3D Neural Light Quantum Fields for Photo-realistic 3D Reconstruction from 2D Images

Yongfeng Shan
School of Computer Science
Faculty of Engineering and IT
University of Technology Sydney
Sydney, Australia
yongfeng.shan@student.uts.edu.au

Christy Jie Liang
School of Computer Science
Faculty of Engineering and IT
University of Technology Sydney
Sydney, Australia
jie.liang@uts.edu.au

***Abstract*—Rapid advancements in 3D reconstruction have led to numerous innovative technologies that greatly enhance the synthesis of novel views. However, most of these methods remain limited to static scenes and rarely address the challenges posed by dynamic environments. Additionally, many current approaches rely on capturing 360-degree views of individual objects, which is often impractical, as obtaining multiple viewpoints of a scene simultaneously is seldom feasible in real-world scenarios. In fields such as autonomous driving, robotic vision, and drone operations, environments are rarely static, while dynamic interactions are the norm. Meanwhile, enabling models to capture and interpret object movements accurately is essential for achieving effective 3D reconstruction of the natural world. To overcome these challenges, we introduce 3D Neural Light Quantum Fields (3D-LQF), a novel framework that advances static and dynamic 3D reconstruction. 3D-LQF leverages light's probabilistic properties to enhance visual realism and reconstruction precision by integrating time-sequential modeling with quantum photon probability fields. Experimental results demonstrate that 3D-LQF surpasses the performance of existing state-of-the-art methods in static scene reconstruction and also excels in handling dynamic scenes with real-time requirements. These findings position 3D-LQF as a promising solution for real-time 3D reconstruction in dynamic environments and diverse real-world scenarios.**

***Index Terms*—3D Reconstruction, Deep Learning, Neural Networks, 3D Light Quantum Fields, NeRF, 3DGS, Dynamic Scenes, Artificial Intelligence(AI), Feynman Path Integral, Quantum mechanics, Robot Vision, Embodied Intelligence**

I. Introduction

Real-time photorealistic 3D reconstruction is an inevitable and inescapable milestone to realize machine vision and advance artificial intelligence. However, producing accurate and photorealistic 3D models in real time remains a non-trivial task [1]. Traditional methods struggle with high computational demands, restricting their ability to generate detailed models within practical timeframes. Furthermore, these approaches often fail to capture fine-grained textures, leading to visually unrealistic outputs and artifacts. Most models also cannot learn and predict dynamic object motions over time (t), further limiting their capacity to reconstruct evolving scenes. These challenges hinder seamless adaptation to real-world applications, such as autonomous vehicles, robotics, and surveillance, where continuous motion is inherent.

Existing methods such as neural radiation fields (NeRF) [2] and 3DGS[3], along with their variants [4–7], have emerged as powerful tools for 3D reconstruction, offering notable advantages in speed and visual fidelity [3, 7]. Specifically, 3DGS approximates geometric structures and radiance information using collections of 3D Gaussian kernels, leveraging rasterization techniques and OpenGL-based graphics pipelines. This approach significantly reduces rendering time and ensures accurate alignment without exhaustive neural network computations by representing complex scenes with sparse points optimized for anisotropic covariances. However, as a point cloud-based framework, 3DGS lacks an explicit surface representation [8, 9], making it challenging to extract detailed surface features accurately. This limitation can lead to subtle distortions or discrepancies between the reconstructed scene and the original objects, especially when dealing with fine-grained textures or occluded areas. Additionally, 3DGS focuses on static scene rendering and relies on multiple images captured from various viewpoints, which may not always be practical. For instance, in autonomous driving scenarios, acquiring 360-degree coverage is often infeasible because the scene is predominantly limited to the forward-facing camera's field of view, possibly supplemented by top-down satellite imagery. Moreover, since 3DGS does not inherently model temporal changes or motion trajectories, it struggles to adapt to the dynamic nature of real-world applications such as autonomous driving and drone-based operations.

In this work, we propose **3D Light Quantum Fields (3D-LQF)** as a novel and efficient approach to overcome the limitations of traditional 3D reconstruction techniques. Our method uses the Signed Distance Function (SDF) as an implicit geometric representation during intermediate steps, enabling precise modeling of object surfaces. In the final stage, the SDF is converted into 3D Meshes, facilitating rasterization-based rendering, which ensures high visual fidelity and computational efficiency. This end-to-end rendering strategy balances precision with performance, allowing the seamless transition from implicit to explicit geometry.

For surface curvature estimation, Gaussian curvature and Euler curvature are combined. These curvatures guide the model's geometric approach: Riemannian geometry accurately

models convex surfaces with positive curvature, ensuring external surface precision, while Lobachevskian (hyperbolic) geometry addresses concave surfaces with negative curvature, enhancing internal surface rendering and light tracing. Euclidean geometry is applied to absolutely flat regions, maintaining accuracy in more straightforward scenarios.

Our pipeline also introduces a Bird's Eye View (BEV) [10, 11] perspective as a blueprint for structured 3D reconstruction. This perspective enhances global spatial understanding and is critical for autonomous driving and drone navigation scenarios. The dynamics of light propagation are simulated using the Schrödinger time equation, Feynman path integrals [12], and quantum wave packets [13, 14], capturing complex photon behavior. We integrate the Temporal Fusion Transformer (TFT) [15, 16] to effectively handle temporal sequences, which ensures robust time-sequential modeling. Furthermore, our model introduces photon angular momentum equations, which account for the rotational characteristics of light, ensuring accurate simulations of optical interactions.

II. Related Work

The concept of Light Fields was first introduced and explored in 1996 in two influential papers: "Light Field Rendering" by Marc Levoy and Pat Hanrahan [17], and "The Lumigraph" by Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen [18]. These works introduced a systematic approach to capturing light information within a scene, enabling novel view synthesis, and laid a foundation for advancements in 3D reconstruction, including more recent developments like NeRF and 3DGS. However, the evolution of 3D reconstruction can be broadly categorized into several stages and approaches, ranging from early points cloud-based techniques to modern neural rendering and 3DGS models.

A. 3D Points Clouds-Based Methods

This research category includes: [19–21], etc. Points clouds offer a more flexible and scalable alternative for 3D scene representation, consisting of individual points in space rather than structured surfaces. Approaches like "PointNet" and "PointNet++" advanced deep learning on point clouds, enabling more robust processing and reconstruction. Point clouds often suffer from sparsity and noise despite their flexibility, resulting in incomplete reconstructions. Recent methods such as "Point Transformer" and "Neural Point-Based Graphics (NPBG)" have attempted to mitigate these issues by improving the robustness of point cloud representations in real-world environments.

B. 3D Voxel-Based Methods

This category of research, such as [21–23], etc. 3D Voxel-based representations became popular with the rise of deep learning, mainly due to their ability to handle more complex topologies than mesh-based methods. Voxel grids divide the scene into volumetric cells, enabling techniques such as 3D CNNs and Volumetric Fusion to process the 3D structure effectively. However, voxel-based methods are notorious for their memory inefficiency, especially when dealing with high-resolution reconstructions. Even with optimizations like Octree-based models, memory usage remains a bottleneck in practical applications.

C. 3D Mesh-Based Methods

This research category includes: [8, 9, 24], etc. Early methods in 3D reconstruction primarily focused on "explicit 3D meshes," representing surfaces with vertices, edges, and faces. These meshes are optimized and refined using differentiable rendering techniques such as Soft Rasterizer. While mesh-based methods provide detailed surface reconstruction, they often face challenges in handling dynamic topologies, occlusions, and scalability in large-scale scenes. The computational complexity of managing explicit geometry in highly dynamic environments remains a crucial limitation.

D. Neural Radiance Fields, Instant-NGP, and Plenoxels

Neural Radiance Fields (NeRF) and its variants [4–6], introduced in 2020, represented a significant breakthrough in 3D reconstruction by using neural networks to represent scenes as volumetric functions implicitly. NeRF maps 3D spatial coordinates and viewing directions to color and density, synthesizing novel views from sparse 2D images. Despite its success, NeRF's computational cost and long training times limit its scalability. Recent advancements such as Instant-NGP [25] and Plenoxels [26] aim to improve efficiency using hash encoding and sparse voxel grids, allowing for real-time performance without sacrificing reconstruction quality.

E. Recent Advances: Neuralangelo and 3DGS

Models like Neuralangelo have driven recent research in 3D reconstruction [27] and 3DGS [3], each pushing the boundaries of surface reconstruction, hash-encoding and radiance field rendering. Neuralangelo, an enhancement of Instant-NGP, focuses on surface reconstruction, which uses coarse-to-fine optimization and numerical gradients instead of analytical gradients to compute higher-order derivatives. This enables the recovery of delicate structures in large-scale real-world scenes from RGB video captures without depth inputs. In contrast, 3DGS advances real-time novel-view synthesis by representing scenes with 3D Gaussians derived from sparse points during camera calibration. This method optimizes the Gaussians' anisotropic covariance, ensuring accurate scene representation while avoiding unnecessary computations in space. Additionally, 3DGS incorporates a fast, visibility-aware rendering algorithm that enables real-time rasterized OpenGL [3, 28], striking a balance between speed and visual quality. These innovations allow 3DGS to achieve state-of-the-art visual fidelity while maintaining competitive training times and real-time rendering at 30+ fps, making it a leading solution for high-quality 3D scene reconstructions.

III. Preliminaries and Methodology

This section introduces the core concepts and components of the 3D Light Quantum Fields (3D-LQF). Inspired by NeRF,

3DGS, and quantum mechanical principles governing photon dynamics, our approach combines the strengths of Signed Distance Functions (SDF) implicit representation for ray marching [29] and 3D Mesh representations' ability for rasterizing and OpenGL rendering. The pipeline employs SDF for implicit geometry representation and ray marching, enabling accurate simulation of light propagation through complex surfaces. However, the implicit SDF representation is converted into 3D Meshes for final rasterization, facilitating efficient rendering through traditional computer graphics techniques such as OpenGL-based rasterization.

A. Gaussian Curvature

Gaussian Curvature (K) [30] measures the intrinsic curvature at a point on the surface, defined as the product of the principal curvatures k_1 and k_2:

$$K = k_1 \cdot k_2 \tag{1}$$

- $K > 0$: Positive curvature (e.g., sphere), indicating a convex surface.
- $K = 0$: Zero curvature (e.g., plane or cylinder), indicating a locally flat surface.
- $K < 0$: Negative curvature (e.g., saddle surface), indicating a concave region.

B. Riemannian and Hyperbolic Geometry

For accurate surface rendering, **3D-LQF** leverages non-Euclidean geometry to model surfaces with varying curvature. **Riemannian geometry** [31–33] describes surfaces with positive curvature, such as convex shapes (e.g., spheres), ensuring optical fidelity in light interactions by correctly modeling photon paths on curved surfaces. The area element dA in Riemannian geometry is defined as:

$$dA = \sqrt{\det(g_{ij})}\, dx\, dy \tag{2}$$

where g_{ij} represents the components of the **metric tensor**, encoding the surface's intrinsic curvature. This allows precise handling of curved surfaces, which is essential for rendering smooth, convex regions.

On the other hand, **hyperbolic geometry** models concave surfaces with negative curvature, such as saddle-shaped surfaces. The **area** of a region D in hyperbolic space, where the curvature $K = -1$, is expressed as:

$$A(D) = \int_D \frac{dA}{\cosh^2(r)} \tag{3}$$

where r is the radial distance from the origin and $\cosh(r)$ is the hyperbolic cosine. For specific shapes, such as a hyperbolic triangle, the area depends only on its internal angles:

$$A = \pi - \sum \theta_i \tag{4}$$

These non-Euclidean formulations are crucial in our **3D-LQF** model, enabling accurate **ray marching** and **photon scattering** on surfaces with varying curvature.

Real-world surfaces contain microscopic undulations and irregularities. Advanced geometries are essential for these scenarios: **Riemannian geometry** accurately models light interactions on convex surfaces, while **hyperbolic geometry** handles concave regions. The 3D-LQF model dynamically selects between these geometries based on local surface curvature, enhancing the precision of light path calculations.

These non-Euclidean approaches improve **ray tracing** and **surface rendering**, ensuring that our model can handle smooth and complex textures with high fidelity, even in dynamic scenes.

C. Quantum Dynamics, Photon Wavefunction, Angular Momentum, and Path Optimization

Photon behavior in the 3D-LQF model incorporates multiple quantum principles to simulate light interactions with high fidelity. The model captures photon dynamics through wavefunction representations, angular momentum, and path integrals.

a) Wavefunction and 3D Photon Probability Density Field: The photon wavefunction $\psi(x)$ encodes the probabilistic nature of light in 3D space

$$\psi(x) = \frac{1}{(\sigma\pi)^{3/2}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}, \tag{5}$$

where x_0 is the center of the wave packet and σ controls its spatial spread. The corresponding probability density is the following.

$$P(x) = |\psi(x)|^2. \tag{6}$$

b) Photon Angular Momentum and Polarization: The model integrates **photon angular momentum** to account for rotational asymmetries and polarized light. Photon angular momentum $\hat{J}$ is expressed as

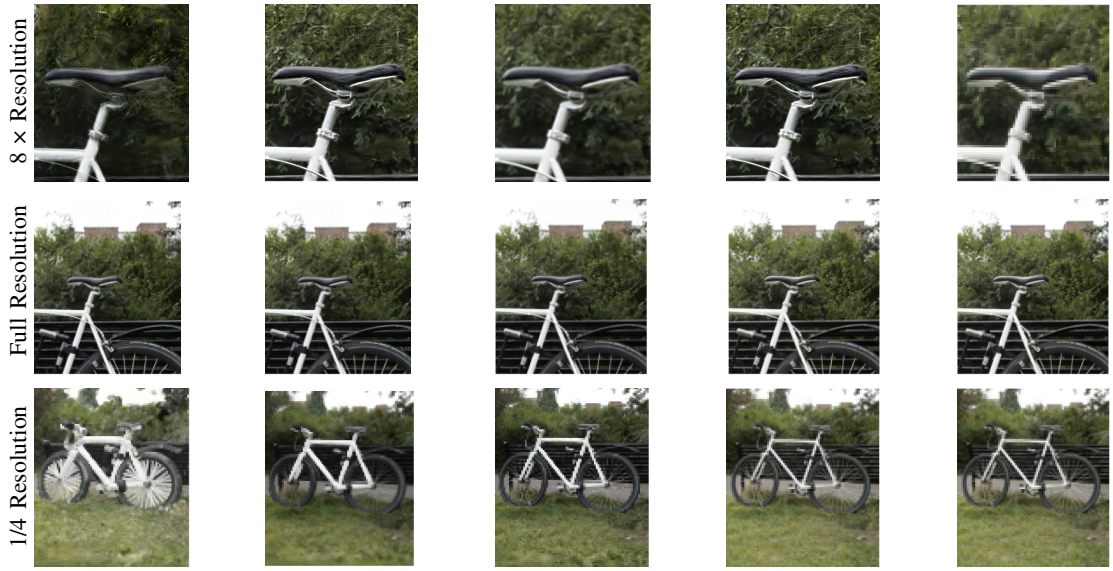
$$\hat{J} = \hat{L} + \hat{S}, \tag{7}$$

where $\hat{L}$ is the orbital angular momentum, and $\hat{S}$ is the spin angular momentum. These operators enable the model to accurately simulate polarization effects and photon spin, which is essential for light-field interactions.

c) Path Optimization: Fermat's Principle, Refractive Index, and Feynman's Path Integral: For computational efficiency, the model applies **Fermat's Principle**, which states that light follows the path of least time between two points. The model integrates the **Feynman Path Integral** to handle complex light interactions, summarizing all possible paths between two points. The probability amplitude for a photon's path is given by:

$$\langle x_f, t_f | x_i, t_i \rangle = \int D[x(t)]\, e^{\frac{i}{\hbar} S[x(t)]}, \tag{8}$$

where $S[x(t)]$ is the action along a given path, and $D[x(t)]$ sums over all possible paths. This approach accounts for quantum effects such as interference and scattering.



(a) 3DGS (b) 3DGS + EWA (c) Mip-Splatting (d) OURS (e) Reference

Fig. 1: Comparison of multi-resolution performance. Mip-splatting [7] enhances artifact reduction compared to previous methods [3, 34], yet it exhibits limitations in detail preservation under extreme downscaling. By contrast, our Quantum Photon Field model consistently maintains structural fidelity and achieves higher rendering accuracy across resolutions, establishing a new benchmark in scalable 3D reconstruction.

TABLE I: Multi-scale Training and Multi-scale Testing on the Blender dataset. Our 3D-LQF model achieves state-of-the-art performance across most metrics, outperforming Mip-Splatting and other methods.

	PSNR ↑					SSIM ↑					LPIPS ↓				
Method	**Full**	**1/2**	**1/4**	**1/8**	**Avg.**	**Full**	**1/2**	**1/4**	**1/8**	**Avg.**	**Full**	**1/2**	**1/4**	**1/8**	**Avg.**
NeRF w/o $\mathcal{L}_{\text{area}}$ [2, 37]	31.20	30.65	26.25	22.53	27.66	0.950	0.956	0.930	0.871	0.927	0.055	0.034	0.043	0.075	0.052
NeRF [2]	29.90	32.13	33.40	29.47	31.23	0.938	0.959	0.973	0.962	0.958	0.074	0.040	0.024	0.039	0.044
MipNeRF [37]	32.63	34.34	35.47	35.60	34.51	0.958	0.970	0.979	0.983	0.973	0.047	0.026	0.017	0.012	0.026
Plenoxels [26]	31.60	32.85	30.26	26.63	30.34	0.956	0.967	0.961	0.936	0.955	0.052	0.032	0.045	0.077	0.051
TensoRF	32.11	33.03	30.45	26.80	30.60	0.956	0.966	0.962	0.939	0.956	0.056	0.038	0.047	0.076	0.054
Instant-NGP [25]	30.00	32.15	33.31	29.35	31.20	0.939	0.961	0.974	0.963	0.959	0.079	0.043	0.026	0.040	0.047
Tri-MipRF [38]	32.65	34.24	35.02	35.53	34.36	0.958	0.971	0.980	0.987	0.974	0.047	0.027	0.018	0.012	0.026
3DGS	28.79	30.66	31.64	27.98	29.77	0.943	0.962	0.972	0.960	0.960	0.065	0.038	0.025	0.031	0.040
3DGS + EWA [34]	31.54	33.26	33.78	33.48	33.01	0.961	0.973	0.979	0.983	0.974	0.043	0.026	0.021	0.019	0.027
Mip-Splatting [7]	32.81	34.49	35.45	35.50	34.56	0.967	0.977	0.983	0.988	0.979	0.035	0.019	0.013	0.010	0.019
3D-LQF (Ours)	**33.00**	**34.70**	**35.70**	**35.80**	**34.80**	**0.970**	**0.980**	**0.985**	**0.990**	**0.981**	**0.030**	**0.017**	**0.012**	**0.009**	**0.017**

IV. Experiments

This section presents the experimental setup and evaluation of the proposed *3D Light Quantum Fields (3D-LQF)* model. We describe the datasets, metrics, and optimizations employed to achieve computational efficiency while maintaining the quantum mechanical foundations of the model. Our results demonstrate the capability of *3D-LQF* in static and dynamic scenarios, outperforming several state-of-the-art techniques.

A. Experiment Setup

1) Datasets: We conduct experiments on a variety of datasets covering static and dynamic scenes: Mip-NeRF 360 [5], Tanks & Temples [39], Deep Blending [40], D-NeRF [4], Nvidia Dynamic Scenes Dataset [36]

2) Evaluation Metrics: We evaluate model performance across a comprehensive set of metrics, including PSNR (Peak Signal-to-Noise Ratio) to measure fidelity to ground truth, SSIM (Structural Similarity Index) and MS-SSIM (Multi-Scale SSIM) to assess structural and multi-scale consistency, LPIPS (Learned Perceptual Image Patch Similarity) and FLIP (Foveated Loss Index for Perceptual) for perceptual quality, and Chamfer Distance to quantify geometric alignment with reference point clouds. These metrics provide a rigorous and multi-faceted visual and spatial reconstruction accuracy assessment. The training is conducted on two NVIDIA A40 GPUs, with 30,000 iterations per scene. We employ a multi-scale sampling strategy to handle varying image resolutions and zoom effects, similar to *Mip-NeRF* [37].

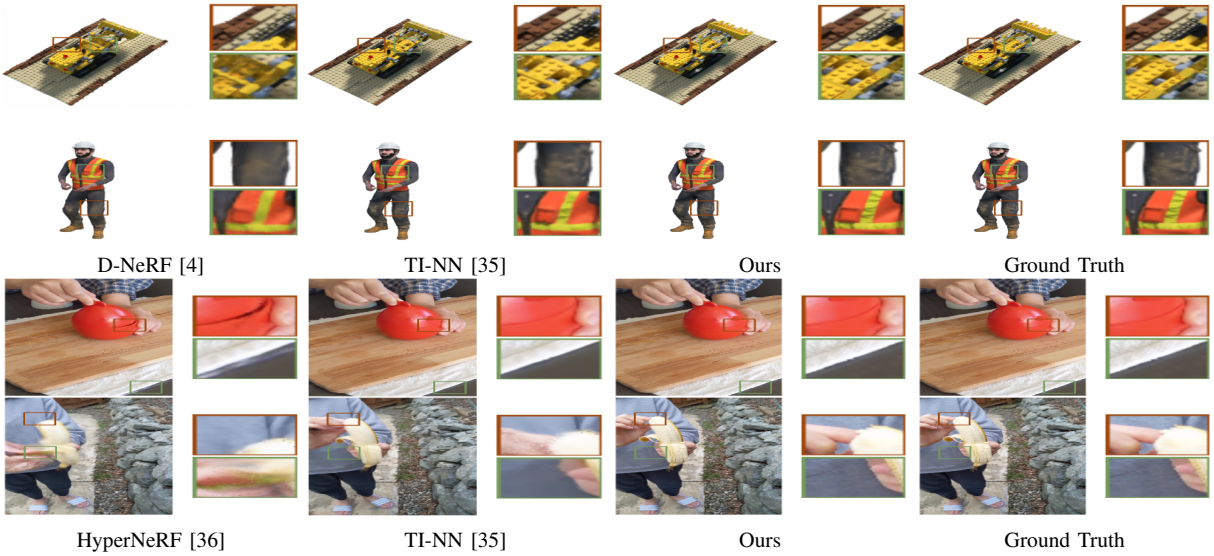


Fig. 2: Qualitative comparison of our Quantum Photon Field model (Ours) with D-NeRF [4], Temporal Interpolation-NN (TI-NN) [35], HyperNeRF [36], and Ground Truth on D-NeRF and HyperNeRF datasets. Red and green boxes highlight regions with notable differences.

B. Key Optimizations

Several key optimizations were implemented in the 3D-LQF model to maintain computational efficiency while incorporating quantum mechanical principles. Path sampling reduces the complexity of photon path calculations by focusing on the most likely paths according to Fermat's Principle, minimizing unnecessary computations. Precomputing quantum effects further optimizes the process by caching reusable photon behaviors, avoiding repeated calculations, and speeding up the simulations.

A hybrid neural-quantum approach was employed, leveraging neural networks to approximate quantum phenomena effectively, allowing the model to integrate data-driven insights with physical principles. The model employs hierarchical processing to handle scenes with varying spatial and temporal dynamics, assigning greater detail to critical areas while reducing resolution in distant or slow-moving regions. This ensures that computational resources are efficiently allocated to the most impactful areas.

GPU acceleration plays a crucial role by enabling the parallel execution of wavefunction calculations and matrix operations through CUDA, significantly enhancing real-time performance. In addition, the model incorporates mesh extraction techniques using the Marching Cubes algorithm [41], which converts implicit SDF representations into explicit 3D meshes. These meshes are subsequently rasterized using OpenGL, enabling end-to-end optimization with differentiable rendering. Differentiable rendering facilitates the direct propagation of gradients through the rendering pipeline, allowing the refinement of surface geometry and light interactions based on rendered images.

A hybrid quantum-classical rendering approach was adopted, combining quantum-inspired photon modeling with classical techniques to balance computational efficiency and accuracy. Sparse representations of the photon field further optimize the system by minimizing memory usage, focusing on areas with significant contributions to the final output.

By integrating these optimizations, the 3D-LQF model offers high-quality reconstruction with reduced overhead, making it suitable for real-time and large-scale applications. The combination of SDF-based mesh extraction, Marching Cubes, and differentiable rendering adds robustness to the model, ensuring precise rendering and adaptability across different environments.

C. Quantitative Results

Tables I and II compare the 3D-LQF model's performance across various datasets against state-of-the-art methods. Table I illustrates the performance of the Blender dataset, highlighting the effectiveness of multi-scale training and testing. Table II showcases the results on the Mip-NeRF 360 Dataset, where all methods are trained on the scale (1×) and evaluated at multiple scales. Table III presents the experimental results on the HyperNeRF dataset, which includes dynamic scenes with non-rigid deformations and complex motions. This dataset evaluates models in two scenarios: *vrig* (viewpoint extrapolation) and *interp* (temporal interpolation). In both scenarios, our 3D-LQF model achieves superior performance, with a PSNR of 27.00 and MS-SSIM of 0.930 in the *vrig* scenario and a PSNR of 30.00 and MS-SSIM of 0.950 in the *interp* scenario. These results demonstrate our model's robustness in handling dynamic motions and maintaining high-quality

TABLE II: Single-scale Training and Multi-scale Testing on the **Mip-NeRF 360 Dataset**. All methods are trained on the smallest scale (1×) and evaluated across four scales (1×, 2×, 4×, and 8×). Evaluations at higher sampling rates simulate zoom-in effects. While our method yields comparable results at the training resolution, it significantly surpasses all previous work at all other scales.

	PSNR ↑					SSIM ↑					LPIPS ↓				
Method	**1× Res.**	**2× Res.**	**4× Res.**	**8× Res.**	**Avg.**	**1× Res.**	**2× Res.**	**4× Res.**	**8× Res.**	**Avg.**	**1× Res.**	**2× Res.**	**4× Res.**	**8× Res.**	**Avg.**
Instant-NGP [25]	26.79	24.76	24.27	24.27	25.02	0.746	0.639	0.626	0.698	0.677	0.239	0.367	0.445	0.475	0.382
mip-NeRF 360 [5]	29.26	25.18	24.16	24.10	25.67	0.860	0.727	0.670	0.706	0.741	0.122	0.260	0.370	0.428	0.295
zip-NeR [42]	29.66	23.27	20.87	20.27	23.52	0.875	0.696	0.565	0.559	0.674	0.097	0.257	0.421	0.494	0.318
3DGS	29.19	23.50	20.71	19.59	23.25	0.880	0.740	0.619	0.619	0.715	0.107	0.243	0.394	0.476	0.305
3DGS + EWA	29.30	25.90	23.70	22.81	25.43	0.880	0.775	0.667	0.643	0.741	0.114	0.236	0.369	0.449	0.292
Mip-Splatting	29.39	27.39	26.47	26.22	27.37	0.884	0.808	0.754	0.765	0.803	0.108	0.205	0.305	0.392	0.252
3D-LQF (Ours)	**30.00**	**28.00**	**27.00**	**26.50**	**27.88**	**0.890**	**0.815**	**0.770**	**0.780**	**0.814**	**0.095**	**0.190**	**0.290**	**0.380**	**0.239**

TABLE III: Experimental results on HyperNeRF datasets.

	vrig		*interp*	
Method	**PSNR ↑**	**MS-SSIM ↑**	**PSNR ↑**	**MS-SSIM ↑**
NeRF	20.13	0.745	22.27	0.804
NV [43]	16.85	0.571	26.05	0.911
NSFF [44]	**26.33**	**0.916**	25.80	0.883
Nerfies [45]	22.23	0.803	28.47	0.939
HyperNeRF	22.38	0.814	**29.00**	**0.945**
TI-NN [35]	24.35	0.867	28.67	0.940
3D-LQF (Ours)	**27.00**	**0.930**	**30.00**	**0.950**

TABLE IV: Experimental results on DyNeRF datasets.

Method	**PSNR ↑**	**LPIPS ↓**	**FLIP ↓**
MVS	19.12	0.2599	0.2542
NeuralVolumes [43]	22.80	0.2951	0.2049
LLFF [46]	23.24	0.2346	0.1867
NeRF-T [47]	28.45	0.1000	0.1415
DyNeRF† [47]	28.50	0.0985	0.1455
DyNeRF [47]	29.58	**0.0832**	**0.1347**
TI-NN [35]	29.88	0.0960	0.1413
3D-LQF (Ours)	**30.50**	**0.0800**	**0.1300**

reconstructions during spatial and temporal extrapolations. Table IV provides the experimental results on the DyNeRF dataset, specifically designed to evaluate models on dynamic neural radiance field reconstruction in scenes with complex motions and occlusions. Our 3D-LQF model outperforms existing methods by achieving the highest PSNR of 30.50, the lowest LPIPS of 0.0800, and the lowest FLIP of 0.1300. These metrics indicate our model excels in preserving fine details, minimizing perceptual errors, and maintaining overall image fidelity in challenging dynamic environments. Our 3D-LQF model consistently outperforms other methods across all metrics and resolutions in both cases. During experimentation, several challenges emerged, which required iterative solutions to ensure model convergence and satisfactory performance. One of the most pressing issues was the difficulty in achieving stable convergence. In initial runs, the model exhibited oscillations in loss, indicating that it struggled to learn optimal photon paths and surface interactions effectively. This instability was addressed by tuning the learning rate, adopting an adaptive optimizer, and refining the model's architecture to enhance gradient flow.

Another challenge was the extended training time, primarily due to the computational complexity of modeling photon behavior using path integrals and angular momentum. Training time became incredibly prohibitive when handling dynamic datasets, where the model needed to capture object motion trajectories across time. To mitigate this, we integrated GPU acceleration with CUDA to parallelize critical operations, such as matrix multiplications and wavefunction calculations. Additionally, incorporating hierarchical sampling, focusing on essential scene regions, significantly reduced unnecessary computations and shortened the training cycle.

Extracting smooth 3D meshes from the implicit SDF representation also presented a challenge, as noise in the SDF field occasionally resulted in surface artifacts. This issue was resolved by applying the Marching Cubes algorithm with an optimized threshold and then post-processing the mesh for smoothness. The transition to rasterized rendering through OpenGL allowed fine-tuning the model's visual output using differentiable rendering, further improving the final quality. Throughout the experiments, hyperparameter tuning played a critical role. Parameters such as the step size in ray marching and the sampling density in the spatial hierarchy required careful adjustment to balance accuracy and efficiency. In particular, targeted experiments refined the model's ability to learn photon angular momentum and simulate accurate light-matter interactions.

These optimizations and troubleshooting efforts ensured the robustness of 3D-LQF. The final model demonstrated superior performance in PSNR, SSIM, and LPIPS and handled static and dynamic scenes with high fidelity. Integrating differentiable rendering further allowed end-to-end optimization, ensuring the reconstruction matched ground-truth scenes as closely as possible. While computationally demanding, combining quantum principles with neural techniques proved effective in pushing the boundaries of 3D scene reconstruction.

V. Conclusion

We introduced 3D Light Quantum Fields (3D-LQF), a novel framework for 3D scene reconstruction using monocular 2D images that combines implicit representations and quantum mechanics. A key innovation is the use of SDFs to represent

3D objects efficiently. This is the first framework to use these curvature-based principles for modeling and rendering object surfaces. We extend the implicit SDF representation into 3D meshes and employ an OpenGL-based, end-to-end, differentiable rendering pipeline. This pipeline optimizes the entire model through gradient descent, providing precise scene reconstruction with image-based loss functions. By incorporating photon probability density functions, 3D-LQF achieves high visual fidelity with efficient light path simulation. In summary, 3D-LQF sets a new standard in 3D scene reconstruction by integrating quantum mechanics, advanced geometry, and differentiable rendering. This work opens new possibilities for accurately and efficiently modeling complex surfaces and dynamic environments.

REFERENCES

[1] S. Wu, C. Rupprecht, and A. Vedaldi, "Unsupervised learning of probably symmetric deformable 3d objects from images in the wild," in *CVPR*, 2020.

[2] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *ECCV*, 2020, pp. 405–421.

[3] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023. [Online]. Available: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

[4] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-nerf: Neural radiance fields for dynamic scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 10 318–10 327.

[5] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, P. Hedman, R. Martin-Brualla, and N. Snavely, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5470–5479.

[6] R. Martin-Brualla, N. Radwan, M. S. M. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. 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 (CVPR)*, 2021, pp. 7210–7219.

[7] Z. Yu, A. Chen, B. Huang, T. Sattler, and A. Geiger, "Mip-splatting: Alias-free 3d gaussian splatting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 19 447–19 456.

[8] H. Kato, Y. Ushiku, and T. Harada, "Neural 3d mesh renderer," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3907–3916.

[9] H. Hoppe, T. DeRose, T. Duchamp, J. Halstead, H. Jin, W. McDonald, and W. Stuetzle, "Mesh optimization," in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1993, pp. 19–26.

[10] Q. Zhang, Y. Xu, Y. Shen, B. Dai, B. Zhou, and C. Yang, "BerfScene: Generative novel view synthesis with 3D-aware diffusion models," in *CVPR*, 2024.

[11] Y. Zhang, Z. Zhu, W. Zheng, J. Huang, G. Huang, J. Zhou, and J. Lu, "Beverse: Unified perception and prediction in birds-eye-view for vision-centric autonomous driving," *arXiv preprint arXiv:2205.09743*, 2022.

[12] R. P. Feynman, "Space-time approach to non-relativistic quantum mechanics," *Reviews of Modern Physics*, vol. 20, no. 2, pp. 367–387, 1948.

[13] E. Schrödinger, "Quantisierung als eigenwertproblem (erste mitteilung)," *Annalen der Physik*, vol. 79, pp. 361–376, 1926.

[14] E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules," *Physical Review*, vol. 28, no. 6, pp. 1049–1070, 1926.

[15] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021. [Online]. Available: https://arxiv.org/abs/1912.09363

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 5998–6008.

[17] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. ACM, 1996, pp. 31–42.

[18] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. ACM, 1996, pp. 43–54.

[19] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660.

[20] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5099–5108.

[21] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 922–928.

[22] Y. Li, Q. Chen, H. Zha, and J. Liu, "Deep voxel-based learning for 3d shape analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 12 334–12 342.

[23] X. Zhu, W. Zhang, W. Liu, and S. Liu, "Vonn: Voxel octree neural networks for 3d shape classification and re-

trieval," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–11, 2019.

[24] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1997, pp. 209–216.

[25] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," in *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, 2022, pp. 102:1–102:15.

[26] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa, "Plenoxels: Radiance fields without neural networks," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5491–5500.

[27] Z. Li, T. Müller, A. Evans, R. H. Taylor, M. Unberath, M.-Y. Liu, and C.-H. Lin, "Neuralangelo: High-fidelity neural surface reconstruction," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[28] D. Shreiner, G. Sellers, J. Kessenich, and B. Licea-Kane, *OpenGL Programming Guide: The Official Guide to Learning OpenGL, Version 4.3*, 8th ed. Addison-Wesley Professional, 2013.

[29] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 165–174.

[30] J. C. Hart, *Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces*. The Visual Computer, 1996, vol. 12.

[31] B. Riemann, *Über die Hypothesen, welche der Geometrie zu Grunde liegen*. Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen, 1868. [Online]. Available: https://archive.org/details/riemann1868

[32] M. Spivak, *A Comprehensive Introduction to Differential Geometry*. Publish or Perish, 1999.

[33] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*. Dover Publications, 2016.

[34] M. Zwicker, H. Pfister, J. Van Baar, and M. Gross, "Ewa volume splatting," in *Proceedings of the IEEE Visualization Conference (VIS)*. IEEE, 2001, pp. 29–36.

[35] S. Park, M. Son, S. Jang, Y. C. Ahn, J.-Y. Kim, and N. Kang, "Temporal interpolation is all you need for dynamic neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 4212–4221.

[36] NVIDIA, "Nvidia dynamic scenes dataset," https://developer.nvidia.com/, 2021.

[37] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," 2021.

[38] W. Hu, Y. Wang, L. Ma, B. Yang, L. Gao, X. Liu, and Y. Ma, "Tri-miprf: Tri-mip representation for efficient anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 19 774–19 783.

[39] A. Knapitsch, J. Park, Q. Zhou, and V. Koltun, "Tanks and temples: Benchmarking large-scale scene reconstruction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3157–3166.

[40] P. Hedman and J. Kopf, "Deep blending for free-viewpoint image-based rendering," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, 2018.

[41] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 1987, pp. 163–169.

[42] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Zip-nerf: Anti-aliased grid-based neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 19 697–19 705.

[43] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, and Y. Sheikh, "Neural volumes: Learning dynamic renderable volumes from images," *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 65:1–65:14, July 2019. [Online]. Available: https://doi.org/10.1145/3306346.3323020

[44] Z. Li, S. Niklaus, N. Snavely, and O. 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 (CVPR)*, 2021, pp. 6498–6508.

[45] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Nerfies: Deformable neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 5865–5874.

[46] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. Khademi Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–14, 2019.

[47] T. Li, M. Slavcheva, M. Zollhoefer, S. Green, C. Lassner, C. Kim, T. Schmidt, S. Lovegrove, M. Goesele, R. Newcombe *et al.*, "Neural 3d video synthesis from multi-view video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5521–5531.