistic dataset.

The Cycles engine has been validated in numerous scientific studies [49–51] for its ability to simulate complex optical phenomena. Additionally, the open-source nature of Blender, its extensive online support, and the convenience of programming through Python make it an ideal choice for automated dataset generation. Despite some limitations with caustics and computational cost, Cycles provides a balanced trade-off between usability and precision, making it suitable for our purposes. By leveraging the advanced features of Blender's Cycles engine, we ensure that the synthesized data appear realistic.

A fixed camera, fixed pattern and surface settings were configured to replicate a realistic scenario for deflectometry, including the use of an orthogonal sinusoidal fringe pattern. This pattern is projected onto a specular object, and the reflected fringes are recorded by a camera. A substantial number of images were captured under these settings, providing a robust foundation for training based on fringe patterns. Figure 3 illustrates this environment for visual clarity.

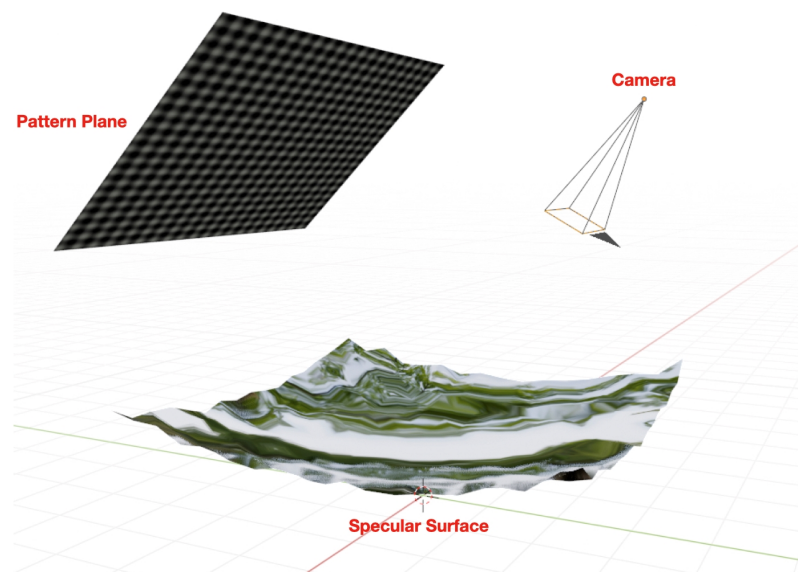


Figure 3. Simulation environment setup for generating the dataset. The environment includes a fixed camera, a fixed pattern, and various surface settings to replicate realistic deflectometry scenarios. An orthogonal sinusoidal fringe pattern is projected onto specular objects, and the reflected fringes are captured by the camera. This setup ensures the generation of a robust and varied dataset, essential for training the VUDNet to accurately reconstruct 3D surfaces from single-shot 2D images. Since this image is a direct screenshot from the Blender environment, the surface is reflecting the simulated world background. The reflection of the pattern plane on the surface is visible to the camera.

4.2. Dataset Generation

To generate the dataset, 2D images were rendered from the simulated environment using Blender's Cycles ray-tracing engine. For each image, the surface was deformed (detailed in Section 4.4) to ensure variability in the dataset with an associated depth map \mathcal{D} .

Each 2D image was rendered based on the specific surface reflectance characteristics, camera settings, and surface topology. The optical ray tracing setup in Cycles was configured with a maximum sample count of 4096, ensuring high-quality image generation. Total bounces were set to 12, diffuse bounces to 4, glossy bounces to 4, transmission bounces to 12, and transparent bounces to 8. In this work, the measurement range is defined as the relative depth of the reconstructed surface, which was normalized to the range $[0, 1]$ during preprocessing. This normalization ensures consistency and stability across varying surface geometries and enables the network to generalize to different scales. As a result, the proposed method can handle surfaces with different depth ranges without requiring additional re-calibration or re-scaling [52].

4.3. Pattern

We utilize a fixed orthogonal sinusoidal fringe pattern \mathcal{P} which remains constant throughout the measurement process. A fixed pattern design is particularly advantageous for industrial applications where the pattern needs to be forged, printed, or otherwise permanently established. Unlike dynamic systems, such as in [22], where an LCD changes the pattern with every shot, our fixed pattern ensures consistency and reliability in scenarios where altering the pattern is impractical or impossible.

The fixed orthogonal sinusoidal fringe pattern \mathcal{P} is represented as

$$\mathcal{P}(x, y) = \mathcal{I}_0 + \mathcal{I}_m \sin\left(\frac{2\pi}{T_x}x + \varphi_x\right) + \mathcal{I}_m \sin\left(\frac{2\pi}{T_y}y + \varphi_y\right) \quad (8)$$

for mean intensity \mathcal{I}_0 , modulation amplitude \mathcal{I}_m , periods T_x, T_y , and phase shifts φ_x, φ_y .

This pattern is projected onto the surface, and the reflected image is captured. The orthogonal nature of the pattern ensures that gradient information is obtained in both x and y directions, which is crucial for 3D reconstruction. By maintaining a fixed pattern, we ensure that our system can be reliably deployed in industrial applications without the need for complex pattern generation mechanisms, providing a stable and efficient solution for 3D surface measurement.

4.4. Shape Deformations

To have a diverse array of geometric features, we systematically introduced deformations to a plane surface $\bar{S} \in \mathbb{R}^2$:

$$\Pi : \bar{S} \times \Theta \rightarrow \mathcal{S}, \mathcal{S}_\theta = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z) = \Pi(u, v; \theta), (u, v) \in \bar{S}, \theta \in \Theta\} \quad (9)$$

Here, Θ represents the set of parameters that conditions the projection. These deformations were designed to simulate realistic scenarios and to increase the likelihood of network generalizability from training data to real-world applications.

4.4.1. Surface Generation

The surface generation process involved modifying a base planar surface by integrating various shapes, such as hemispheres, to resemble geometric structures and by applying parametric deformations to create deformation forms. Each surface was assigned a specular material to mimic realistic lighting interactions, crucial for representing real-world scenarios.

1. **Shape Integration:** To generate the *Geometric Surfaces* set, multiple convex shapes, primarily hemispheres, were superimposed onto a base planar surface \bar{S} at randomized positions and scales.
2. **Parametric Deformations:** To generate the *Deformation Surfaces* set, the planar surface was deformed using a set of parametric functions Π , such as parabolic and sinusoidal transformations. These functions control the depth and curvature of the deformation, sculpting the surface into various forms.

3. **Randomization:** To enhance diversity, the parameters involved in both geometric shape integration (size θ_1 , position θ_2 , and the number of hemispheres θ_3) and deformations were randomized, resulting in a wide range of surface topologies \mathcal{S}_θ , featuring variations from subtle indentations to pronounced curvatures.
4. **Specular Material Assignment:** All generated surfaces \mathcal{S} were assigned a specular material to ensure that the generated data accurately represented real-world scenarios, where surface reflectivity significantly impacts depth perception.

4.4.2. Data Capture

Following the creation of these surfaces, we rendered 2D images and captured corresponding depth maps. This process involved two steps:

1. **Rendering 2D Images:** Each deformed surface was rendered under controlled lighting conditions to capture the specular reflections characteristic of real-world environments.
2. **Depth Map Acquisition:** Depth maps \mathcal{D} were generated for each configuration, providing the precise geometric ground-truth information essential for training.

This methodical approach to generating a diverse and realistic dataset enables our deep learning model to learn robust features applicable across a wide range of real-world surfaces.

4.5. Depth Map Standardization

Depth maps generated from simulations often require standardization to ensure consistency and accuracy in subsequent processing steps. Algorithm 1 illustrates the process we employ for this standardization. The algorithm begins with the calculation of the centroid of the depth map, providing a central reference point for further transformations. Principal Component Analysis (PCA) is then used to determine the primary axes of variance in the data, identifying the principal components that describe the orientation of the surface in the 3D space. The PCA eigenvectors lead to a rotation matrix, ensuring that the surface data are oriented with the XY plane, with the Z-axis representing height. The transformation aligns the centroid of the depth map to the origin, standardizing the position and orientation. Finally, the depth values are normalized to a predefined range of $[0, 1]$.

Algorithm 1 Normalize and Align Depth Map

- 1: **procedure** NORMALIZEDDEPTHMAP(*DepthMap*)
 - 2: **Centroid Calculation:**
 - 3: $C \leftarrow \frac{1}{N} \sum_{i=1}^N \mathcal{D}_k^i$ ▷ Calculate the centroid C of the depth map \mathcal{D}
 - 4: **Principal Component Analysis (PCA):**
 - 5: $\Sigma \leftarrow \frac{1}{N-1} \sum_{i=1}^N (\mathcal{D}_k^i - C)(\mathcal{D}_k^i - C)^T$ ▷ Compute the covariance matrix Σ
 - 6: $[\vec{v}_1, \vec{v}_2, \vec{v}_3] \leftarrow \text{Eigenvectors}(\Sigma)$ ▷ Eigenvectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ of Σ
 - 7: **Rotation Calculation:**
 - 8: **Normalization and Vector Operations:**
 - 9: Normalize the normal vector $\vec{n} = \frac{\vec{v}_3}{\|\vec{v}_3\|}$
 - 10: $\vec{v} = \vec{n} \times [0, 0, 1]$ ▷ Cross product with Z-axis
 - 11: $s = \|\vec{v}\|, c = \vec{n} \cdot [0, 0, 1]$ ▷ Sine and cosine of the angle
 - 12: **Skew-Symmetric Matrix and Rodrigues' Formula:**
 - 13: $\text{Skew}(\vec{v}) = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$
 - 14: $R = I + \text{Skew}(\vec{v}) + \text{Skew}(\vec{v})^2 \frac{1-c}{s^2}$ ▷ Compute the rotation matrix
 - 15: **Transformation Application:**
 - 16: $\mathcal{D}_k^i \leftarrow R(\mathcal{D}_k^i - C)$ ▷ Apply the rotation and translation to each point in the depth map
-

Algorithm 1 *Cont.*

```

17: Normalization:
18:  $a, b \leftarrow$  range values (0, 1)
19:  $\mathcal{D}_k^{i''} \leftarrow a + \frac{(\mathcal{D}_k^{i'} - \min(\mathcal{D}_k^{i'}))(\max(\mathcal{D}_k^{i'}) - \min(\mathcal{D}_k^{i'}))}{b - a}$   $\triangleright$  Normalize the depth values to the
    range  $[a, b]$ 
20: return NormalizedMap  $\leftarrow \mathcal{D}_k^{i''}$ 
21: end procedure

```

4.6. Data Preparation

The dataset used in our study consists of a diverse set of depth maps generated from simulated environments, totaling 400 images. We include 250 deformation surface samples and 150 geometric ones, as our research indicated that deformation surfaces present more complexity, and so were emphasized in our simulations. Each image is 320×240 pixels in size and is rendered in greyscale. The dataset is divided into training, validation, and test sets, with 80% used for training, 10% for validation, and 10% for testing. This division ensures a robust evaluation of the model's performance across different data splits and helps in assessing generalization.

5. Results and Discussion

All experiments were conducted on a system equipped with an NVIDIA GeForce RTX 3060 GPU, 12th Gen Intel(R) Core(TM) i7-12700 CPU, and 32 GB of RAM. The training process for the VUDNet was implemented using TensorFlow (version 2.12.0) and Keras (version 2.12.0). The implementation of VUDNet is available on GitHub at <https://github.com/hadious/VUDNet>, accessed on 16 November 2024. The entire training phase took approximately 4 h to complete. Inference times are reported in Table 1.

Table 1. Performance comparison of different models on our generated dataset. VUDNet has reconstruction errors approximately one fourth that of competing methods, under all three criteria of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Log Error, and with an inference time (computational complexity) similar to or less than competing methods. Bolded numbers indicate the best performance on the General data set.

Model	MAE	RMSE	LogError	Inference Time (ms)
VUDNet (On Deformation)	0.0443	0.0600	0.0122	19
VUDNet (On Geometric)	0.0168	0.0218	0.0038	19
VUDNet (General)	0.0355	0.0470	0.0090	19
DYNet++ * (General)	0.1607	0.1987	0.0790	26
D-UNet ** (General)	0.2052	0.2245	0.0451	21

* DYNet++ plus reconstructing the surface by zonal integration for reconstructing the surfaces as mentioned in [17]. Since there are no publicly available data on the DYNET++ dataset, we were unable to test on their dataset, so we implemented their method on our dataset. ** For D-UNet [29], as there are no publicly available data or code, we implemented it based on [29] and tested it on our proposed dataset.

Figure 4 qualitatively demonstrates the success of our model. The predicted depth map exhibits a high degree of smoothness, indicating effective regularization and the absence of noise or artifacts. The image shows no visible outliers, suggesting consistent accuracy across the surface. Most significantly, there is no discernible presence of the orthogonal fringe pattern that would have been strikingly present in the input image, highlighting VUNet's ability to reconstruct the underlying surface geometry without retaining or passing through the pattern information.

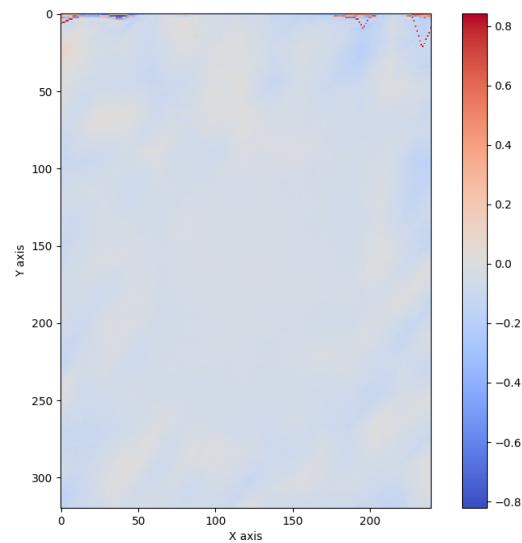


Figure 4. Mean absolute difference between the ground truth depth map and the predicted depth map from our VUDNet model for a selected sample. The error pattern demonstrates smoothness, effective regularization, and an overall minimal presence of outliers. The top region contains localized patterns of higher error values (illustrated as collection of red pixels), however despite these localized artifacts, the remainder of the image demonstrates the model’s accuracy with no visible orthogonal fringe patterns.

Figure 5 presents a comparison between the ground truth and predicted depth maps for a randomly selected sample. The left images display the ground truth depth map, while the right images show the depth map predicted by our VUDNet model. The predicted depth map is smoother than the original, due to the regularization term. Nevertheless, the estimates still retain a high degree of fine detail. The visual comparison suggests that the model effectively captured the essential features of the surface, aligning well with the expected characteristics observed in the ground truth data. This smoothing effect helps to manage the noise and artifacts typically associated with specular surfaces.

The t-SNE (t-distributed Stochastic Neighbor Embedding) method [53] is a powerful dimensionality reduction technique commonly used to visualize high-dimensional data in a two- or three-dimensional space. By applying t-SNE, we can effectively observe how well-separated the latent space is, learned by the VAE. Figure 6 demonstrates such a separation of the representation space, whereby the clustering effect indicates that the VAE successfully formed a coherent representation space where intrinsic features of different surfaces are well captured and recognizable. In this context, a cluster in the VAE’s representation space refers to a grouping of data points (input images) that share similar underlying features. Ideally, cluster separation is evidence of meaningful feature learning, as it implies that the VAE is able to effectively distinguish between various surface characteristics.

In Figure 6, the well-defined clusters suggest that the VAE can identify and learn the underlying patterns in the data, leading to potent feature extraction. This capability is essential for generalizing the network’s performance across diverse and complex surface geometries.

Figure 7 builds on Figure 6, but now based on a network trained exclusively on the deformation data as opposed to the entire dataset in Figure 6. The scattering of the samples based on their features shows that similar data points (similar surface shapes) are located near to each other in the latent space. This proximity indicates that the VAE can effectively group similar surfaces, enhancing the network’s ability to perform accurate depth estimation. The close grouping of similar samples demonstrates the VAE’s proficiency in feature extraction, ensuring that surfaces with similar characteristics are represented similarly.

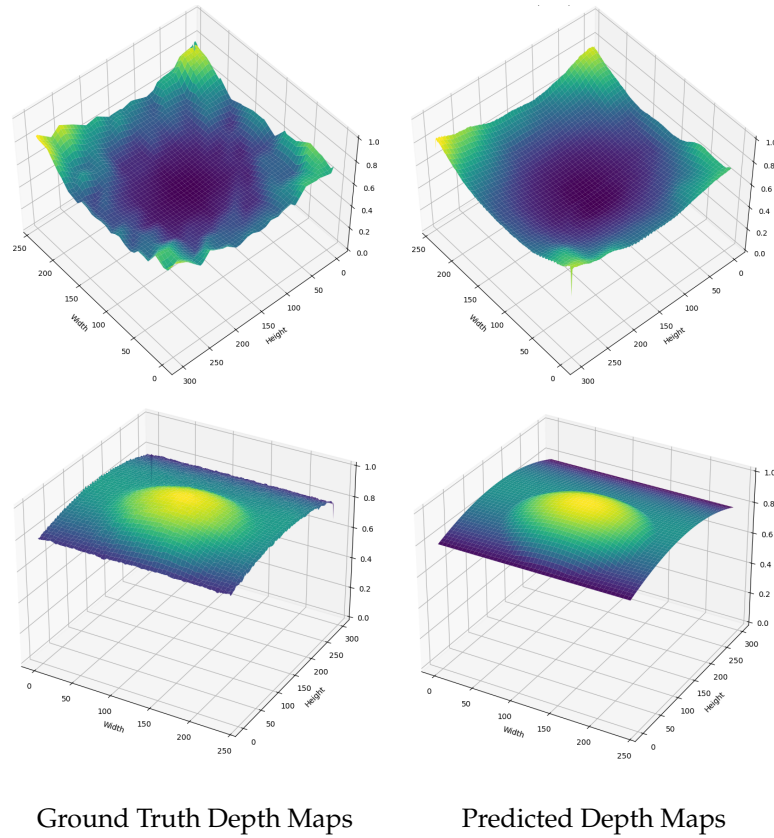


Figure 5. Comparison of ground truth (left) and VUDNet-estimated depth maps (right), showing effective noise reduction and fine detail retention. The top images showcase a result for a deformation sample, while the bottom images represent an example of the geometric case.

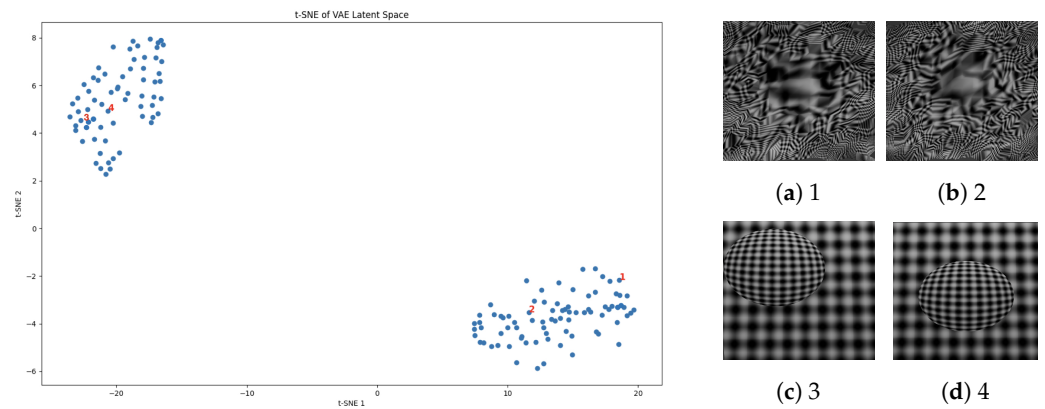


Figure 6. VAE representation (left) showcasing distinct separation of clusters in the latent space, visualized as the first and second components of t-SNE. The clustering indicates effective differentiation of surface characteristics and potent feature extraction. The four images on the right correspond to the selected points in the latent space (right). It is evident that images from a given cluster share related surface characteristics, highlighting the network’s ability to identify underlying similarities despite variations in surface characteristics.

The VAE’s ability to develop such clusters and similarity behaviours signifies its strength in capturing and representing the essential features of the input data. This representation plays a critical role in the subsequent stages of depth estimation, where accurate and reliable feature extraction is paramount.

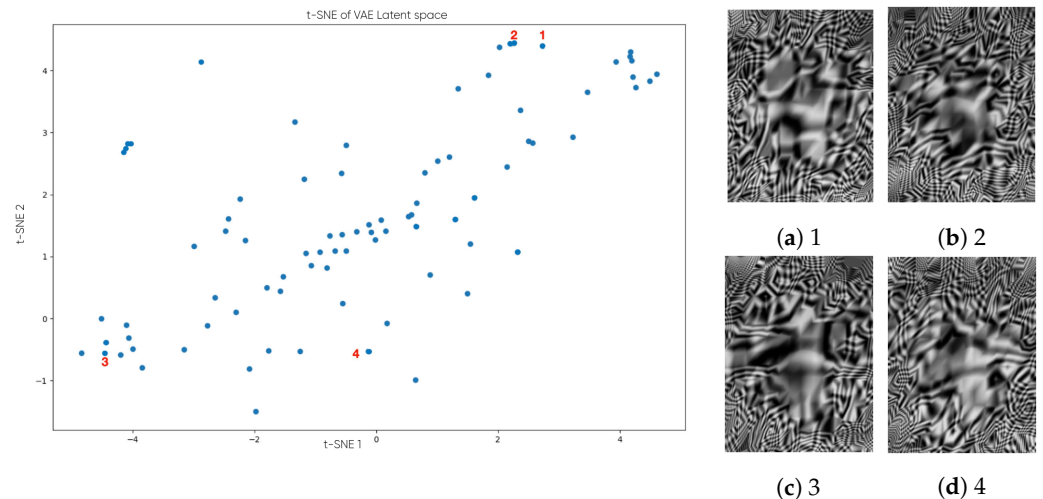


Figure 7. The image panels here are organized the same as in Figure 6, with the latent space (left) visualized from the first and second components of t-SNE, and the images (right) corresponding to the selected points in the latent space. The difference is that Figure 6 was trained on the entire dataset, whereas here it is trained exclusively on the deformation data. It is clear that images positioned closer in the latent space share more similarities in reflection shape, while those farther apart are less similar, even though they belong to the same overall category of deformation surfaces. This emphasizes the network’s capability to capture underlying similarities despite variations in surface characteristics within the same category.

Although our method employs an ensemble of a VAE and U-Net, we focus on visualizing the latent representation of the VAE. As highlighted in prior research [32], VAEs are designed to learn compact, informative latent spaces [54–56] that capture the underlying structure of the data. In contrast, U-Net primarily performs pixel-level refinement, working directly in the image space without explicitly learning a latent representation [33]. Therefore, the U-Net’s contribution is more appropriately assessed through its impact on the final depth maps, rather than through latent space visualization.

Table 1 demonstrates the remarkable performance of VUDNet. The quantitative metrics presented in the table demonstrate the network’s capability to achieve high accuracy and consistency in depth estimation. Specifically, the VUDNet model shows significantly lower Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and LogError compared to the comparator DYNNet++ [17] and D-UNet [29] models. The lower errors across all metrics, at the same time at reduced computational complexity, highlight the superior performance of the VUDNet’s hybrid architecture in providing precise, reliable, and efficient 3D surface reconstructions.

The effective combination of VAE and U-Net components, along with the MLP ensemble, allows VUDNet to capture intricate surface details while maintaining overall structural coherence. The distinct separation in the VAE representation space indicates successful feature extraction, validating the model’s ability to generalize across diverse and complex surface geometries.

6. Conclusions

In this paper, we introduced VUDNet, an innovative deep neural network designed for the end-to-end 3D reconstruction of specular free-form surfaces using single-shot deflectometry. Our approach uniquely combines the strengths of both generative and discriminative models by integrating a Variational Autoencoder (VAE) and a modified U-Net. This hybrid architecture excels in depth estimation and detail refinement, providing high-fidelity reconstructions that surpass the performance of existing methods. Our dataset, which allows for network training on a diverse range of scenarios, will be made publicly available to support future work in this domain.

Experimental results demonstrate that VUDNet achieves incredible reconstruction accuracy from single-shot 2D images, with an RMSE error one-fourth as large (or better) and at reduced computational complexity in comparison to competing state-of-the-art methods. The integration of both MSE and TV regularization terms in the loss function promotes spatial smoothness and reduces noise, leading to visually coherent depth maps. The findings of this study set a new standard for single-shot deflectometry, showcasing the potential of deep learning in advancing optical metrology for specular surfaces.

In future work, we plan to explore the application of VUDNet to other forms of surface measurement and extend our approach to dynamic and real-time scenarios. Additionally, we aim to investigate the integration of more advanced generative models to further improve the robustness of depth estimation in diverse and complex environments.

Author Contributions: Conceptualization, M.S., S.M. and P.F.; Formal analysis, M.S.; Funding acquisition, C.P., A.M.D.L. and P.F.; Methodology, M.S., S.M. and A.N.; Project administration, P.F.; Resources, C.P. and A.M.D.L.; Software, M.S.; Supervision, P.F.; Validation, M.S.; Visualization, M.S.; Writing—original draft, M.S.; Writing—review and editing, S.M., A.N., C.P., A.M.D.L. and P.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work is financially supported by the MITACS Accelerate program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors gratefully acknowledge the financial support of the MITACS Accelerate program.

Conflicts of Interest: Authors Claire Preston, Anthony M. D. Lee were employed by the company General Fusion Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kwak, H.; Kim, J. Semiconductor multilayer nanometrology with machine learning. *Nanomanufacturing Metrol.* **2023**, *6*, 15. [[CrossRef](#)]
2. Li, T.; Wang, S.; Luo, Y.; Wan, J.; Luo, Z.; Chen, M. 3D Vision and Intelligent On-line Inspection in SMT Microelectronic Packaging: A Review. *IEEE J. Emerg. Sel. Top. Ind. Electron.* **2024**, *5*, 779–789. [[CrossRef](#)]
3. Jangra, P.; Duhan, M. Comparative analysis of devices working on optical and spintronic based principle. *J. Opt.* **2024**, *53*, 1629–1649. [[CrossRef](#)]
4. Flores-Fuentes, W.; Arellano-Vega, E.; Sergiyenko, O.; Alba-Corpus, I.Y.; Rodríguez-Quinonez, J.C.; Castro-Toscano, M.J.; González-Navarro, F.F.; Vasavi, S.; Miranda-Vega, J.E.; Hernández-Balbuena, D.; et al. Surface color estimation in 3D spatial coordinate remote sensing by a technical vision system. *Opt. Quantum Electron.* **2024**, *56*, 406. [[CrossRef](#)]
5. Li, T.; Polette, A.; Lou, R.; Jubert, M.; Nozais, D.; Pernot, J.P. Machine Learning-Based 3D Scan Coverage Prediction for Smart-Control Applications. *Comput.-Aided Des.* **2024**, *176*, 103775. [[CrossRef](#)]
6. Prauzek, M.; Hercik, R.; Konecny, J.; Mikolajek, M.; Stankus, M.; Koziorek, J.; Martinek, R. An optical-based sensor for automotive exhaust gas temperature measurement. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [[CrossRef](#)]
7. Rak, G.; Hočevar, M.; Kolbl Repinc, S.; Novak, L.; Bizjan, B. A review on methods for measurement of free water surface. *Sensors* **2023**, *23*, 1842. [[CrossRef](#)]
8. Zhang, T.; Xia, R.; Zhao, J.; Wu, J.; Fu, S.; Chen, Y.; Sun, Y. Low coherence measurement methods for industrial parts with large surface reflectance variations. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 7006514. [[CrossRef](#)]
9. Burke, J.; Pak, A.; Höfer, S.; Ziebarth, M.; Roschani, M.; Beyerer, J. Deflectometry for specular surfaces: An overview. *Adv. Opt. Technol.* **2023**, *12*, 1237687. [[CrossRef](#)]
10. Huang, L.; Idir, M.; Zuo, C.; Asundi, A. Review of phase measuring deflectometry. *Opt. Lasers Eng.* **2018**, *107*, 247–257. [[CrossRef](#)]
11. Häusler, G.; Faber, C.; Olesch, E.; Ettl, S. Deflectometry vs. interferometry. In Proceedings of the Optical Measurement Systems for Industrial Inspection VIII, Munich, Germany, 13–17 May 2013; Volume 8788, pp. 367–377.
12. 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](#)]
13. Wójcik, A.; Niemczewska-Wójcik, M.; Śladek, J. Assessment of free-form surfaces' reconstruction accuracy. *Metrol. Meas. Syst.* **2017**, *24*, 303–312. [[CrossRef](#)]

14. Orumi, M.A.B.; Sepanj, M.H.; Famouri, M.; Azimifar, Z.; Wong, A. Unsupervised Deep Shape from Template. In Proceedings of the Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, 27–29 August 2019; Proceedings, Part I 16; Springer: Berlin/Heidelberg, Germany, 2019; pp. 440–451.
15. Jiang, X.J.; Scott, P.J. *Advanced Metrology: Freeform Surfaces*; Academic Press: Cambridge, MA, USA, 2020.
16. Qiao, G.; Huang, Y.; Song, Y.; Yue, H.; Liu, Y. A single-shot phase retrieval method for phase measuring deflectometry based on deep learning. *Opt. Commun.* **2020**, *476*, 126303. [[CrossRef](#)]
17. Nguyen, M.T.; Ghim, Y.S.; Rhee, H.G. DYnet++: A deep learning based single-shot phase-measuring deflectometry for the 3D measurement of complex free-form surfaces. *IEEE Trans. Ind. Electron.* **2023**, *71*, 2112–2121. [[CrossRef](#)]
18. Liang, H.; Sauer, T.; Faber, C. Using wavelet transform to evaluate single-shot phase measuring deflectometry data. In Proceedings of the Applications of Digital Image Processing XLIII, Online, 24 August–4 September 2020; Volume 11510, pp. 404–410.
19. Mangione, N.S.; Wu, H.; Preston, C.; Lee, A.M.; Entezami, S.; Ségas, R.; Forysinski, P.W.; Suponitsky, V. Shape manipulation of a rotating liquid liner imploded by arrays of pneumatic pistons: Experimental and numerical study. *Fusion Eng. Des.* **2024**, *198*, 114087. [[CrossRef](#)]
20. Qian, J.; Feng, S.; Li, Y.; Tao, T.; Han, J.; Chen, Q.; Zuo, C. Single-shot absolute 3D shape measurement with deep-learning-based color fringe projection profilometry. *Opt. Lett.* **2020**, *45*, 1842–1845. [[CrossRef](#)]
21. Chang, H.T.; Lin, T.Y.; Chuang, C.H.; Chen, C.Y.; Ho, C.C.; Chang, C.Y. Separation of two-dimensional mixed circular fringe patterns based on spectral projection property in fractional Fourier transform domain. *Appl. Sci.* **2021**, *11*, 859. [[CrossRef](#)]
22. Wu, Z.; Wang, J.; Jiang, X.; Fan, L.; Wei, C.; Yue, H.; Liu, Y. High-precision dynamic three-dimensional shape measurement of specular surfaces based on deep learning. *Opt. Express* **2023**, *31*, 17437–17449. [[CrossRef](#)]
23. Dupont, E.; Whye Teh, Y.; Doucet, A. Generative Models as Distributions of Functions. In Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, Virtual, 28–30 March 2022; Camps-Valls, G., Ruiz, F.J.R., Valera, I., Eds.; Proceedings of Machine Learning Research; PMLR: Birmingham, UK, 2022; Volume 151, pp. 2989–3015.
24. Lavrač, N.; Podpečan, V.; Robnik-Šikonja, M. *Representation Learning: Propositionalization and Embeddings*; Springer: Berlin/Heidelberg, Germany, 2021.
25. Nguyen, M.T.; Ghim, Y.S.; Rhee, H.G. One-shot deflectometry for high-speed inline inspection of specular quasi-plane surfaces. *Opt. Lasers Eng.* **2021**, *147*, 106728. [[CrossRef](#)]
26. Wang, J.; Wang, T.; Xu, B.; Willomitzer, O.C. Accurate Eye Tracking from Dense 3D Surface Reconstructions using Single-Shot Deflectometry. *arXiv* **2023**, arXiv:2308.07298.
27. Li, W.; Liu, T.; Tai, M.; Zhong, Y. Three-dimensional measurement for specular reflection surface based on deep learning and phase measuring profilometry. *Optik* **2022**, *271*, 169983. [[CrossRef](#)]
28. Suresh, V.; Zheng, Y.; Li, B. PMENet: Phase map enhancement for Fourier transform profilometry using deep learning. *Meas. Sci. Technol.* **2021**, *32*, 105001. [[CrossRef](#)]
29. Dou, J.; Wang, D.; Yu, Q.; Kong, M.; Liu, L.; Xu, X.; Liang, R. Deep-learning-based deflectometry for freeform surface measurement. *Opt. Lett.* **2022**, *47*, 78–81. [[CrossRef](#)] [[PubMed](#)]
30. Lopez, R.; Regier, J.; Jordan, M.I.; Yosef, N. Information constraints on auto-encoding variational bayes. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6117–6128.
31. Pu, Y.; Gan, Z.; Henaou, R.; Yuan, X.; Li, C.; Stevens, A.; Carin, L. Variational autoencoder for deep learning of images, labels and captions. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2360–2368.
32. Pinheiro Cinelli, L.; Araújo Marins, M.; Barros da Silva, E.A.; Lima Netto, S. Variational autoencoder. In *Variational Methods for Machine Learning with Applications to Deep Networks*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 111–149.
33. 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—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
34. Dietterich, T.G. Ensemble methods in machine learning. In Proceedings of the International Workshop on Multiple Classifier Systems, Berlin/Heidelberg, Germany, 20–23 June 2000; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
35. Ye, Y.; Ji, S. A Hybrid Generative and Discriminative PointNet on Unordered Point Sets. *arXiv* **2024**, arXiv:2404.12925.
36. Garcia Satorras, V.; Akata, Z.; Welling, M. Combining generative and discriminative models for hybrid inference. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 13820–13830.
37. Kou, G.; Chen, H.; Hefni, M.A. Improved hybrid resampling and ensemble model for imbalance learning and credit evaluation. *J. Manag. Sci. Eng.* **2022**, *7*, 511–529. [[CrossRef](#)]
38. Roelofs, R. *Measuring Generalization and Overfitting in Machine Learning*; University of California: Berkeley, CA, USA, 2019.
39. Ying, X. An overview of overfitting and its solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [[CrossRef](#)]
40. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4484–4502. [[CrossRef](#)]
41. Kobler, E.; Effland, A.; Kunisch, K.; Pock, T. Total deep variation: A stable regularization method for inverse problems. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 9163–9180. [[CrossRef](#)] [[PubMed](#)]
42. Baiju, P.; Antony, S.L.; George, S.N. An intelligent framework for transmission map estimation in image dehazing using total variation regularized low-rank approximation. *Vis. Comput.* **2022**, *38*, 2357–2372. [[CrossRef](#)]

43. Ibrahim, M.M.; Liu, Q.; Khan, R.; Yang, J.; Adeli, E.; Yang, Y. Depth map artefacts reduction: A review. *IET Image Process.* **2020**, *14*, 2630–2644. [[CrossRef](#)]
44. Aotani, T.; Kobayashi, T.; Sugimoto, K. Meta-optimization of bias-variance trade-off in stochastic model learning. *IEEE Access* **2021**, *9*, 148783–148799. [[CrossRef](#)]
45. Pourtaheri, Z.K.; Zahiri, S.H. Ensemble classifiers with improved overfitting. In Proceedings of the 2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIIEC), Bam, Iran, 9–11 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 93–97.
46. Abimannan, S.; El-Alfy, E.S.M.; Chang, Y.S.; Hussain, S.; Shukla, S.; Satheesh, D. Ensemble multifeatured deep learning models and applications: A survey. *IEEE Access* **2023**, *11*, 107194–107217. [[CrossRef](#)]
47. Bernardo, J.; Bayarri, M.; Berger, J.; Dawid, A.; Heckerman, D.; Smith, A.; West, M. Generative or discriminative? getting the best of both worlds. *Bayesian Stat.* **2007**, *8*, 3–24.
48. Ganaie, M.A.; Hu, M.; Malik, A.K.; Tanveer, M.; Suganthan, P.N. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151. [[CrossRef](#)]
49. Koch, M.; Rosselló, J.M.; Lechner, C.; Lauterborn, W.; Eisener, J.; Mettin, R. Theory-assisted optical ray tracing to extract cavitation-bubble shapes from experiment. *Exp. Fluids* **2021**, *62*, 1–19. [[CrossRef](#)]
50. Kiuchi, S.; Koizumi, N. Simulating the appearance of mid-air imaging with micro-mirror array plates. *Comput. Graph.* **2021**, *96*, 14–23. [[CrossRef](#)]
51. Villa, J.; McMahan, J.; Nesnas, I. Image Rendering and Terrain Generation of Planetary Surfaces Using Source-Available Tools. In Proceedings of the 46th Annual AAS Guidance, Navigation & Control Conference, Breckenridge, CO, USA, 2023; pp. 1–24.
52. Eigen, D.; Puhersch, C.; Fergus, R. Depth map prediction from a single image using a multi-scale deep network. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2366–2374.
53. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
54. Yee-King, M. Latent spaces: A creative approach. In *The Language of Creative AI: Practices, Aesthetics and Structures*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 137–154.
55. Gelada, C.; Kumar, S.; Buckman, J.; Nachum, O.; Bellemare, M.G. Deepmdp: Learning continuous latent space models for representation learning. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 2170–2179.
56. Fries, W.D.; He, X.; Choi, Y. Lasdi: Parametric latent space dynamics identification. *Comput. Methods Appl. Mech. Eng.* **2022**, *399*, 115436. [[CrossRef](#)]

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