OpticFusion: Multi-Modal Neural Implicit 3D Reconstruction of Microstructures by Fusing White Light Interferometry and Optical Microscopy

Shuo Chen   Yijin Li   Guofeng Zhang[†]

State Key Lab of CAD&CG, Zhejiang University

{chenshuo.eric, eugenelee, zhangguofeng}@zju.edu.cn

Abstract

White Light Interferometry (WLI) is a precise optical tool for measuring the 3D topography of microstructures. However, conventional WLI cannot capture the natural color of a sample's surface, which is essential for many microscale research applications that require both 3D geometry and color information. Previous methods have attempted to overcome this limitation by modifying WLI hardware and analysis software, but these solutions are often costly. In this work, we address this challenge from a computer vision multi-modal reconstruction perspective for the first time. We introduce OpticFusion, a novel approach that uses an additional digital optical microscope (OM) to achieve 3D reconstruction with natural color textures using multi-view WLI and OM images. Our method employs a two-step data association process to obtain the poses of WLI and OM data. By leveraging the neural implicit representation, we fuse multi-modal data and apply color decomposition technology to extract the sample's natural color. Tested on our multi-modal dataset of various microscale samples, OpticFusion achieves detailed 3D reconstructions with color textures. Our method provides an effective tool for practical applications across numerous microscale research fields. The source code and our real-world dataset are available at https://github.com/zju3dv/OpticFusion.

1. Introduction

The 3D structure of microscale samples is crucial in numerous scientific research fields and industrial production processes. White light interferometry (WLI) [9, 49] is widely used for high-precision 3D surface topography measurement, offering a lateral resolution of approximately 0.5 micrometers and a vertical accuracy at the sub-nanometer scale across diverse surfaces. Consequently, WLI is utilized in ultra-precision machining [22, 25, 55], integrated circuit inspection [8, 15, 38], biological structure analy-

sis [10, 37, 40], and other applications. In WLI (Fig. 1a), broadband white light with a very short coherence length serves as the light source. A beamsplitter divides the light into a reference beam and a measurement beam, directed toward a reference mirror and the sample surface. High-contrast interference fringes are captured by a charged-coupled device (CCD) camera only when the optical path difference (OPD) is close to zero. As the piezoelectric tube (PZT) scans along the Z-axis, the zero-OPD position shifts according to the sample's 3D structure. By determining the Z-value of the PZT that maximized the interference intensity at each point on the sample surface during the scanning, the corresponding 3D topography of the microstructure's surface can be obtained (Fig. 1b).

Although WLI uses white light to illuminate the sample, conventional WLI instruments typically employ monochrome CCD cameras to capture grayscale interference patterns, resulting in the loss of the sample's color information. However, the natural color of the sample can be valuable in many applications, such as observing color changes on laser-treated metal surfaces [2], identifying defects and corrosion on samples [42], and distinguishing boundaries between different materials [16]. Currently, several methods exist to obtain color images in WLI. One direct approach is to use a three-chip color CCD camera [26] or a single-chip color CCD camera with a Bayer filter [12, 20]. However, this approach significantly impacts WLI's lateral resolution, accuracy, and throughput [5]. Another method employs switchable RGB light sources [5]. Generally, these approaches require modifications to the WLI hardware and interference pattern analysis software, making them impractical for existing commercial monochromatic WLI instruments. An alternative approach [18, 27] utilizes a reference sample with known reflectance and Fourier spectral analysis of the monochromatic interference patterns to obtain the sample's reflectance spectrum. Nonetheless, some commercial WLI instruments only provide the final topography map and do not allow users to access the raw interference patterns. In this paper, we propose a novel approach that does not rely on costly hardware modifications to WLI.
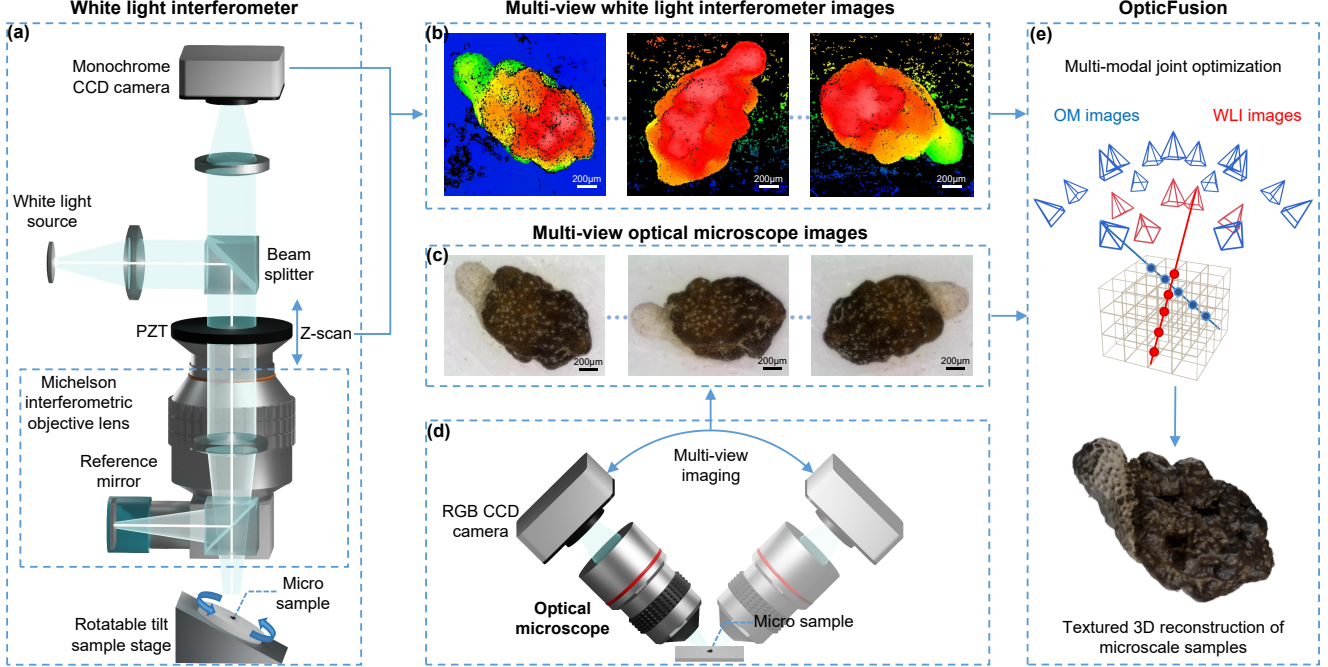

[†]Corresponding Author

White light interferometer

(a)

Monochrome CCD camera

White light source

Beam splitter

PZT — Z-scan

Michelson interferometric objective lens

Reference mirror

Micro sample

Rotatable tilt sample stage

Multi-view white light interferometer images

(b)

200μm

Multi-view optical microscope images

(c)

200μm

(d)

Multi-view imaging

RGB CCD camera

Optical microscope

Micro sample

OpticFusion

(e)

Multi-modal joint optimization

OM images   WLI images

Textured 3D reconstruction of microscale samples

Figure 1. **Multi-modal neural implicit reconstruction of microscale samples with multi-view images from white light interferometer and optical microscope.** **(a)** The principle of multi-view WLI scanning. **(b, c)** The output of white light interferometer and optical microscope. **(d)** The multi-view imaging with an optical microscope. **(e)** OpticFusion reconstructs 3D models of microstructures with their natural surface colors, without costly modifications of the WLI system.

For the first time, we introduce a multi-modal 3D reconstruction method utilizing only a commercial WLI instrument and a digital optical microscope (OM, Fig. 1d). This method enables accurate 3D reconstruction of microscale samples with their natural surface colors.

We propose OpticFusion, a method for neural implicit surface reconstruction of microscale samples using multi-view WLI (Fig. 1b) and OM images (Fig. 1c). Unlike typical multi-sensor systems with data alignment or fixed relative poses, our approach acquires WLI and OM images independently. Therefore, we employ a two-step data association method to determine the poses of all WLI and OM images within an absolute-scale coordinate system. In the reconstruction process, we jointly optimize a neural network-based Signed Distance Function (SDF) using both multi-view WLI and OM images (Fig. 1e). This approach allows OM images to fill voids where WLI data lack measurements, enhancing the reconstruction quality and producing complete, detailed surface models of microscale samples. Notably, in microscale research, the focus is not on synthesizing novel views but on capturing the intrinsic colors of the sample itself, independent of the view direction. To achieve this, we incorporate color decomposition strategies from Intrinsic-NeRF [51] and Color-NeuS [56], adding a view-dependent residual branch into the color network, which results in more natural colors for the model texture.

To validate the effectiveness of our method, we collect a real dataset consisting of multi-view WLI and OM images of samples, with surface feature sizes ranging from tens to hundreds of micrometers, including a butterfly wing, flower seeds, a circuit board, and a microsensor. The results show that OpticFusion accurately reconstructs the surface details of various samples while texturing natural colors to the model. Additionally, we simulate the imaging characteristics of WLI to create a synthetic dataset. We demonstrate the superiority of combining WLI and OM data by quantitatively evaluating the reconstruction quality. We also show that our reconstructed results can be used for roughness measurement of microscale surfaces and detailed analysis of biological samples.

Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to propose a method for textured 3D reconstruction of microstructures using only a commercial WLI and an OM, without costly hardware modifications needed in previous methods.
- We propose a novel pipeline that includes a two-step data association method and a neural implicit representation with a residual term to fuse multi-modal data, enabling accurate geometric reconstruction and natural surface color acquisition.
- We evaluate the proposed method on a dataset of multi-view WLI and OM data of various real-world microscale samples, as well as on a synthetic dataset, demonstrating the effectiveness and practical value of our approach.

2. Related Work

Multi-Modal 3D Reconstruction. Multi-modal 3D reconstruction methods enhance the accuracy and speed or provide richer 3D information by combining data from different sensors and leveraging their unique advantages. Beyond commonly used RGB cameras, these methods also incorporate sensors such as depth cameras [3, 24, 31, 32], Lidar [21], infrared (IR) [43] and thermal imaging [14, 33]. In our work, OM and WLI can be approximated as a combination of an RGB camera and a depth sensor. Conventional RGBD reconstruction methods use depth information for model reconstruction [31, 32] and RGB for subsequent appearance reconstruction steps to texture the 3D model [47, 57]. Recent methods [3, 24, 58] use neural networks to implicitly represent 3D models, enabling the fusion of multi-sensor information into the 3D model through an end-to-end optimization process. Our work shifts the focus from macroscale to microscale objects. Currently, research on 3D reconstruction of microscale surfaces is relatively limited, with existing studies [7, 44] typically relying on a single instrument. In microscale research, a range of available instruments exhibit entirely different characteristics. Our work is the first attempt to fuse WLI and OM data, exploring the feasibility of these multi-modal 3D reconstruction methods in the microscale domain.

Neural Implicit Representation. Neural implicit representations are widely applied to various tasks. DeepSDF [34] employs neural networks to represent 3D shapes as signed distance functions. NeRF [29] and its subsequent works [4, 28, 52] model scenes as radiance fields, allowing for realistic novel view synthesis. Some approaches [51, 54] have implemented the decomposition of lighting and color attributes within a scene, thereby supporting more flexible applications. Further developments, such as NeuS [48] and VolSDF [50], focus on accurate surface reconstruction, enhancing the fidelity of 3D models. Other works [23, 35] extend these representations to dynamic scenes, enabling the capture of temporal changes. Moreover, neural implicit representations have proven effective in handling data from other modalities, such as CT [53] and MRI [41]. Our work further demonstrates the effectiveness of neural implicit representations in fusing multi-view WLI and OM data. Our method captures the fine surface geometric features and intrinsic colors of microstructures, which are essential for microscale research and applications.

3. Method

In this section, we introduce the characteristics of WLI data in Sec. 3.1. Here, we design the first pipeline for textured surface reconstruction of microscale samples using WLI and OM multi-modal data, as shown in Fig. 2. Specifically, given multi-view WLI and OM data of a microscale sample

as input, we obtain the camera poses for these two sets of multi-modal data within an absolute-scale coordinate system using a two-step data association process (Sec. 3.2). In Sec. 3.3, we use both conventional reconstruction and texture mapping workflows (Sec. 3.3.1), as well as neural implicit surface reconstruction workflows (Sec. 3.3.2), to fuse the multi-modal data into a textured 3D model of the sample. Finally, we present the implementation details of the neural implicit surface reconstruction in Sec. 3.4.

3.1. Preliminary: The Characteristics of WLI Data

The WLI utilizes interference from a broad-spectrum light source to achieve high-precision 3D measurements of microscale sample surfaces. The output of a commercial WLI instrument is a height map with an accurate scale, where each pixel's measurement precision can reach subnanometer levels. With the known XY scanning range corresponding to the objective lens used in WLI, the absolute coordinates of each pixel can be calculated. Thus, each WLI image can be represented as a point cloud or, as discussed in subsequent sections, a depth map from an orthographic camera. Although WLI offers high measurement accuracy, it encounters limitations when measuring smooth micro surfaces with steep angles, resulting in incomplete data with significant voids (more details in Sec. 4.4). Additionally, WLI instruments are expensive and relatively slow when scanning areas with large height variations. Consequently, in real experiments, it is challenging to obtain a substantial number of WLI scans from multiple viewpoints for a single sample. Therefore, the multi-view WLI data used in our experiments can be described as a set of high-precision depth maps with limited views and containing data voids.

3.2. Multi-Modal Data Association

First, we need to obtain the poses of virtual orthographic cameras corresponding to the multi-view WLI images, as well as the camera poses of the multi-view OM images. Most multi-modal 3D reconstruction methods involve multiple sensors with fixed relative rigid-body transformations and synchronized, aligned data. For example, an RGBD camera provides aligned color and depth images, or different sensors fixed on the same platform with unchanging relative poses. However, in our approach, the acquisitions of WLI and OM images are completely independent. Moreover, WLI and OM images have different resolutions, different numbers of inputs, and different imaging models, making it challenging to directly establish data associations between the raw WLI and OM images. To address this issue, we propose a two-step data association method.

In the first step, we separately calculate the relative camera poses among multiple WLI images and among multiple OM images. Multi-view WLI images are converted into 3D point clouds, and then the Iterative Closest Point (ICP) al-
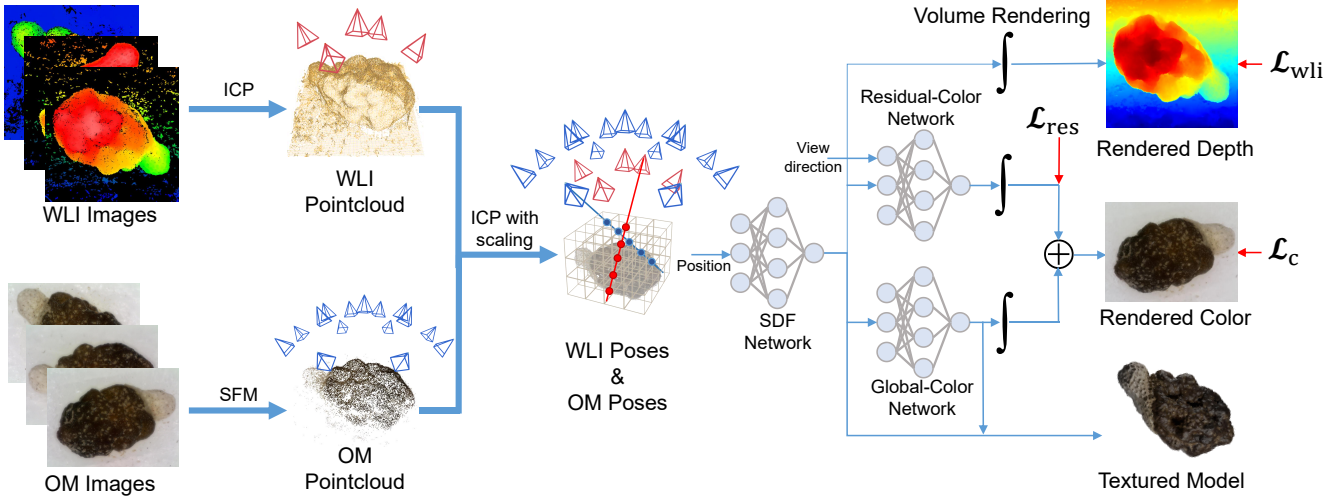
Figure 2. **System pipeline of OpticFusion.** Our method takes multi-view WLI and OM images of a microscale sample as input. Through a two-step data association process, we compute the pose of each WLI and OM image in the same absolute-scale coordinate system. Using differentiable volume rendering techniques, we train the neural implicit representation of the microstructure's geometry and view-independent color under the supervision of the multi-modal data. The result is a textured 3D surface model of the microscale sample.

gorithm [39] is used to align these point clouds within a unified coordinate system. This process calculates the relative poses of the virtual orthographic cameras of the WLI images and produces a merged WLI point cloud, denoted as P_{wli}. For multi-view OM images, we use traditional structure from motion (SFM) [1] to compute the camera poses and generate a sparse point cloud of the microstructure. We then filter out low-confidence points, retaining only high-confidence spatial points, which results in the OM point cloud, denoted as P_{om}. In the second step, we calculate the relative poses between the WLI and OM images by aligning two intermediate point clouds. Since WLI data contains absolute-scale information while the OM reconstruction result has scale ambiguity, we employ the ICP algorithm with scale transformation to align P_{om} to the P_{wli} coordinate system. Finally, this alignment allows us to obtain the camera poses for each WLI and OM image within the same absolute-scale coordinate system.

3.3. Multi-Modal 3D Reconstruction

3.3.1 Conventional Solution

A conventional method for fusing multi-modal data for 3D reconstruction involves two sequential steps: surface reconstruction and texture mapping. Here, we utilized this conventional approach to fuse WLI and OM data, reconstructing textured 3D models of microscale samples. First, we used the Poisson surface reconstruction method [17] to convert the fused point cloud obtained from the previous steps into a surface mesh of the microstructure. Then, we applied the MVS-Texturing [47] method to map the OM images onto the mesh, resulting in a textured 3D model.

3.3.2 Neural Implicit Multi-Modal 3D Reconstruction with View-Independent Color

Our multi-modal neural implicit representation consists of three networks: an SDF network, a global color network, and a residual color network. The SDF network converts input coordinate $\mathbf{x}$ into the SDF value at that spatial location and a geometry feature vector $\mathbf{f}$. We follow the approaches of NeuS [48] in the design of this part. In the original NeuS, $\mathbf{f}$ is immediately input into a color network along with the view direction $\mathbf{v}$ and normal $\mathbf{n}$ to map the spatial point to its color value in the specified view direction. Importantly, unlike rendering applications for macroscale objects, research in the microscale domain does not require the synthesis of novel views. Instead, it focuses on the geometry of the sample and its intrinsic colors. Therefore, we do not want the reconstructed surface color of the 3D model including surface highlights, due to lighting or changes in viewing direction. Inspired by the color decomposition strategies in Intrinsic-NeRF [51] and Color-NeuS [56], we modified NeuS to decouple the view-dependent color components and eliminate the specular highlights from the object surface in the OM images. The original color network is further decomposed into a view-dependent residual color network $\mathcal{R}$ and a view-independent global color network $\mathcal{G}$.

$$c_{\text{res}} = \mathcal{R}(\mathbf{x}, \mathbf{n}, \mathbf{f}, \mathbf{v}), \tag{1}$$

$$c_{\text{global}} = \mathcal{G}(\mathbf{x}, \mathbf{n}, \mathbf{f}). \tag{2}$$

The outputs of the two networks are added together to get the complete color $c = c_{\text{res}} + c_{\text{global}}$. We follow the volume rendering of NeuS [48] to accumulate the output into a rendered color C. When we generate the mesh with the

marching cube algorithm, only the global color network is used to infer the vertex colors of the mesh to get the final textured 3D model of the microscale sample.

3.4. Optimization

To optimize the implicit representation networks for the geometry and color of microstructures, we minimize a series of losses, including color loss, residual loss, WLI loss, and SDF regularization loss, supervised by the posed WLI and OM images. We randomly sample pixels on both WLI and OM images and generate sampling rays based on the corresponding camera projection model and pose. In our experiments, WLI and OM image pixels have the same probability of being sampled. We denoted m as the batch size and n as the number of sample points. First, the color loss is defined as the difference between the rendered color C along the ray and the pixel color $\widehat{C}$ in the OM image:

$$\mathcal{L}_{\mathrm{c}} = \frac{1}{m} \sum_i \left\| \widehat{C}_i - C_i \right\|_2^2. \tag{3}$$

Second, the residual loss aims to minimize the output of the residual color network:

$$\mathcal{L}_{\mathrm{res}} = \frac{1}{m} \sum_i \| C_{\mathrm{res}_i} \|_2^2. \tag{4}$$

This guides the view-independent global color network in capturing the primary color information of the OM image. In contrast, the residual network only represents other view-dependent components, such as specular highlights on the object's surface.

The WLI loss is the difference between the depth D along the ray direction and the WLI depth $\widehat{D}$:

$$\mathcal{L}_{\mathrm{wli}} = \frac{1}{m} \sum_i \left\| \widehat{D}_i - D_i \right\|_2^2. \tag{5}$$

Finally, a commonly used regularization term $\mathcal{L}_{\mathrm{reg}}$ [11], is applied to constrain the network's SDF output:

$$\mathcal{L}_{\mathrm{reg}} = \frac{1}{mn} \sum_{i,j} \left(\| \mathbf{n}_{i,j} \| - 1 \right)^2. \tag{6}$$

However, unlike RGBD images where each pixel has both color and depth information, each sampled point on WLI and OM images only has depth or color information, respectively. Therefore, when sampling a pixel on an OM image, the loss function is $\mathcal{L} = \lambda_{\mathrm{c}} \mathcal{L}_{\mathrm{c}} + \lambda_{\mathrm{res}} \mathcal{L}_{\mathrm{res}} + \lambda_{\mathrm{reg}} \mathcal{L}_{\mathrm{reg}}$. When sampling a pixel on a WLI image, the loss function is $\mathcal{L} = \lambda_{\mathrm{wli}} \mathcal{L}_{\mathrm{wli}} + \lambda_{\mathrm{reg}} \mathcal{L}_{\mathrm{reg}}$.

4. Experiments

4.1. Experimental Setup

In our experiments, we executed the OpticFusion code on a workstation computer equipped with an AMD 5950X CPU

and an Nvidia RTX4090 GPU. Our multi-modal neural implicit surface reconstruction is built upon the existing PyTorch [36] implementation [13] of hash encoding [30] and NeuS [48]. The loss weights for our experiments are set as follows: $\lambda_{\mathrm{c}} = 1.0$, $\lambda_{\mathrm{res}} = 0.001$, $\lambda_{\mathrm{wli}} = 1.0$, and $\lambda_{\mathrm{reg}} = 0.1$. We used the Adam optimizer [19] with a learning rate of 0.01 to train the network, conducting 20,000 iterations. Each iteration involved randomly selecting 256 rays and sampling 1024 points per ray. Please refer to the supplementary material for more implementation details.

4.2. Evaluation on Real World Dataset

WLI-OM Dataset. We first constructed a dataset comprising multi-view OM images and WLI scans of real-world microscale samples. This dataset includes five sequences collected from typical microstructures studied in microscale research, such as flower seeds, a butterfly wing, a circuit board, and an electronic component. The surface geometric features of these samples range in size from tens to hundreds of micrometers. The OM images were captured by a commercial digital microscope, Dino-Lite AM7915MZT. Each sample was imaged by OM at tilt angles of 15°, 30°, and 45°, with a photo taken every 30° around the z-axis, plus one top-down view. This process resulted in 37 OM images per sample, each with a resolution of 1296×972. For the multi-view WLI data, we used a commercial WLI instrument, ZYGO NewView 8200. Since the WLI can only move along the z-axis and cannot tilt, we placed the samples on a 30° tilt turntable and performed scans every 90°. Together with a scan without tilt, we obtained 5 WLI images per sample, each with a resolution of 1024×1024.

Results Analysis. We applied three methods to the WLI-OM dataset: the conventional solution described previously, the NeuS method, and our OpticFusion method. The results are shown in Fig. 3. While WLI offers very high measurement accuracy, it often results in significant voids in the data, with missing measurements on some parts of the sample surface. Consequently, Poisson reconstruction in the conventional solution causes unsatisfactory geometry, as seen in the apparent surface defects in Fig. 3g and 3i and the discontinuities on the butterfly's tiny surface burr in Fig. 3j. Additionally, the conventional texture mapping method can sometimes fail, resulting in inconsistencies with the OM images. For NeuS, we used only OM images as input. However, due to the limited number and angles of the OM images, the reconstructed geometry is poor in some texture-less areas of the microstructure surface. Furthermore, because the optical microscope cannot capture very subtle shape variations on the sample surface, some surface details are lost in the reconstructed results, such as the white areas in Fig. 3k and the depressions on the surface in Fig. 3l. In contrast, our OpticFusion method performs better in fusing multi-view WLI and OM data into a textured 3D model.
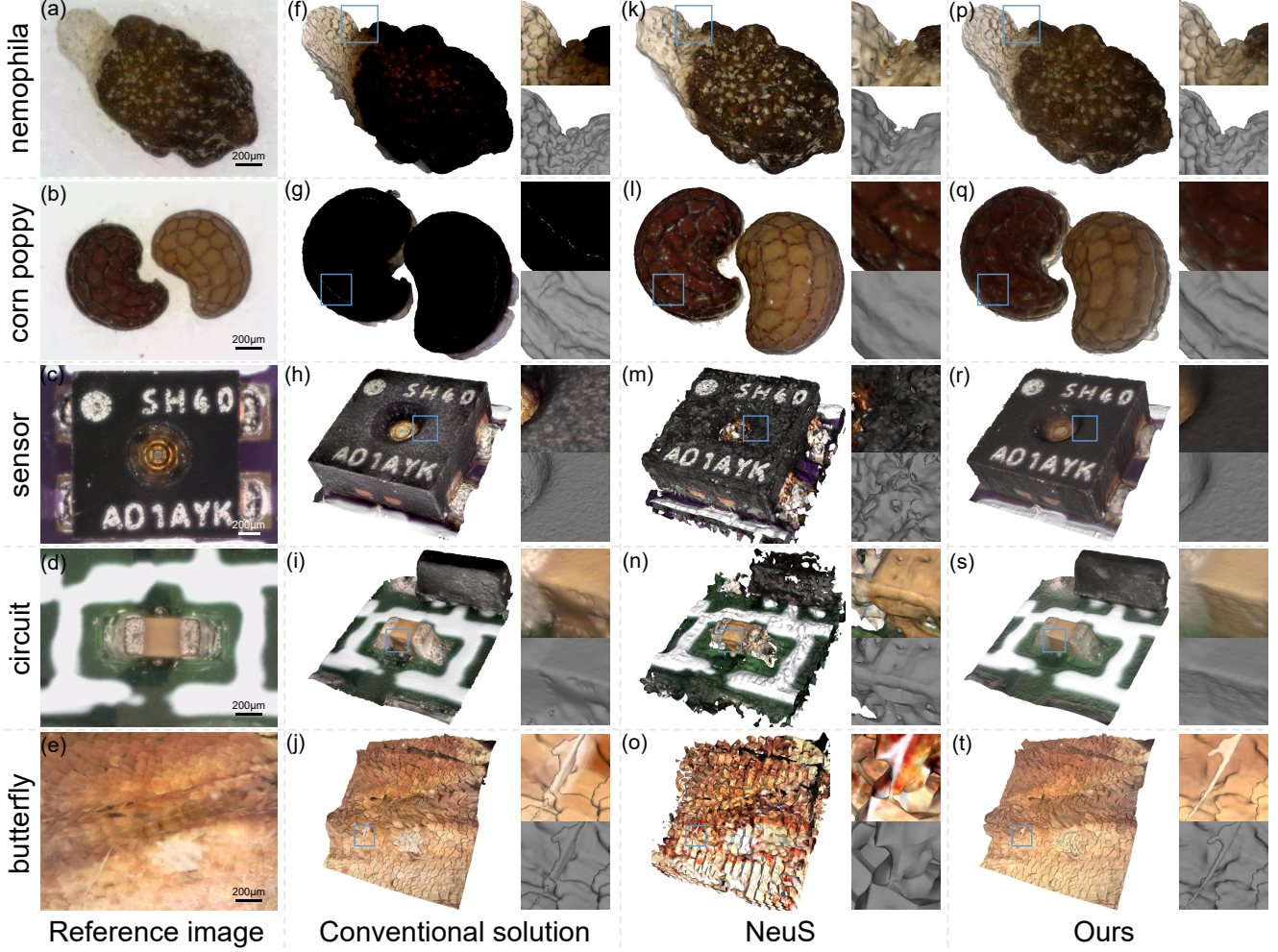
Figure 3. **Qualitative reconstruction results on our WLI-OM real-world dataset.** Compared with other methods, our reconstruction results have more accurate surface geometry and natural texture colors. Please see more visualizations in the supplementary material.

The multi-view OM images help fill in the areas where WLI data is missing, such as the surfaces in Fig. 3q and 3s and the complete tiny burr on the butterfly wing in Fig. 3t. The WLI data provides accurate constraints for surface reconstruction, preventing the surface collapse in texture-less regions, as observed in NeuS. The few input OM images are enough to texture the surface with natural colors. These results demonstrate that our OpticFusion method effectively combines the advantages of WLI and OM data, achieving high-fidelity geometry and texture reconstruction even in challenging microscopic scenarios, ensuring robustness and detail retention across varying surface features.

4.3. Application

Our method has broad applications in microscale research, as demonstrated by the following examples. Surface roughness analysis of microstructures is a common standard for evaluating the precision of micro-nanoscale processing [22, 25, 55]. By obtaining a complete 3D model of the microscale sample with our method, we can easily analyze the surface roughness of any sample region. Here, we use two roughness parameters, the arithmetical mean height of the surface (Sa) and the root mean square height of the surface (Sq), with calculation details provided in the supplementary material. The roughness analysis results on different surfaces of a microsensor are shown in Fig. 4.

Our reconstruction results are also highly beneficial for analyzing biological surface details [40]. Cross-sectional images of the butterfly wing in different directions clearly show the distribution, spacing, and size of surface scales, as well as the thickness and height of wing veins (Fig. 5). Researchers can even select and analyze specific structures on the biological surface, such as an individual scale or a tiny burr. The reconstructed complete 3D models with natural colors provide a more intuitive and detailed perspective for researchers in the microscopic field, facilitating further analysis and understanding of microstructures.
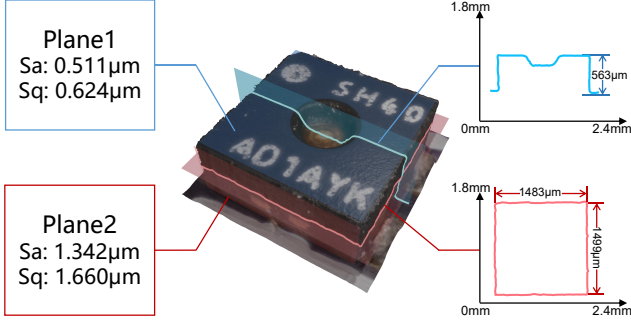
Figure 4. **Roughness measurement of microscale surfaces.** We calculate the roughness of two surfaces and obtain cross-sections in two directions.
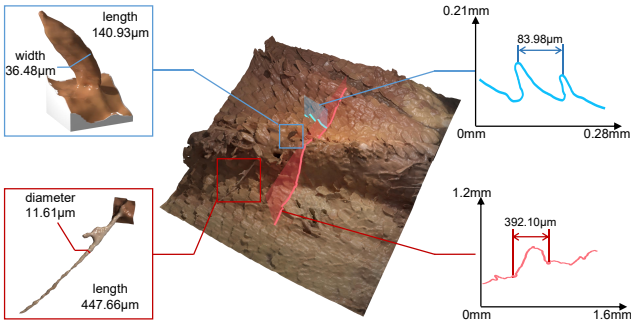

Figure 5. **Surface detail analysis of a biological sample.** We isolate separate 3D models of a single scale and a tiny burr on the butterfly's wing. Additionally, we provide cross-sections of multiple continuous scales and veins.

4.4. Evaluation on Synthetic Dataset

Construction of Synthetic Dataset. In real experimental settings, obtaining the ground truth 3D model of microscale samples is challenging, as WLI is among the most precise instruments for 3D measurements in microstructures. To quantitatively evaluate our method, we constructed a set of synthetic multimodal data based on the configurations of our WLI-OM dataset. We utilized 3D models from the NeRF Synthetic dataset [29] in our synthetic dataset. We rendered RGB images with known poses in Blender [6], matching the number, tilt angles, and resolution of OM images in the WLI-OM dataset. The main challenge is simulating multi-view WLI data. WLI data can be regarded as noise-free depth maps acquired by an orthographic camera since WLI's sub-nanometer measurement error is negligible compared to the geometric features of microscale samples. A key characteristic of WLI is that the numerical aperture (NA) of the objective lens limits the measurable surface tilt angles [45, 46], leading to voids in the scan data (Fig. 6). As shown in Fig. 7a, when the tilt angle α of a specular micro surface exceeds $\theta = arcsin(\mathrm{NA})$, the incident light cannot be reflected back to the WLI objective lens, resulting in no observable interference pattern on the CCD camera. Thus, many areas on the smooth substrate of the sample at

large tilt angles lack measurement values in Fig. 6. Conversely, rough micro surfaces can exceed the NA limitation because the diffuse and back-scattered light from the rough tilted surface can be captured by the objective lens (Fig. 7b). High dynamic range measurement techniques capture interference fringes, providing measurements on the rough seed surface in Fig. 6. In our synthetic data, we approximate this relationship between the proportions of the voids in WLI data and the surface tilt angle and reflectance by selectively removing depth values from the original depth maps (more details are included in the supplementary).
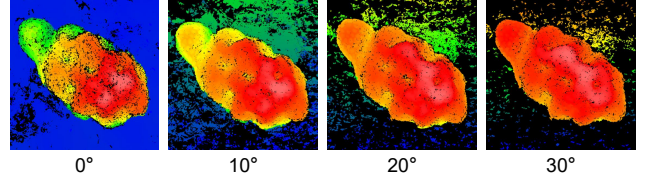

Figure 6. **WLI scanning results at different angles.** As the scanning angle increases, the proportion of valid measurement data from WLI on the specular surface (bottom plane) gradually decreases. In contrast, the rough scattering surface (seed surface) does not change significantly.
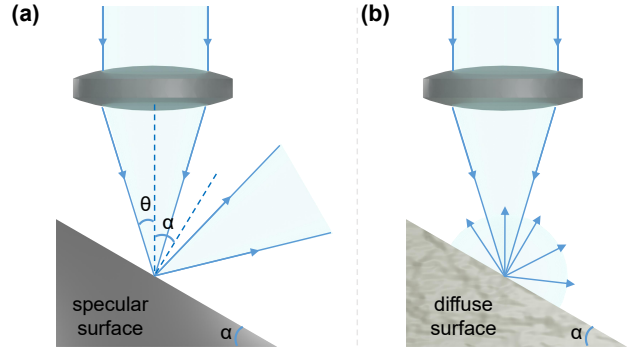

Figure 7. **WLI measurements of two different types of surfaces at tilt angles exceeding the numerical aperture (NA) limit. (a)** on specular surface. **(b)** on diffuse surface.

Results Analysis. Here, we use Chamfer Distance to evaluate the reconstruction quality of different methods: conventional solution (Poisson reconstruction of the WLI point cloud), training a neural implicit representation (NeuS) with either multi-view OM data or WLI data separately, using both WLI and OM for supervision without the residual term, and our complete reconstruction method, OpticFusion. The results are shown in Tab. 1. First, the Poisson reconstruction of the incomplete WLI point cloud yields relatively poor results. Second, The results demonstrate that multimodal reconstruction using both WLI and OM data significantly outperforms reconstruction using either data type alone, consistent with our observations in the real data experiment. These results indicate that fusing WLI and OM data through neural implicit representation is an effective

Chamfer distance ↓	lego	drums	hotdog	ficus	chair	mic	ship	average
Conventional solution	0.511	1.248	0.341	0.178	1.188	1.858	0.227	0.793
Ours w/o WLI supervision	0.160	0.034	1.067	0.021	0.193	0.026	2.265	0.538
Ours w/o OM supervision	0.148	0.491	**0.053**	0.049	0.614	0.534	2.855	0.678
Ours w/o residual term	0.038	0.033	0.065	**0.017**	0.071	0.016	**0.069**	0.044
Ours	**0.031**	**0.030**	0.071	0.018	**0.064**	**0.015**	0.070	**0.043**

Table 1. **Quantitative comparison of reconstruction results on the synthetic multi-view WLI and OM dataset.** We use chamfer distance as our evaluation metric. Our method performs better than other methods. Moreover, the residual item in our method has no noticeable impact on the reconstruction quality.

approach for reconstructing complex 3D structures at the microscopic level. Finally, for the residual term used in our OpticFusion, we intend for it to improve only the texture color of the final model without negatively impacting the geometric reconstruction. As expected, the residual term does not significantly affect the reconstruction error. Please refer to our supplementary material for more visualizations.

4.5. Ablation Study
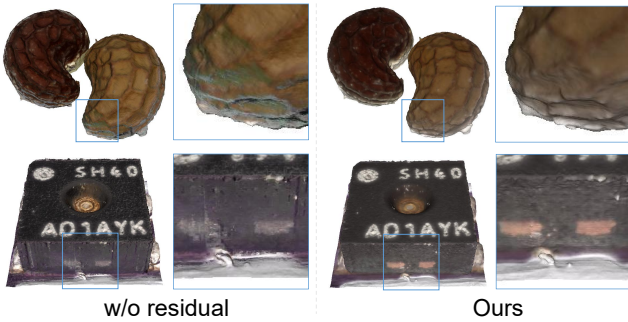

w/o residual   Ours

Figure 8. **Effect of the residual term.**

As shown in Fig. 8, we demonstrate the necessity and effectiveness of introducing a residual term during the generation of surface textures. A common method for calculating the color of vertices extracted in neural implicit surface representation is to use the direction opposite to the surface normal as the view direction and then infer the color using the color network. However, for microscale samples, the limited number and angles of optical microscope observations make it difficult to obtain comprehensive views of each point from various directions. Additionally, highlights at particular angles on the sample surface may obscure the original color of the sample. Therefore, in our real experimental data, we clearly observe that this common texture generation method is insufficient, leading to some strange colors appearing on our model. By introducing the residual term, the color of each vertex is obtained by inputting the coordinates of the point into the view-independent global color network, resulting in a more natural surface color.

We further explored the performance of dense reconstruction relying solely on OM images with a sufficient number of views. Here, we collected 109 OM images of one
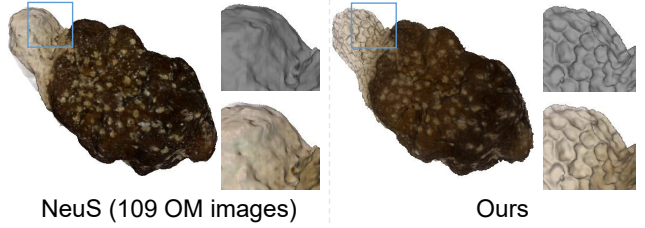

NeuS (109 OM images)   Ours

Figure 9. **Reconstruction result with more OM images.**

sample from the WLI-OM dataset, three times the number used in our previous experiments. As shown in Fig. 9, the OM images still fail to reconstruct certain detailed grooves on the sample's surface. This is because the geometric information in these areas is not represented as texture information in the OM images, and thus cannot be obtained through multi-view reconstruction. On the other hand, the ability of WLI to measure surface topography is independent of the sample's surface texture and mainly depends on whether the sample can reflect the incident white light back into the objective lens. This further demonstrates the necessity and effectiveness of fusing WLI and OM data.

5. Conclusion

We propose a novel method for neural implicit surface reconstruction of microstructures using WLI and OM multi-modal data to obtain textured 3D models of microscale samples. To get the camera poses of independently collected WLI and OM data, we employ a two-step data association method. We then use multi-modal data as supervision to optimize the neural implicit representation, effectively combining the advantages of WLI and OM data to achieve precise reconstruction of microscale surface geometry. We use a color decomposition strategy that introduces a view-dependent residual term into the implicit representation and extracts the view-independent color component from OM images, resulting in more natural texture colors. Our experiments demonstrate OpticFusion's capability to reconstruct high-quality textured 3D models of various microscale samples and its practical value in microscopic research.

References

[1] Agisoft. *Agisoft Metashape.* https://www.agisoft.com, 2024. 4

[2] Arkadiusz J Antończak, Dariusz Kocoń, Maciej Nowak, Paweł Kozioł, and Krzysztof M Abramski. Laser-induced colour marking—sensitivity scaling for a stainless steel. *Applied Surface Science*, 264:229–236, 2013. 1

[3] Dejan Azinović, Ricardo Martin-Brualla, Dan B Goldman, Matthias Nießner, and Justus Thies. Neural RGB-D surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6290–6301, 2022. 3

[4] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479, 2022. 3

[5] JL Beverage, X Colonna De Lega, and MF Fay. Interferometric microscope with true color imaging. In *Interferometry XVII: Techniques and Analysis*, pages 216–225. SPIE, 2014. 1

[6] Blender. *Blender.* https://www.blender.org, 2024. 7

[7] Shuo Chen, Mao Peng, Yijin Li, Bing-Feng Ju, Hujun Bao, Yuan-Liu Chen, and Guofeng Zhang. Multi-view neural 3D reconstruction of micro-and nanostructures with atomic force microscopy. *Communications Engineering*, 3(1):131, 2024. 3

[8] Mark Davidson, Kalman Kaufman, Isaac Mazor, and Felix Cohen. An application of interference microscopy to integrated circuit inspection and metrology. In *Integrated Circuit Metrology, Inspection, & Process Control*, pages 233–249. SPIE, 1987. 1

[9] Peter De Groot. Coherence scanning interferometry. In *Optical measurement of surface topography*, pages 187–208. Springer, 2011. 1

[10] Yang Dong, Shu-Qian Fan, Yu Shen, Ji-Xiang Yang, Peng Yan, You-Peng Chen, Jing Li, Jin-Song Guo, Xuan-Ming Duan, Fang Fang, et al. A novel bio-carrier fabricated using 3D printing technique for wastewater treatment. *Scientific reports*, 5(1):12400, 2015. 1

[11] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *Proceedings of the 37th International Conference on Machine Learning*, pages 3789–3799, 2020. 5

[12] Tong Guo, Feng Li, Jinping Chen, Xing Fu, and Xiaotang Hu. Multi-wavelength phase-shifting interferometry for micro-structures measurement based on color image processing in white light interference. *Optics and Lasers in Engineering*, 82:41–47, 2016. 1

[13] Yuan-Chen Guo. Instant neural surface reconstruction. *Github.* https://github.com/bennyguo/instant-nsr-pl, 2022. 5

[14] Youngjib Ham and Mani Golparvar-Fard. An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery. *Advanced Engineering Informatics*, 27(3):395–409, 2013. 3

[15] Akira Hirabayashi, Hidemitsu Ogawa, and Katsuichi Kitagawa. Fast surface profiler by white-light interferometry by use of a new algorithm based on sampling theory. *Applied Optics*, 41(23):4876–4883, 2002. 1

[16] Wei Huang, Jianhua Chen, Yao Yao, Ding Zheng, Xudong Ji, Liang-Wen Feng, David Moore, Nicholas R Glavin, Miao Xie, Yao Chen, et al. Vertical organic electrochemical transistors for complementary circuits. *Nature*, 613(7944):496–502, 2023. 1

[17] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 4

[18] Jin-Yong Kim, Seungjae Kim, Min-Gyu Kim, and Heui Jae Pahk. Generating a true color image with data from scanning white-light interferometry by using a fourier transform. *Current Optics and Photonics*, 3(5):408–414, 2019. 1

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[20] U Paul Kumar, Wang Haifeng, N Krishna Mohan, and MP Kothiyal. White light interferometry for surface profiling with a colour CCD. *Optics and Lasers in Engineering*, 50 (8):1084–1088, 2012. 1

[21] Mathieu Labbé and François Michaud. RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *Journal of field robotics*, 36(2):416–446, 2019. 3

[22] Duo Li, Zhen Tong, Xiangqian Jiang, Liam Blunt, and Feng Gao. Calibration of an interferometric on-machine probing system on an ultra-precision turning machine. *Measurement*, 118:96–104, 2018. 1, 6

[23] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6498–6508, 2021. 3

[24] Xinyang Liu, Yijin Li, Yanbin Teng, Hujun Bao, Guofeng Zhang, Yinda Zhang, and Zhaopeng Cui. Multi-modal neural radiance field for monocular dense slam with a light-weight tof sensor. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1–11, 2023. 3

[25] Don A Lucca, Matthew J Klopfstein, and Oltmann Riemer. Ultra-precision machining: cutting with diamond tools. *Journal of Manufacturing Science and Engineering*, 142 (11):110817, 2020. 1, 6

[26] Suodong Ma, Chenggen Quan, Rihong Zhu, Cho Jui Tay, and Lei Chen. Surface profile measurement in white-light scanning interferometry using a three-chip color CCD. *Applied Optics*, 50(15):2246–2254, 2011. 1

[27] Sébastien Marbach, Rémy Claveau, Fangting Wang, Jesse Schiffler, Paul Montgomery, and Manuel Flury. Wide-field parallel mapping of local spectral and topographic information with white light interference microscopy. *Optics Letters*, 46(4):809–812, 2021. 1

[28] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF*

conference on computer vision and pattern recognition, pages 7210–7219, 2021. 3

[29] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 3, 7

[30] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 5, 1

[31] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE international symposium on mixed and augmented reality*, pages 127–136. IEEE, 2011. 3

[32] Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Marc Stamminger. Real-time 3D reconstruction at scale using voxel hashing. *ACM Transactions on Graphics (ToG)*, 32 (6):1–11, 2013. 3

[33] Mert Özer, Maximilian Weiherer, Martin Hundhausen, and Bernhard Egger. Exploring multi-modal neural scene representations with applications on thermal imaging. *arXiv preprint arXiv:2403.11865*, 2024. 3

[34] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 3

[35] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. 3

[36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. PyTorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 5

[37] Nora E Paul, Claudia Skazik, Marc Harwardt, Matthias Bartneck, Bernd Denecke, Doris Klee, Jochen Salber, and Gabriele Zwadlo-Klarwasser. Topographical control of human macrophages by a regularly microstructured polyvinylidene fluoride surface. *Biomaterials*, 29(30):4056–4064, 2008. 1

[38] M Roy, P Svahn, L Cherel, and CJR Sheppard. Geometric phase-shifting for low-coherence interference microscopy. *Optics and Lasers in Engineering*, 37(6):631–641, 2002. 1

[39] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the ICP algorithm. In *Proceedings third international conference on 3-D digital imaging and modeling*, pages 145–152. IEEE, 2001. 4

[40] Clemens F Schaber, Stanislav N Gorb, and Friedrich G Barth. Force transformation in spider strain sensors: white light interferometry. *Journal of the Royal Society Interface*, 9(71):1254–1264, 2012. 1, 6

[41] Liyue Shen, John Pauly, and Lei Xing. NeRP: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):770–782, 2022. 3

[42] Guang-Ling Song and ZhenQing Xu. The surface, microstructure and corrosion of magnesium alloy az31 sheet. *Electrochimica Acta*, 55(13):4148–4161, 2010. 1

[43] Patrick Stotko, Michael Weinmann, and Reinhard Klein. Albedo estimation for real-time 3D reconstruction using RGB-D and IR data. *ISPRS journal of photogrammetry and remote sensing*, 150:213–225, 2019. 3

[44] Ahmad P Tafti, Andrew B Kirkpatrick, Zahrasadat Alavi, Heather A Owen, and Zeyun Yu. Recent advances in 3D SEM surface reconstruction. *Micron*, 78:54–66, 2015. 3

[45] Matthew Thomas, Rong Su, Peter de Groot, and Richard Leach. Optical topography measurement of steeply-sloped surfaces beyond the specular numerical aperture limit. In *Optics and Photonics for Advanced Dimensional Metrology*, pages 29–36. SPIE, 2020. 7

[46] Matthew Thomas, Rong Su, Peter De Groot, Jeremy Coupland, and Richard Leach. Surface measuring coherence scanning interferometry beyond the specular reflection limit. *Optics Express*, 29(22):36121–36131, 2021. 7

[47] Michael Waechter, Nils Moehrle, and Michael Goesele. Let there be color! large-scale texturing of 3D reconstructions. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 836–850. Springer, 2014. 3, 4

[48] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 34:27171–27183, 2021. 3, 4, 5

[49] James C Wyant. White light interferometry. In *Holography: a tribute to yuri denisyuk and emmett leith*, pages 98–107. SPIE, 2002. 1

[50] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. 3

[51] Weicai Ye, Shuo Chen, Chong Bao, Hujun Bao, Marc Pollefeys, Zhaopeng Cui, and Guofeng Zhang. IntrinsicNeRF: Learning intrinsic neural radiance fields for editable novel view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 339–351, 2023. 2, 3, 4

[52] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNeRF: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4578–4587, 2021. 3

[53] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1960–1970, 2021. 3

[54] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 3

[55] Zhiyu Zhang, Jiwang Yan, and Tsunemoto Kuriyagawa. Manufacturing technologies toward extreme precision. *International Journal of Extreme Manufacturing*, 1(2):022001, 2019. 1, 6

[56] Licheng Zhong, Lixin Yang, Kailin Li, Haoyu Zhen, Mei Han, and Cewu Lu. Color-NeuS: Reconstructing neural implicit surfaces with color. In *2024 International Conference on 3D Vision (3DV)*, pages 631–640. IEEE, 2024. 2, 4

[57] Qian-Yi Zhou and Vladlen Koltun. Color map optimization for 3D reconstruction with consumer depth cameras. *ACM Transactions on Graphics (ToG)*, 33(4):1–10, 2014. 3

[58] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R Oswald, and Marc Pollefeys. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12786–12796, 2022. 3