NeuFG: Neural Fuzzy Geometric Representation for 3D Reconstruction

Qingqi Hong, *Member*, *IEEE*, Chuanfeng Yang, Jiahui Chen, Zihan Li, Qingqiang Wu, Qingde Li, and Jie Tian, *Fellow*, *IEEE*

***Abstract*—3D reconstruction from multi-view images is considered as a longstanding problem in computer vision and graphics. In order to achieve high-fidelity geometry and appearance of 3D scenes, this paper proposes a novel geometric object learning method for multi-view reconstruction with fuzzy set theory. We establish a new neural 3D reconstruction theoretical frame called neural fuzzy geometric representation (NeuFG), which is a special type of implicit representation of geometric scene that only takes value in [0, 1]. NeuFG is essentially a volume image, and thus can be visualized directly with the conventional volume rendering technique. Extensive experiments on two public datasets, i.e., DTU and BlendedMVS, show that our method has the ability of accurately reconstructing complex shapes with vivid geometric details, without the requirement of mask supervision. Both qualitative and quantitative comparisons demonstrate that the proposed method has superior performance over the state-of-the-art neural scene representation methods. The code will be released on GitHub soon.**

***Index Terms*—3D Reconstruction, Multi-View, Neural Rendering, Fuzzy Set Theory**

I. Introduction

Reconstructing the geometry and appearance of 3D scenes from multi-view images is one of the fundamental problems in computer vision and graphics. In recent years, 3D reconstruction with neural rendering has become a promising alternative to conventional reconstruction techniques, of which recent works have shown remarkable performance on novel view synthesis [1]–[9] [10], [11] and geometry reconstruction [12]–[25] [26], [27] from a set of images.

Generally, neural rendering techniques can be categorized into two groups: surface rendering based approaches and volume rendering based approaches. Surface rendering based approaches determine the color of a viewing ray by first finding the intersection between the ray and the scene geometry, then obtaining the RGB value of this point by applying a certain surface illumination model. In this type of methods, the gradient is only back propagated to a local region near the intersection, which makes it difficult in reconstructing complex geometric objects involving sudden change in depths or severe self-occlusions [28]. In addition, mask supervision is usually required to converge to a valid surface.

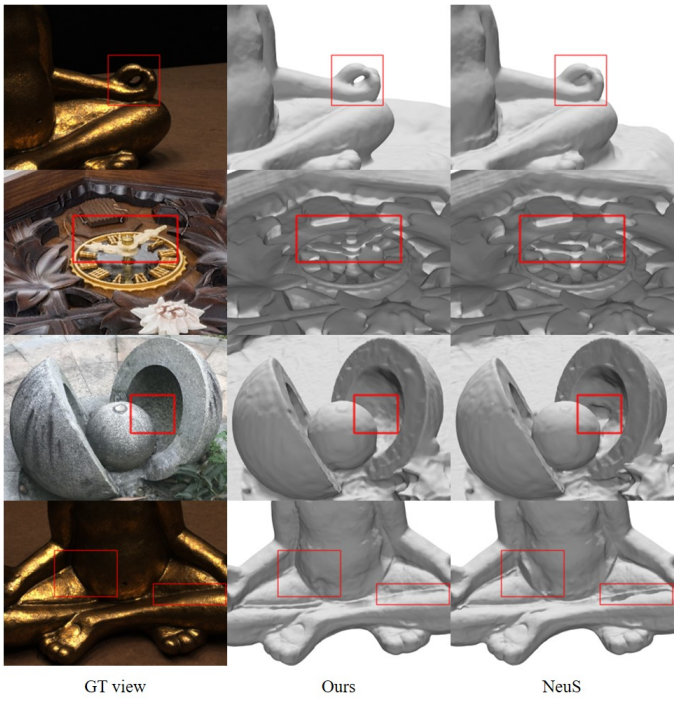


Fig. 1. Visual comparisons between our method and NeuS. First row: shape with tiny hole; Second row: shape with slender structures; Third row: input image with shadow; Fourth row: shape with texutreless region.

On the other hand, volume rendering based approaches represent the scene as a continuous field of volume density or occupancy, and render an image by integrating the amount of light intensity arriving at the camera along each view ray. In recent years, NeRF [3] and its follow-ups have achieved excellent results for novel view synthesis and larger scene rendering. However, accurate geometry extraction from the underlying volume density remains a challenging task. Existing methods often lead to artifacts, redundant surface patches, and incomplete and non-smooth surfaces. This is because the radiance field representation is generally too flexible to constrain the 3D geometry sufficiently, especially in the presence of ambiguities [29].

Consequently, in order to capture high-quality geometry, several works [28]–[30] have attempted to combine the im-

This work was supported in part by National Natural Science Foundation of China (No. 61502402), and the commissioned project of MetaMaker Inc.. *(Chuanfeng Yang and Jiahui Chen have equal comtributions on this work. Corresponding author(s): Qingqi Hong and Qingde Li.)*

Qingqi Hong, Chuanfeng Yang, Jiahui Chen, and Qingqiang Wu are with the Center for Digital Media Computing, School of Film, School of Informatics, Xiamen University, Xiamen 361005, China (e-mail: hongqq@xmu.edu.cn).

Zihan Li is with Xiamen University and University of Washington, Seattle, USA (e-mail: lizihanlizh@foxmail.com).

Qingde Li is with the School of Computer Science, University of Hull, Hull, HU6 7RX, UK (e-mail: Q.li@hull.ac.uk).

Jie Tian is with the School of Engineering Medicine, Beihang University, Beijing, China, and the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: tian@ieee.org).

plicit surface representation with neural volume rendering scheme for multi-view reconstruction. Generally, these methods can extract smooth surface from the learned volume density field. However, as shown in Fig. 1, when there is no mask supervision, they are still struggling in reconstructing subtle surface details such as objects with tiny holes and slender structures. In addition, the reconstruction results are also often greatly affected by the shadow of the input image. They may even fail in reconstructing the texture-less regions for some geometric objects.

In this paper, we propose a novel geometric object learning method for multi-view reconstruction with fuzzy set theory. By considering different volumetric geometric shapes in the scene as the **visible section of the light radiation fields** generated from these geometric objects **in terms of 3D fuzzy point sets**, we establish a **new neural 3D reconstruction theoretical frame** called **neural fuzzy geometric representation (NeuFG)**. The scene learned by NeuFG is essentially a mapping from R^3 to [0, 1], and thus can be considered naturally as a volume image and visualized following the standard volume rendering formulation. Extensive experiments on two public datasets, i.e., DTU and BlendedMVS, show that our method has the ability of accurately reconstructing complex shapes with vivid geometric details, without the requirement of mask supervision. Both qualitative and quantitative comparisons demonstrate that the proposed method has superior performance over the state-of-the-art neural scene representation methods.

In summary, our contributions are as follows:

- We establish a **new neural shape learning theoretical frame** called neural fuzzy geometric representation (NeuFG) by considering different volumetric shapes as the **light radiation fields** generated from these geometric objects **in terms of fuzzy sets**.
- We present how this representation can be used for multi-view reconstruction with the standard volume rendering formulation.
- We experimentally validate that our technique is capable of reconstructing high-quality geometries of 3D scenes and achieving superior results over the state-of-the-art methods.

II. Related Works

A. Multi-view 3D Reconstruction

In the past decades, 3D reconstruction from multi-view images is considered as a longstanding problem in computer vision and graphics. Before the era of deep learning, classical techniques for multi-view stereo (MVS) reconstruction can be divided into depth-based methods [31]–[33] and voxel-based methods [34]–[36]. In depth-based methods, dense point clouds are first generated by fusing depth maps; a point cloud meshing technique is then employed to generate the surface (Please refer to [37] for the recent advances in depth completion and depth maps fusion). Depth-based methods generally require complicated rendering pipelines, which could result in incomplete 3D models due to the accumulated errors of all stages [30]. On the other hand, voxel-based methods directly represent shapes in a volume with a voxel grid, which can generate complete models but limited to low resolution due to high memory requirements. At the era of deep learning, some parts of the classic MVS pipeline can be replaced with learning-based techniques. For instance, several works have conducted on learning feature matching [38]–[41], depth fusion [42], [43] or depth inference [44]–[46] from multi-view images. Generally, these learning-based MVS methods have difficulty to generate high-quality 3D geometries and synthesize photorealistic novel views [29].

B. Neural Rendering

In recent years, 3D reconstruction with neural rendering has become a highly promising technique to achieve excellent performance on novel view synthesis [1]–[9], [47], [48] and geometry reconstruction [13]–[20], [49] from multi-view images. Generally, the related works about neural rendering can be categorized as either surface rendering or volume rendering. Surface rendering-based approaches, like IDR [50] and DVR [51], utilize a differentiable rendering pipeline to generate images from a 3D object for the purpose of supervision. For example, by conditioning on the viewing direction, IDR is able to capture a high level of surface detail and achieve impressive reconstruction results, even in the presence of non-lambertian surfaces. However, both DVR and IDR require pixel-accurate object masks for all views as input and may fail to reconstruct objects with complex structures that causes abrupt changes in depth [28].

On the other hand, volume rendering-based approaches, like NeRF [3] and follow-ups [52]–[59], utilize volume rendering by learning the accumulated radiance intensity along each view ray. This kind of methods is cable of producing excellent results on novel view synthesis without the requirement of mask supervision. However, the recovered 3D geometry based on these methods is far from satisfactory [29]. Aiming at capturing high-quality geometry of objects, several other works [28]–[30] have attempted to combine the implicit surface representation with neural volume rendering scheme for multi-view reconstruction. For instance, Oechsle et al. [29] proposed a principled unified framework, named UNISURF, to unify surface and volume rendering via a hybrid MLP representation. Although UNISURF can improve the reconstruction quality in a certain extent by shrinking the sample region of volume rendering during the optimization, there is still room for further improvement of the reconstructed surface's accuracy [28]. By modeling the volume density as a transformed signed distance function (SDF), VolSDF [30] is able to produce high-quality geometry reconstructions. However, additional sampling algorithm is required for the approximation of opacity for volume rendering of the new density representation. In NeuS [28], 3D surfaces are modelled as neural SDF and volume rendering is applied to train this SDF representation with robustness. NeuS allowed to achieve an unbiased estimate of the corresponding surface without additional sampling algorithm. However, it still has difficulty to accurately reconstruct shapes with tiny holes and to correctly reconstruct the textureless region of the

surface, and the reconstruction results are often affected by the shadow of the input image. Consequently, in order to improve neural implicit surface reconstruction, more recent works have focused on introducing additional constraints of geometry optimization into volume rendering scheme. For example, Geo-Neus [60] introduced SDF optimization constraints and geometry-consistent constraints into Neus to focus on the true surface optimization. Similarly, MonoSDF [61] introduced depth consistency loss and normal consistency loss on the basis of VolSDF.

III. Method

In this section, we firstly propose a new representation of scene's geometry with fuzzy set theory. Then we present the volume rendering scheme based on the new neural fuzzy geometric representation. Finally, the loss function for training our model is presented at the end of this section.

A. Implicit geometric representation in 3D fuzzy sets

a) Neural fuzzy geometric representation: The color value at a pixel of an image taken with a digital camera based on a given configuration of the scene is basically the light intensity received at the pixel when considering the pixel as a pinhole. To accurately reconstruct the scene geometry from a given image is fundamentally a task of estimating the light field of the scene. Given the limitation of the devices commonly used to visualize the light field, in general, only a fraction of the whole light spectrum can be sampled. Therefore, a digital image generated from the light field with a given device is not based on the whole light field, rather, it is based only on a very limited section of the light spectrum that is visible to the device. When represented in floating point value, in practice, the light intensity captured at each position is usually represented with a value in [0, 1], which is a direct mapping from the visible section of the light spectrum to interval [0, 1]. As is commonly known, three basic phenomena will happen when light interacts a geometric object: reflection, refraction, and absorption, which subsequently generate three volumetric fields, namely, the fields of reflection, refraction, and shadow. One typical feature of the three fields is that all these fields are of a soft boundary, with its degree of softness varying according to the distance between the objects and the light sources (see Fig. 2). Thus, the light field captured by a digital imaging device can be naturally modelled as 3D fuzzy point sets represented in terms of fuzzy set membership.

Fig. 2. The light field as a result of light-objects interaction is of a soft boundary with the degree of softness varying according to the distance to the objects.

Our **key insight** is that, by **considering different volumetric geometric shapes** in the scene **as the light radiation fields generated from these geometric objects in terms of fuzzy sets**, we establish a **new neural shape learning theoretical frame** called neural fuzzy geometric representation (NeuFG). Light field captured by a camera is only a fraction of the real light spectrum. The radiance field indicated by multiple views is therefore only a fraction of the real-world light field visible to the camera. Compared with other methods, e.g., NeuS, **our mathematical framework better explains the nature of the light field captured by a camera**.

In fuzzy set theory, a fuzzy set for a given domain space χ is defined as a membership function taking value in [0, 1]. An element with a membership 1 indicates that the element is in the set, and 0 if the element not in the set. For an element with membership value larger than 0 and less than 1, say, 0.8, it indicates that the element is partially included in the set with a degree of membership of 80%, neither definitely in, nor definitely out. Fuzzy set is a natural generalization of conventional concept of subset. An ordinary subset in space χ can be considered as a special type of fuzzy set, whose membership function only takes either value 0 or 1.

A fuzzy set on R^3 is naturally a mapping from $R^3 \to [0, 1]$, which is perfectly suitable for neural scene geometric field representation when the mapping is differentiable. For neural scene representation, a neural network can be trained to represent the geometric scene in terms of fuzzy set membership functions on R^3, a special type of implicit function which only takes value from interval [0, 1]. In this research, we show how to construct a differentiable membership function to assign to every location $\boldsymbol{P} \in R^3$ a membership value varying between 0 and 1 via neural network learning.

For an ordinary 3D geometric object, its membership function can be constructed directly by composing its implicit representation with the Heaviside step function. Let the implicit function representing the geometric shape be $f(\boldsymbol{P})$ ($\boldsymbol{P} \in R^3$), with $f(\boldsymbol{P}) = 0$ representing set of points on the boundary Ω, whereas, $f(\boldsymbol{P}) > 0$ in the interior region Ω^+, and $f(\boldsymbol{P}) < 0$ in the exterior region Ω^-. Let $H(x)$ be the Heaviside step function, that is,

$$H(x) := \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \tag{1}$$

Then the membership function of the 3D object can be defined as the composition:

$$\mu(\boldsymbol{P}) = H \circ f(\boldsymbol{P}) \tag{2}$$

which can naturally perform the mapping from $R^3 \to R \to \{0, 1\}$.

Generally, there are several points for the benefits of using fuzzy set theory in implicit representation:

- Multiple view-based 3D neural learning is fundamentally about radiance field reconstruction, as **the colour at each pixel of an image is essentially the amount of light accumulated at the pixel**. For any object in the scene, once illuminated, it can be considered as a kind of secondary light source, just like the Moon, which does give light, but once illuminated by the Sun, it becomes a light source.

- For a geometric object with diffuse surface, the light field generated by the object with albedo light source can be considered naturally as a 3D fuzzy object without a clear boundary. **Any 3D object without a clear boundary can be directly modelled as an implicit function** (Please refer to [62] for more details).
- In addition, by viewing an object in 3D as **a fuzzy set-based representation can be directly considered as a volume image** when re-represented in a discrete 3D space.

b) Piecewise polynomial smooth step function for neural fuzzy represented implicit shapes: As can be seen directly that the mapping from R^3 to $\{0, 1\}$ obtained by converting an implicit representation of a geometric object using the classical Heaviside step function is not continuous and differentiable, which is not suitable for the training of neural networks [63]. Hence, smooth step function is required to perform a continuous 3D mapping for neural rendering of implicit shapes and to achieve high-quality reconstruction. Fig. 3 presents a visual illustration how an implicitly represented geometric object can be interpreted as a fuzzy set.

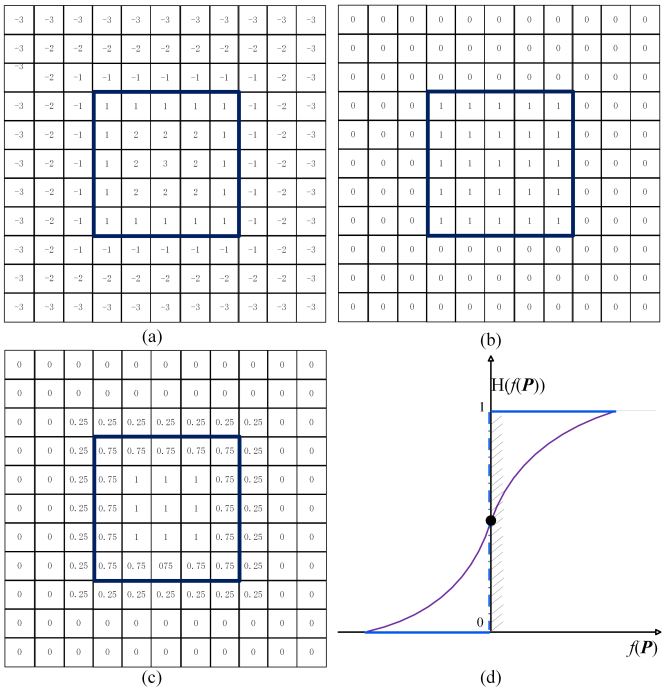


Fig. 3. A visual illustration how an implicitly represented geometric object can be interpreted as a fuzzy set. (a) A square is defined by the signed distance function (a subset of implicit surface function) $f(\boldsymbol{P})$; (b) The mapping is performed by the classical Heaviside step function; (c) The mapping is performed by the smooth step function; (d) The diagrams of classical Heaviside step function (blue dot curve) and smooth step function (purple solid curve).

In order to achieve high-quality reconstruction of implicit surfaces, we apply the piecewise polynomials smooth step function (PPSSF) [64] to the neural rendering of fuzzy set represented implicit functions. **Piecewise polynomial representation** of the smooth step function allows **more accurate and efficient** implementation than non-polynomial represented smooth step functions. PPSSF is defined by starting with the standard Heaviside unit step function and can be expressed either in iterative or explicit forms [64].

The explicit form of degree n PPSSF is expressed as follows:

$$H_n(x) = \frac{1}{n!2^n} \sum_{k=0}^{n} (-1)^k \binom{n}{k} (x + (n - 2k))^n H_0(x + (n - 2k)) \tag{3}$$

which is obtained as the recursive convolution of $H_n(x) \star g(x)$, where $H_0(x)$ is the classical Heaviside step function and $g(x) = (H_0(x + 1) - H_0(x - 1))/2$.

$H_n(x)$ increases from 0 to 1 monotonically over the interval $[-n; n]$, which can be modified to $[-\delta, \delta]$ by introducing a nonnegative number $\delta > 0$.

Finally, our neural fuzzy set implicit representation for a 3D object can be rewritten as:

$$\mu(\boldsymbol{P}) = H_n \circ f(\boldsymbol{P}) = H_n(f(\boldsymbol{P})) \tag{4}$$

B. Rendering scheme based on NeuFG

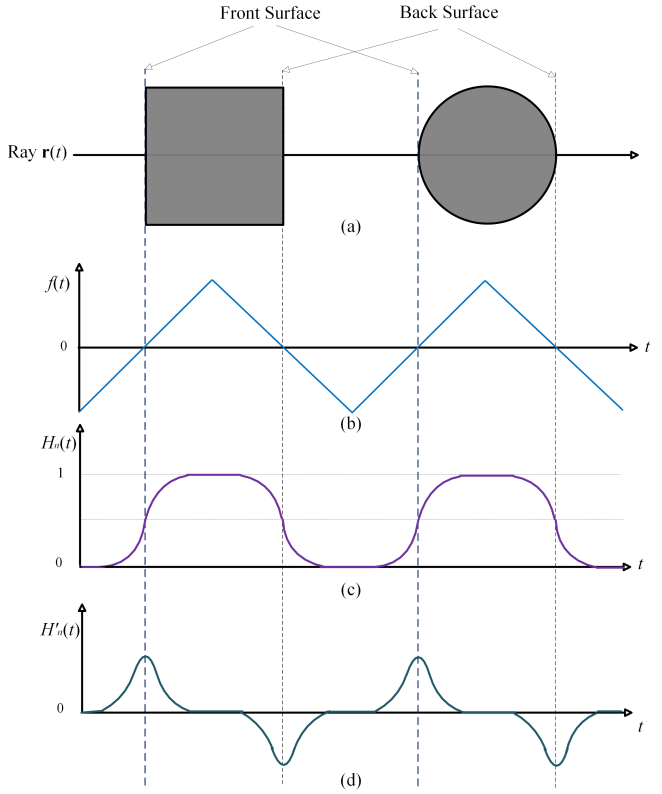


Fig. 4. Illustration of rendering scheme based on neural fuzzy geometric representation. (a) A camera ray penetrating into two objects; (b) The signed distance function of the objects; (c) The neural fuzzy geometric representation of the objects; (d) The neural fuzzy geometric representation is both unbiased and occlusion-aware.

a) Rendering scheme: The technique of volume rendering considers a camera ray $\mathbf{r}$ emanating from a position $\mathbf{o} \in R^3$ in direction $\mathbf{d} \in R^3$, $|\mathbf{d}| = 1$ defined by $\mathbf{r}(t) = \mathbf{o}+\mathbf{d}t, t \leq 0$. In the classical volume rendering [65], the expected color $C(\mathbf{r})$ is accumulated along the ray by

$$C(\mathbf{r}) = \int_0^{+\infty} T(t)\sigma(\mathbf{r}(t))c(\mathbf{r}(t), \mathbf{d})\, \mathrm{d}t \tag{5}$$

where $\sigma(\mathbf{r}(t))$ is the volume density, and $T(t) = \exp\Big(-\int_0^t \sigma(\mathbf{r}(s))\,\mathrm{d}s\Big)$ denotes the accumulated transmittance along the ray.

Given the **obvious physical and geometric meaning about** $H_n(f(\boldsymbol{P}))$, we can just model the opacity $\mathbf{O}$ directly in terms of $H_n(f(\mathbf{r}(t)))$ shown below:

$$\mathbf{O}(t) = 1 - T(t) = H_n(f(\mathbf{r}(t))) \tag{6}$$

According to the **properties of smooth step function** $H_n(x)$, $H_n(x)$ can be **considered as a cumulative distribution function of radiance field's intensity**. Thus, the density function corresponding to the light radiance's intensity can thus be directly expressed as the derivative of H_n:

$$H_n'(t) = \frac{dH_n(f(\mathbf{r}(t)))}{dt} = H_n'(f(\mathbf{r}(t)))\nabla f(\mathbf{r}(t)) \cdot \mathbf{d} \tag{7}$$

which is typically extremely concentrated near the object's front boundary when a geometric object is opaque.

Fig. 4 is the illustration of rendering scheme based on neural fuzzy geometric representation. As can be observed from the figures and the properties of $H_n(x)$, $H_n'(t)$ **is not only unbiased, but also occlusion-aware**. It is obvious that, $H_n'(t)$ **has peak values only when** $f(\mathbf{r}(t)) = 0$, which guarantees that the geometric surfaces can be reconstructed **without bias**. In addition, the peak value of $H_n'(t)$ is **positive at the front surface while negative at the back surface** of the object. That is, only the colors for the front surface need be blended **with occlusion-awareness**.

b) The derived new discrete opacity: According to the discretization of the standard volume rendering formulation [3], we have:

$$\widehat{C} = \sum_{i=1}^{n} T_i \alpha_i c_i \tag{8}$$

where T_i is the discrete accumulated transmittance defined by $T_i = \prod_{j=1}^{i-1}(1-\alpha_j)$, and α_i is the is discrete opacity values defined by

$$\alpha_i = 1 - \exp\Big(-\int_{t_i}^{t_i+1} \sigma(\mathbf{r}(t))\,\mathrm{d}t\Big) \tag{9}$$

Based on our NeuFG, we can conveniently derive the new α_i from $H_n(\mathbf{r}(t))$. According to Eq.(6), we can directly model the transmitance function in terms of $H_n(t)$:

$$T(t) = 1 - H_n(f(\mathbf{r}(t))) \tag{10}$$

Expanding the expression of $T(t)$ yields:

$$\exp\Big(-\int_0^t \sigma(\mathbf{r}(s))\,\mathrm{d}s\Big) = 1 - H_n(f(\mathbf{r}(t))) \tag{11}$$

Discretizating Eq.(11), yields:

$$\begin{aligned}
&\exp\Big(-\int_{t_i}^{t_{i+1}} \sigma(\mathbf{r}(t))\,\mathrm{d}t\Big) \\
&= \exp\Big(-\Big(\int_0^{t_{i+1}} \sigma(\mathbf{r}(t))\,\mathrm{d}t - \int_0^{t_i} \sigma(\mathbf{r}(t))\,\mathrm{d}t\Big)\Big) \\
&= \exp\Big(-\int_0^{t_{i+1}} \sigma(\mathbf{r}(t))\,\mathrm{d}t\Big)\exp\Big(\int_0^{t_i} \sigma(\mathbf{r}(t))\,\mathrm{d}t\Big) \\
&= \frac{\exp\Big(-\int_0^{t_{i+1}} \sigma(\mathbf{r}(t))\,\mathrm{d}t\Big)}{\exp\Big(-\int_0^{t_i} \sigma(\mathbf{r}(t))\,\mathrm{d}t\Big)} \\
&= \frac{1 - H_n(f(\mathbf{r}(t_{i+1})))}{1 - H_n(f(\mathbf{r}(t_i)))}
\end{aligned} \tag{12}$$

Then, by combining Eq. (9) and Eq. (12), we can derive the new α_i in terms of $H_n(t)$:

$$\begin{aligned}
\alpha_i &= 1 - \frac{1 - H_n(f(\mathbf{r}(t_{i+1})))}{1 - H_n(f(\mathbf{r}(t_i)))} \\
&= \frac{1 - H_n(f(\mathbf{r}(t_i))) - (1 - H_n(f(\mathbf{r}(t_{i+1}))))}{1 - H_n(f(\mathbf{r}(t_i)))} \\
&= \frac{H_n(f(\mathbf{r}(t_{i+1}))) - H_n(f(\mathbf{r}(t_i)))}{1 - H_n(f(\mathbf{r}(t_i)))}
\end{aligned} \tag{13}$$

As can be seen from Fig. 4 (d), $H_n'(t)$ is negative at the back surface of the object, i.e., $H_n(f(\mathbf{r}(t_{i+1}))) - H_n(f(\mathbf{r}(t_i))) < 0$. However, α_i should be non-negative, thus we clip it against zero:

$$\alpha_i = \max\Big(\frac{H_n(f(\mathbf{r}(t_{i+1}))) - H_n(f(\mathbf{r}(t_i)))}{1 - H_n(f(\mathbf{r}(t_i)))}, 0\Big) \tag{14}$$

C. Loss function

To train the model of NeuFG, we minimize the difference between the rendered colors and the ground truth colors with the constraining of Eikonal term [66] for geometric regularization. Besides that, masks can also be utilized for supervision if provided. Specifically, we optimize our neural networks and δ by randomly sampling a batch of pixels $\mathcal{P}$ and their corresponding rays in world space. For each pixel $p \in \mathcal{P}$ we have $(C_p, M_p, \mathbf{o}_p, \mathbf{d}_p)$, where $C_p \in R^3$ is its intensity (RGB color), $M_p \in \{0, 1\}$ is its optional mask, $\mathbf{o}_p \in R^3$ is the camera location, and $\mathbf{d}_p \in R^3$ is the viewing direction (camera to pixel). Our training loss consists of three terms:

$$L = L_{RGB} + \lambda L_E + \rho L_M \tag{15}$$

where L_{RGB} is the color loss, L_E is the Eikonal loss, L_M is the optional mask loss, and λ and ρ are hyper-parameters.

The color loss is defined as

$$L_{RGB} = \frac{1}{\mid \mathcal{P} \mid} \sum_p \mid \widehat{C_p} - C_p \mid \tag{16}$$

where $\mid \cdot \mid$ denotes the 1-norm, and $\widehat{C_p}$ is the numerical approximation to the volume rendering integral in Eq. (8).

TABLE I
THE QUANTITATIVE COMPARISONS (CHAMFER DISTANCES) BETWEEN OUR METHOD AND BASELINES ON THE DTU MVS DATASET IN BOTH SETTINGS – WITH/WITHOUT MASK SUPERVISION. BOLD DENOTES THE BEST PERFORMANCE.

	With Mask Supervision					**Without Mask Supervision**					
Scan ID	IDR	NeRF	NeuS	NeuFG (ours)	COLMAP	NeRF	Mip-NeRF	UNISURF	VolSDF	NeuS	NeuFG (ours)
scan24	1.63	1.83	0.83	**0.68**	**0.81**	1.90	1.98	1.32	1.14	1.00	0.96
scan37	1.87	2.39	0.98	**0.93**	2.05	1.60	1.65	1.83	1.26	1.37	**1.22**
scan40	0.63	1.79	0.56	**0.49**	**0.73**	1.85	1.65	1.72	0.81	0.93	0.77
scan55	0.48	0.66	**0.37**	**0.37**	1.22	0.58	1.61	0.44	0.49	0.43	**0.40**
scan63	**1.04**	1.79	1.13	1.18	1.79	2.28	2.90	1.35	1.25	**1.10**	1.16
scan65	0.79	1.44	**0.59**	0.63	1.58	1.27	1.79	0.79	0.70	0.65	**0.62**
scan69	0.77	1.50	**0.60**	**0.60**	1.02	1.47	1.51	0.80	0.72	**0.57**	**0.57**
scan83	1.33	**1.20**	1.45	1.33	3.05	1.67	1.93	1.49	1.29	1.48	**0.97**
scan97	1.16	1.96	**0.95**	1.00	1.40	2.05	2.19	1.37	1.00	1.09	**0.97**
scan105	0.76	1.27	0.78	**0.76**	2.05	1.07	1.32	0.89	**0.70**	0.83	0.94
scan106	0.67	1.44	**0.52**	**0.52**	1.00	0.88	0.83	0.59	0.66	0.52	**0.50**
scan110	**0.90**	2.61	1.43	1.30	1.32	2.53	2.52	1.47	**1.08**	1.20	1.10
scan114	0.42	1.04	**0.36**	**0.36**	0.49	1.06	1.38	0.46	0.42	**0.35**	0.36
scan118	0.51	1.13	0.45	**0.43**	0.78	1.15	1.49	0.59	0.61	**0.49**	**0.49**
scan122	0.53	0.99	**0.45**	0.50	1.17	0.96	1.12	0.62	0.55	0.54	**0.53**
mean	0.89	1.53	0.76	**0.73**	1.36	1.48	1.61	1.01	0.85	0.84	**0.77**

The Eikonal loss is defined as

$$L_E = \mathbb{E}_x(\parallel \nabla_x f(x) \parallel -1)^2 \tag{17}$$

which encourages $f(\cdot)$ to approximate a signed distance function with Implicit Geometric Regularization (IGR) [66]; the samples x are taken to combine a single random uniform space point and a single point from $\mathcal{P}$ for each pixel p.

The mask loss is defined as

$$L_M = BCE(M_p - \widehat{O_p}) \tag{18}$$

where $\widehat{O_p} = \sum_{i=1}^{n} T_{p,i}\alpha_{p,i}$ is the sum of weights along the camera ray, and BCE is the binary cross entropy loss.

IV. EXPERIMENTS

A. Baselines

To validate the effectiveness of our method, we compare it to several different baselines, namely, COLMAP [32], IDR [50], NeRF [53], Mip-NeRF360 [10] and UNISURF [29]/VolSDF [30]/NeuS [28]. COLMAP is considered as a widely-used classical MVS method with impressive performance on multi-view reconstruction. IDR is the state-of-the-art surface reconstruction approach which represents the scene geometric objects as the 0-level set of an implicit function represented in a differential neural network and visualized using a neural surface render. This method can reconstruct high-quality surfaces but requires input masks as supervision. Unlike IDR, Mip-NeRF360 is a state-of-the-art implicit geometric learning method based on volume rendering. While it can produce excellent results on novel view synthesis without the requirement of mask supervision, it is difficult to extract geometric surfaces accurately from the volume density learned by Mip-NeRF. Unisuf/VolSDF/NeuS combine the implicit surface representation with neural volume rendering scheme, which can extract high-quality geometric surfaces from the learned volume density field.

B. Datasets

To evaluate our approach against these baseline methods, we conduct extensive experiments on two public datasets, i.e., DTU MVS [32] and BlendedMVS [67]. The DTU MVS dataset contains multi-view images with respective extrinsic and intrinsic camera parameters at a resolution of 1200×1600. We evaluate our method on the 15 scans that were selected by IDR [50]. The BlendedMVS dataset is a large-scale dataset containing multi-view images with respective camera extrinsics and intrinsics. We use 7 examples from the low-res set containing 31 to 143 different views of unmasked images at a resolution of 576×768.

C. Implementation details

For the loss function (i.e., Eq.(15)), we set $\lambda = 0.1$, and $\rho = 0.1$ if mask is provided. The ADAM optimizer is used to train our neural networks. We assume the region of interest is inside a unit sphere, and sample 512 rays per batch and train our model for $300k$ iterations, which costs about 7 hours for the 'w/ mask' setting and 8.5 hours for the 'w/o mask' setting on a single NVIDIA RTX3090 GPU.

Similar to the hierarchical sampling strategy used in NeRF, we first uniformly sample 64 points along the ray, then we iteratively conduct importance sampling for 4 times. The coarse probability estimation in the $i-th$ iteration ($i = 1, \ldots, 4$) is computed by a fixed δ value, which is set as $1.4/2^i$. Consequently, the probability of fine sampling is computed by the learned δ value.

D. Quantitative comparisons

We conduct quantitative comparisons between our method and baselines on the DTU MVS dataset in both settings – with/without mask supervision. The reconstruction quality is measured with the Chamfer distances in the same way as IDR [50]. Table I reports the measured scores of our method and baselines, which shows that our approach outperforms the baseline methods on the DTU dataset in both settings – w/ and w/o mask. In detail, for mask supervision, our method

gets smaller average values of Chamfer distances with 0.15, 0.81, and 0.03 than that of IDR, NeRF, and NeuS, respectively. Particularly, in the setting of without mask supervision, our method achieves superior performance to all other methods, including COLMAP, NeRF, UNISURF, VolSDF, and NeuS. Specifically, our method can reduce the average value of Chamfer distances to 0.77, which is 0.07 less than that of the former best method (i.e., NeuS).

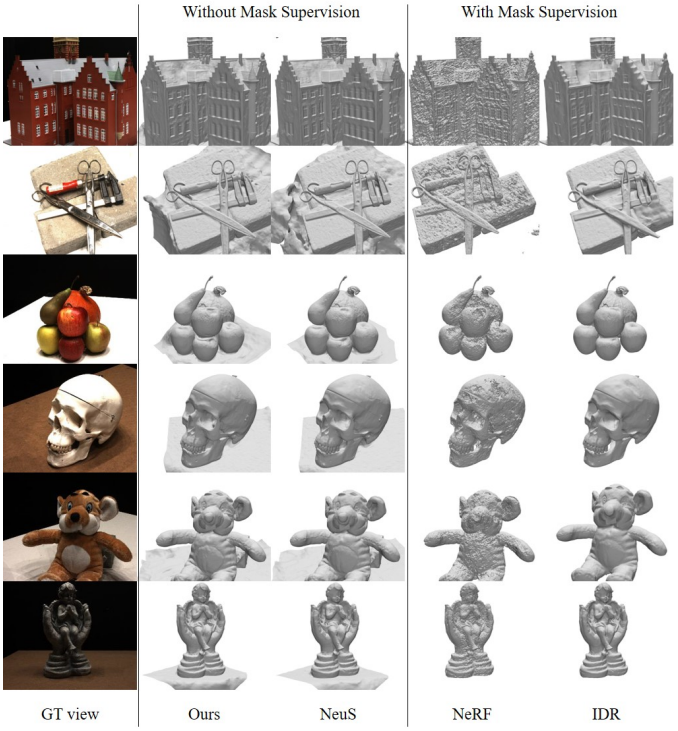


Fig. 5. Qualitative comparison of reconstructed surfaces from the DTU dataset.

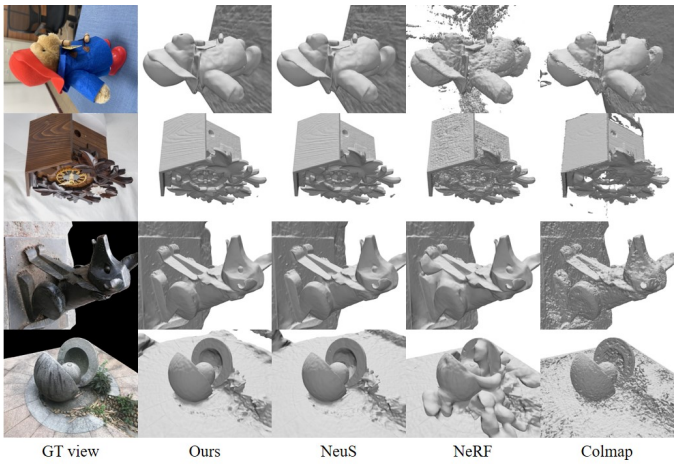


Fig. 6. Qualitative comparison of reconstructed surfaces from the BlendedMVS dataset. Note that the reconstructed results are all based on the setting of without mask supervision.

E. Qualitative comparisons

We conduct the qualitative comparisons on both the DTU dataset and the BlendedMVS in Fig. 5 and Fig. 6, respectively. As shown in the figures, compared to NeRF and COLMAP, our method, NeuS, and IDR can capture the overall spatial arrangement of the scene accurately and produce high-quality surfaces. In addition, our method does not require mask supervision when compared to IDR, and can capture more geometric details, e. g., tiny holes, slender structures, texture-less region, and shadow region, than NeuS does. The detail comparisons with NeuS will presented in the next section.

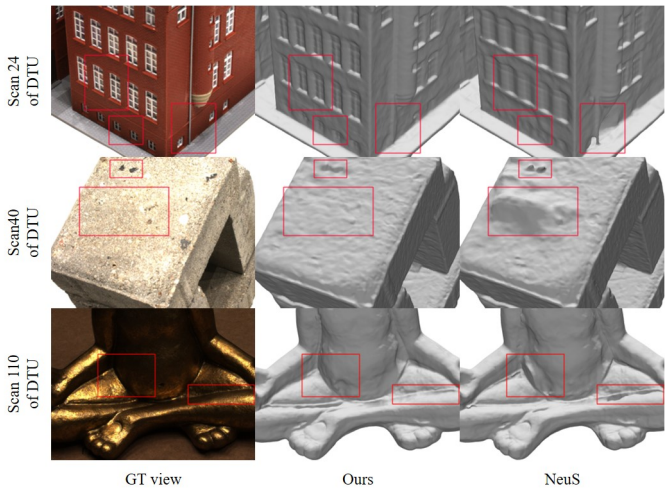


Fig. 7. Detail comparisons with NeuS for the setting of with mask supervision.

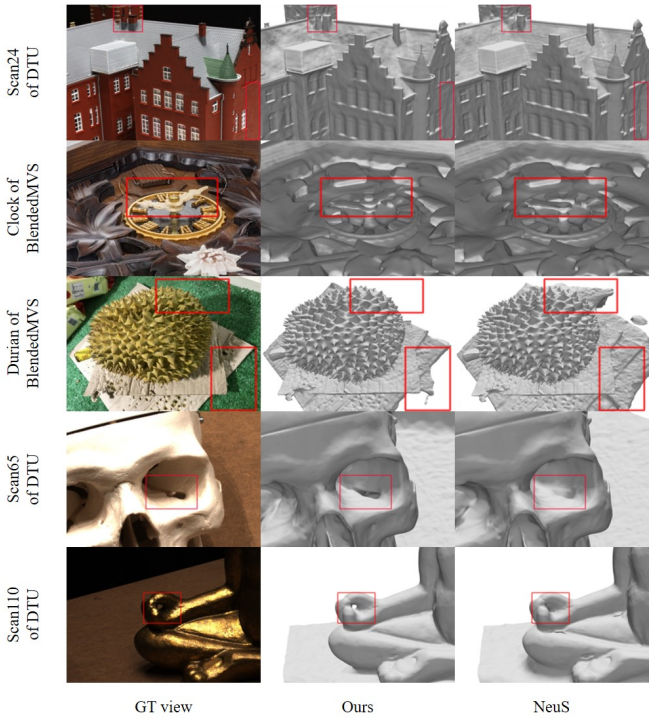


Fig. 8. Detail comparisons with NeuS for the setting of without mask supervision.

TABLE II
THE QUANTITATIVE COMPARISONS (CHAMFER DISTANCES) BETWEEN OUR METHOD COMBINED WITH EXPLICIT GEOMETRY SUPERVISION AND OTHER SOTA METHODS ON THE DTU DATASET. BOLD DENOTES THE BEST PERFORMANCE.

Scan ID	HF-Neus	VOXURF	2DGS	Geo-Neus	NeuFG$^+$(Ours)
scan24	0.76	0.65	0.48	0.50	**0.48**
scan37	1.32	0.74	0.91	**0.72**	0.76
scan40	0.7	0.39	0.39	**0.38**	**0.38**
scan55	0.39	**0.35**	0.39	0.37	0.38
scan63	1.06	0.96	1.01	0.92	**0.88**
scan65	0.63	0.64	0.83	**0.51**	**0.51**
scan69	0.63	0.85	0.81	0.50	**0.49**
scan83	1.15	1.58	1.36	1.27	**1.15**
scan97	1.12	1.01	1.27	0.89	**0.73**
scan105	0.80	0.60	0.76	**0.66**	**0.66**
scan106	0.52	1.11	0.70	0.51	**0.48**
scan110	1.22	0.37	1.40	0.80	**0.75**
scan114	0.33	0.45	0.40	**0.31**	**0.31**
scan118	0.49	0.47	0.76	0.42	**0.40**
scan122	0.50	0.72	0.52	0.44	**0.43**
mean	0.77	0.72	0.80	0.61	**0.58**

F. Detail comparisons with NeuS

This section demonstrates the detailed comparisons of our method with NeuS, which is an excellent method to reconstruct high-fidelity surfaces, in both settings – with/without mask supervision. As shown in Fig. 7 for the setting of w/ mask, the reconstruction result of NeuS is greatly affected by the shadow of the input image, while our method can correctly reconstruct the geometry under the shadow (Scan 24 of DTU). In addition, NeuS has difficulty in estimating the texture-less regions. This is because in the absence of textures, the parameters related to light reflection are unknown, and the materials required for rendering are also a part of learning. However, the proposed fuzzy set representation can be directly considered a volumetric image. It can be assumed that the surfaces of all 3D objects are diffused and can be rendered with or without material information. As shown in Scan 110 of DTU, our method is able to achieve much better reconstruction result than NeuS does for the textureless region of the surface. Particularly, for the setting of w/o mask, our method is much more powerful than NeuS in the aspect of capturing more geometric details. As shown in Fig. 8, our method can accurately reconstruct shapes with tiny holes and slender structures, such as the chimney of the house (Scan 24 of DTU), the holes of the bunny's hand (Scan 110 of DTU) and skull's eye (Scan 65 of DTU) , and the minute hand of the clock (Clock of BlendedMVS), while NeuS has difficulty to capture these geometric details.

G. Combination with geometry optimization

Generally, our neural fuzzy geometric representation can be **considered as a baseline and combined with explicit geometry supervision** to further improve implicit surface reconstruction. We conduct the experiment on the combination of our neural fuzzy geometric representation with the additional constraints of SDF optimization and geometry-consistent designed in Geo-NeuS [60]. For fair comparison, the experiment results of Geo-Neus are based on its official implementation repo[1]. As presented in Table II, our approach outperforms other SOTA methods, such as HF-Neus [68], VOXURF [69], 2DGS [70], and Geo-Neus [60]. In addition, we also conduct the qualitative comparisons with Geo-Neus, as shown in Fig. 9. Compared to Geo-Neus, our method has more powerful capabilities for capturing realistic geometric details (first three rows of Fig. 9) and repairing geometric integrity (last row of Fig. 9).

[1]https://github.com/GhiXu/Geo-Neus

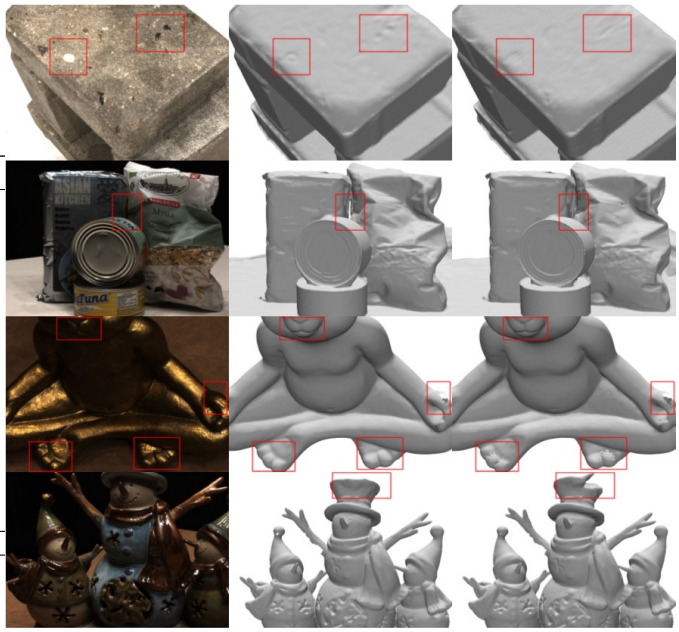


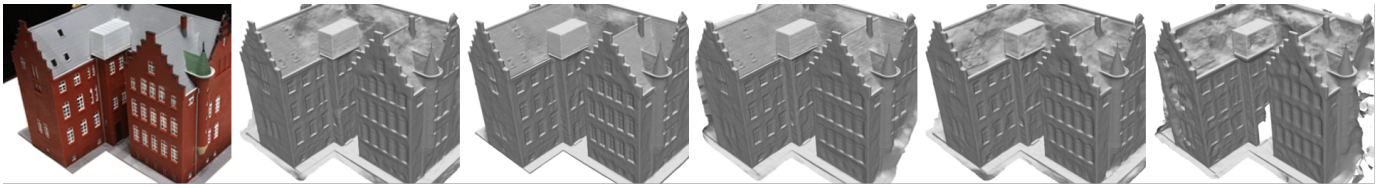
Fig. 9. Qualitative comparison of reconstructed surfaces between our method combined with explicit geometry supervision and Geo-Neus on the DTU dataset.

H. Ablation study

To demonstrate the effectiveness of the proposed method, we experimented the following different implementations: (a) $H_n(t)$ with $n = 2$; (b) $H_n(t)$ with $n = 3$; (c) $H_n(t)$ with $n = 4$; (d) replacing $H_n(t)$ by Sigmoid function $\sigma(t)$; (e) without Eikonal loss. The qualitative and quantitative results are shown in Figure 10. From the results, it is evident that a larger value of n leads to better reconstruction results but also requires more computation. As depicted in the figure, when n is set to 4, the roof of Scan 24 becomes remarkably flat, demonstrating the potential efficacy of our method. In our paper, we chose a balanced approach by selecting $n = 3$, which allows us to achieve an enhanced optimization boost without imposing excessive computational demands. In addition, our $H_n(t)$ outperforms Sigmoid function $\sigma(t)$, which is a popular function with similar properties and widely used for multi-view reconstruction. It is also worth noting that the omission of the Eikonal function can significantly influence the final results.

V. CONCLUSION

This paper presents **NeuFG**, a **novel geometric object learning method** for multi-view reconstruction with **3D fuzzy set theory**. The scene learned by NeuFG is essentially a



Refrence Image	$H_n(t)$, n=2	$H_n(t)$, n=3	$H_n(t)$, n=4	$\sigma(t)$	**w/o** Eikonal
Chamfer Distance	0.97	0.96	0.92	1.00	1.43

Fig. 10. Ablation studies. We depict the qualitative results and report the quantitative metrics in Chamfer distance.

mapping from R^3 to [0, 1], and thus can be considered as a volume image and visualized following the standard volume rendering formulation. As the conversion of the sdf function is done using a **piecewise polynomial smooth step function**, there is **no loss of accuracy** during the conversion process. Experiments demonstrate that the proposed method has superior performance over the state-of-the-art neural scene representation methods. Particularly, our method has the ability of accurately reconstructing complex shapes with tiny holes or textureless region without mask supervision. In addition, our neural fuzzy geometric representation can be considered as a baseline and combined with geometry optimization. Compared with the framework and explanations provided in NeuS, **our solution is much more elegant and simpler**.

Generally, our model is limited to represent solid, non-transparent objects. One of our future works is to extend our current model for the representation of transparent objects. Our method is of high computational resource consumption for network training, like many other learning-based works. Thus, another future work we will focus on is the reduction of computational costs of our model by adopting certain accelerated algorithms, such as instant NGP [71].

REFERENCES

[1] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, and Y. Sheikh, "Neural volumes: Learning dynamic renderable volumes from images," *arXiv preprint arXiv:1906.07751*, 2019.

[2] S. Kaza *et al.*, "Differentiable volume rendering using signed distance functions," Ph.D. dissertation, Massachusetts Institute of Technology, 2019.

[3] 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," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[4] L. Liu, J. Gu, K. Zaw Lin, T.-S. Chua, and C. Theobalt, "Neural sparse voxel fields," *Advances in Neural Information Processing Systems*, vol. 33, pp. 15 651–15 663, 2020.

[5] S. Saito, T. Simon, J. Saragih, and H. Joo, "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 84–93.

[6] A. Trevithick and B. Yang, "Grf: Learning a general radiance field for 3d representation and rendering," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 182–15 192.

[7] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa, "Plenoxels: Radiance fields without neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5501–5510.

[8] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, and J. T. Barron, "Nerf in the dark: High dynamic range view synthesis from noisy raw images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 190–16 199.

[9] X. Huang, Q. Zhang, Y. Feng, H. Li, X. Wang, and Q. Wang, "Hdr-nerf: High dynamic range neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 398–18 408.

[10] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5470–5479.

[11] ——, "Zip-nerf: Anti-aliased grid-based neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19 697–19 705.

[12] K. Genova, F. Cole, D. Vlasic, A. Sarna, W. T. Freeman, and T. Funkhouser, "Learning shape templates with structured implicit functions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7154–7164.

[13] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3d reconstruction in function space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4460–4470.

[14] M. Michalkiewicz, J. K. Pontes, D. Jack, M. Baktashmotlagh, and A. Eriksson, "Implicit surface representations as layers in neural networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4743–4752.

[15] M. Niemeyer, L. Mescheder, M. Oechsle, and A. Geiger, "Occupancy flow: 4d reconstruction by learning particle dynamics," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5379–5389.

[16] 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*, 2019, pp. 165–174.

[17] S. Peng, M. Niemeyer, L. Mescheder, M. Pollefeys, and A. Geiger, "Convolutional occupancy networks," in *European Conference on Computer Vision*. Springer, 2020, pp. 523–540.

[18] K. Li, Y. Tang, V. A. Prisacariu, and P. H. Torr, "Bnv-fusion: Dense 3d reconstruction using bi-level neural volume fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6166–6175.

[19] J. Liu, P. Ji, N. Bansal, C. Cai, Q. Yan, X. Huang, and Y. Xu, "Planemvs: 3d plane reconstruction from multi-view stereo," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8665–8675.

[20] W. Lin, C. Zheng, J.-H. Yong, and F. Xu, "Occlusionfusion: Occlusion-aware motion estimation for real-time dynamic 3d reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1736–1745.

[21] A.-D. Nguyen, S. Choi, W. Kim, J. Kim, H. Oh, J. Kang, and S. Lee, "Single-image 3-d reconstruction: Rethinking point cloud deformation," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022.

[22] X. Long, C. Lin, P. Wang, T. Komura, and W. Wang, "Sparseneus: Fast generalizable neural surface reconstruction from sparse views," in *Computer Vision – ECCV 2022*, 2022, pp. 210–227.

[23] Z. Zhou and S. Tulsiani, "Sparsefusion: Distilling view-conditioned diffusion for 3d reconstruction," in *CVPR*, 2023.

[24] L. Melas-Kyriazi, C. Rupprecht, I. Laina, and A. Vedaldi, "Realfusion: 360° reconstruction of any object from a single image," in *Arxiv*, 2023.

[25] M. Liu, C. Xu, H. Jin, L. Chen, M. V. T, Z. Xu, and H. Su, "One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization," 2023.

[26] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 8456–8465.

[27] J. Tang, H. Zhou, X. Chen, T. Hu, E. Ding, J. Wang, and G. Zeng, "Delicate textured mesh recovery from nerf via adaptive surface refinement,"

in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17 739–17 749.

[28] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," *NeurIPS*, 2021.

[29] M. Oechsle, S. Peng, and A. Geiger, "Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5589–5599.

[30] L. Yariv, J. Gu, Y. Kasten, and Y. Lipman, "Volume rendering of neural implicit surfaces," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4805–4815, 2021.

[31] M. Agrawal and L. S. Davis, "A probabilistic framework for surface reconstruction from multiple images," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 2. IEEE, 2001, pp. II–II.

[32] J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, "Pixelwise view selection for unstructured multi-view stereo," in *European conference on computer vision*. Springer, 2016, pp. 501–518.

[33] S. Galliani, K. Lasinger, and K. Schindler, "Massively parallel multiview stereopsis by surface normal diffusion," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 873–881.

[34] J. S. De Bonet and P. Viola, "Poxels: Probabilistic voxelized volume reconstruction," in *Proceedings of International Conference on Computer Vision (ICCV)*, vol. 2, 1999.

[35] A. Broadhurst, T. W. Drummond, and R. Cipolla, "A probabilistic framework for space carving," in *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, vol. 1. IEEE, 2001, pp. 388–393.

[36] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 151–173, 1999.

[37] Z. Xie, X. Yu, X. Gao, K. Li, and S. Shen, "Recent advances in conventional and deep learning-based depth completion: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21, 2022.

[38] W. Hartmann, S. Galliani, M. Havlena, L. Van Gool, and K. Schindler, "Learned multi-patch similarity," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1586–1594.

[39] V. Leroy, J.-S. Franco, and E. Boyer, "Shape reconstruction using volume sweeping and learned photoconsistency," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 781–796.

[40] B. Ummenhofer, H. Zhou, J. Uhrig, N. Mayer, E. Ilg, A. Dosovitskiy, and T. Brox, "Demon: Depth and motion network for learning monocular stereo," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5038–5047.

[41] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4353–4361.

[42] S. Donne and A. Geiger, "Learning non-volumetric depth fusion using successive reprojections," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7634–7643.

[43] G. Riegler, A. O. Ulusoy, H. Bischof, and A. Geiger, "Octnetfusion: Learning depth fusion from data," in *2017 International Conference on 3D Vision (3DV)*. IEEE, 2017, pp. 57–66.

[44] P.-H. Huang, K. Matzen, J. Kopf, N. Ahuja, and J.-B. Huang, "Deepmvs: Learning multi-view stereopsis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2821–2830.

[45] Y. Yao, Z. Luo, S. Li, T. Fang, and L. Quan, "Mvsnet: Depth inference for unstructured multi-view stereo," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 767–783.

[46] Y. Yao, Z. Luo, S. Li, T. Shen, T. Fang, and L. Quan, "Recurrent mvsnet for high-resolution multi-view stereo depth inference," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5525–5534.

[47] V. Sitzmann, M. Zollhöfer, and G. Wetzstein, "Scene representation networks: Continuous 3d-structure-aware neural scene representations," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[48] V. Sitzmann, J. Thies, F. Heide, M. Nießner, G. Wetzstein, and M. Zollhofer, "Deepvoxels: Learning persistent 3d feature embeddings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2437–2446.

[49] M. Atzmon, N. Haim, L. Yariv, O. Israelov, H. Maron, and Y. Lipman, "Controlling neural level sets," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[50] L. Yariv, Y. Kasten, D. Moran, M. Galun, M. Atzmon, B. Ronen, and Y. Lipman, "Multiview neural surface reconstruction by disentangling geometry and appearance," *Advances in Neural Information Processing Systems*, vol. 33, pp. 2492–2502, 2020.

[51] M. Niemeyer, L. Mescheder, M. Oechsle, and A. Geiger, "Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3504–3515.

[52] M. Boss, R. Braun, V. Jampani, J. T. Barron, C. Liu, and H. Lensch, "Nerd: Neural reflectance decomposition from image collections," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 684–12 694.

[53] R. Martin-Brualla, N. Radwan, M. S. 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*, 2021, pp. 7210–7219.

[54] M. Niemeyer and A. Geiger, "Giraffe: Representing scenes as compositional generative neural feature fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 453–11 464.

[55] S. Peng, Y. Zhang, Y. Xu, Q. Wang, Q. Shuai, H. Bao, and X. Zhou, "Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9054–9063.

[56] 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*, 2021, pp. 10 318–10 327.

[57] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger, "Graf: Generative radiance fields for 3d-aware image synthesis," *Advances in Neural Information Processing Systems*, vol. 33, pp. 20 154–20 166, 2020.

[58] P. P. Srinivasan, B. Deng, X. Zhang, M. Tancik, B. Mildenhall, and J. T. Barron, "Nerv: Neural reflectance and visibility fields for relighting and view synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7495–7504.

[59] K. Zhang, G. Riegler, N. Snavely, and V. Koltun, "Nerf++: Analyzing and improving neural radiance fields," *arXiv preprint arXiv:2010.07492*, 2020.

[60] Q. Fu, Q. Xu, Y.-S. Ong, and W. Tao, "Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction," *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[61] Z. Yu, S. Peng, M. Niemeyer, T. Sattler, and A. Geiger, "Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction," *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[62] Q. Li and J. Tian, "Blending implicit shapes using fuzzy set operations," *WSEAS Trans. Info. Sci. and App.*, vol. 5, no. 7, p. 1230–1240, jul 2008.

[63] N. Ravi, J. Reizenstein, D. Novotny, T. Gordon, W.-Y. Lo, J. Johnson, and G. Gkioxari, "Accelerating 3d deep learning with pytorch3d," *arXiv preprint arXiv:2007.08501*, 2020.

[64] Q. Li, "Smooth piecewise polynomial blending operations for implicit shapes," in *Computer Graphics Forum*, vol. 26, no. 2. Wiley Online Library, 2007, pp. 157–171.

[65] J. T. Kajiya and B. P. Von Herzen, "Ray tracing volume densities," *ACM SIGGRAPH computer graphics*, vol. 18, no. 3, pp. 165–174, 1984.

[66] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. Lipman, "Implicit geometric regularization for learning shapes," *arXiv preprint arXiv:2002.10099*, 2020.

[67] Y. Yao, Z. Luo, S. Li, J. Zhang, Y. Ren, L. Zhou, T. Fang, and L. Quan, "Blendedmvs: A large-scale dataset for generalized multi-view stereo networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1790–1799.

[68] Y. Wang, I. Skorokhodov, and P. Wonka, "Hf-neus: Improved surface reconstruction using high-frequency details," *Advances in Neural Information Processing Systems*, vol. 35, pp. 1966–1978, 2022.

[69] T. Wu, J. Wang, X. Pan, X. Xu, C. Theobalt, Z. Liu, and D. Lin, "Voxurf: Voxel-based efficient and accurate neural surface reconstruction," in *International Conference on Learning Representations (ICLR)*, 2023.

[70] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao, "2d gaussian splatting for geometrically accurate radiance fields," in *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery, 2024.

[71] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *arXiv preprint arXiv:2201.05989*, 2022.