APSIPA Transactions on Signal and Information Processing, 2023, 12, e6

Original Paper

Deep Review and Analysis of Recent NeRFs

Fang Zhu[1,2*], Shuai Guo[3], Li Song[3,4], Ke Xu[1,2] and Jiayu Hu[5]

[1] *State Key Laboratory of Mobile Network and Mobile Multimedia Technology, China*
[2] *Sanechips Technology Co., LTD, China*
[3] *Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, China*
[4] *Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, China*
[5] *University of California, Los Angeles, CA, USA*

ABSTRACT

Neural radiance fields (NeRFs) refer to a suit of deep neural networks that are used to learn and represent objects or scenes. Generally speaking, NeRFs have five main characters: volumetric rendering, novel view synthesis, factorizable embedded space, multi-view consistency and weighted importance sampling. Recently, NeRFs have drawn great attention and are now important cornerstones of metaverse and augmented reality research, as is their stronger efficiency and more imaginative rendering performance. There have been many reviews of NeRFs, most of them focus on different applications of NeRFs. In this paper, we provide a deep review and analysis of recent NeRF related works, according to the main characters of NeRFs they make further progress in. Then we introduce some new application innovations of NeRFs, and illustrate future opportunities of them. We hope this paper can provide an insightful organization of current developments in NeRFs, identify their limitations, and give suggestions for further research.

*Corresponding author: Fang Zhu, zhu.fang@sanechips.com.cn. This work was supported by National Key R&D Project of China (2021YFF0900500).

Received 06 September 2022; Revised 07 December 2022
ISSN 2048-7703; DOI 10.1561/116.00000162

Keywords: NeRF, review, volumetric rendering, factorizable embedding, future innovations.

1 Introduction

3D scene modeling and prior-based rendering are important directions in augmented reality, meta-universe and controllable digital twin research. The first try of a 2D view mapping recording of a 3D world was recorded in the 1920s. John Logie Baird, one of the TV pioneers, demonstrated the idea of high-quality 3D television [18, 19, 52]. However, the long-running practice of modeling and physical rendering in computer graphics (CG) scenes is much more mature for such purposes. The beneficial experience stemmed from mesh exercises inside the CG domain includes: (i) high quality of rendering result when the 3D objects have 3D transformations; (ii) low storage consumption of 3D objects and scenes; (iii) controllability of rendering results of 3D scenes based on the Cartesian Coordinates.

However, mesh modeling of 3D scenes in the CG domain is really man-made controllable representations, which are based on the rendering pipeline with specific features such as light, material, occupancy, and so on. The main bottleneck is content creation, i.e., a vast amount of expensive manual work by skilled artists is required for the creation of the underlying scene representations in terms of surface geometry, appearance, light sources, and animations.

Concurrently, powerful spatial representation related perception methods have emerged in the computer vision (CV) and machine learning (ML) communities. The seminal work on multi-view stereo (MVS) by Michael *et al.* [4] has evolved in recent years into paradigms of the creation of high-resolution 3D models of natural scenes, such as structure from motion(SfM) and simultaneous localization and mapping(SLAM). Furthermore, in order to obtain better space occupancy representation, including the continuity of representation quantity and multi-scale self-adaptation, implicit representation technology represented by implicit surface is gradually attracting more attention from [16, 76]. In particular, implicit representation based on ML has become a hot topic of current research and has been widely discussed, such as in these works [48, 66, 73]. However, although high-quality spatial modeling of 3D scenes can be obtained under such a paradigm, explicit reconstruction of scene properties is still hard and error-prone and usually leads to artifacts in the rendered content.

Very recently, the two areas have come together and have been explored under the topic of neural radiance fields (NeRF) [55]. NeRF brings the promise of addressing both reconstruction and rendering by using deep networks to learn complex mappings from captured images to novel image synthesis.

Furthermore, NeRF combines physical knowledge, e.g., mathematical models of projection and imaging, with learned components as scene representations, to yield new and powerful algorithms for controllable content rendering.

Since its first introduction in 2020 [55], NeRFs have received wide attention. Many scholars have carried out deep research and extension around it. Recently, many related high-level papers have been presented.

There have been many reviews of NeRFs, most of them focus on different applications of NeRFs. Tewari *et al.* [85] review the recent trends on neural rendering techniques, and discuss the specific applications of neural rendering and the underlying neural scene representations. Xie *et al.* [99] focus on neural field techniques and applications of neural fields to different problems (e.g., visual computing, robotics, audio). Gao *et al.* [21] provide an introduction to the theory of NeRF based novel view synthesis, and a benchmark comparison of the performance and speed of key NeRFs. In this paper, we try to distill the principles of NeRF and review the different breakthroughs based on principles of NeRFs. The central theme around which we structured this paper is the five characters of NeRF: volumetric rendering, novel view synthesis, factorizable embedded space, multi-view consistency and weighted importance sampling. Based on this, we first review the original NeRF and distill the principles of NeRF. Then we use five sections to introduce the recent outstanding works in NeRF research, according to the main character they make further progress in, as shown in Table 1. Some of the recent works that make further progress in characters and application innovations of NeRFs are shown in Figure 1. We think this is the main difference between our paper and other NeRF reviews. After that, we introduce new application innovations of NeRFs. Finally, we try to summarize the future development direction of NeRF research.

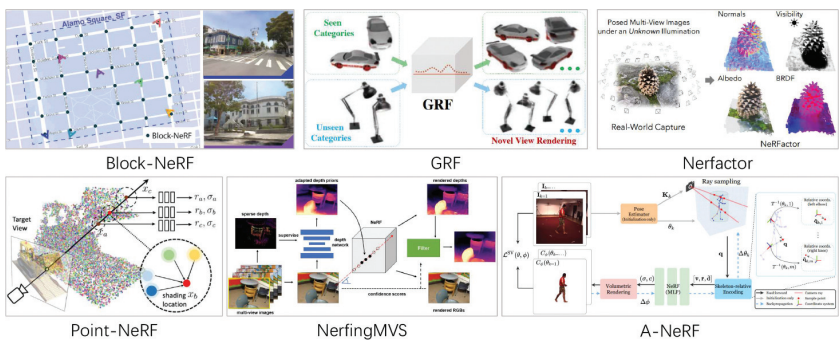

Figure 1: Overviews of some recent works that make further progress in characters and application innovations of NeRFs, including Block-NeRF [83], GRF [86], Nerfactor [112], Point-NeRF [102], NerfingMVS [94] and A-NeRF [79].

Table 1: NeRF's characters and breakthroughs.

Characters	Breakthroughs
Volumetric rendering	Faster training NeRFs Faster inference NeRFs Sparse NeRFs Depth-supervised NeRFs NeRFs for big scene rendering
Novel view synthesis	NeRFs with novel view synthesis paradigms
Factorizable embedded space	Relightable NeRFs Deformable NeRFs NeRFs for Scene Editing
Multi-view consistent	NeRFs that further explore multi-view consistency
Weighted importance sampling	NeRFs that have novel sampling methods

2 Neural Radiance Fields

In this section we review the initial release version of NeRF [55]. A NeRF encodes a static scene θ as a continuous volumetric radiance field g_θ of color c and density σ. Specifically, for a 3D point x and viewing direction unit vector d, as illustrated in (1), g_θ is implicitly expressed by a MLP.

$$(\delta, c) = g_\theta(x, d). \tag{1}$$

In such neural implicit functions, the embedded algorithm reflects the color computing, with respect to local structure, material, and lighting, and works directly in the representation space. The expansion of such a definition is in stark contrast to the implicit surface technology.

The scene representation by NeRF is optimized through a differentiable rendering loss to reproduce the appearance of a set of input images from known camera poses, as the loss given in (2). Figure 2 is an overview of neural radiance field scene representation and differentiable rendering procedure. The original nerf is developed under the assumption of an emit-absorb model and treats every sample location as a light source, so that there is no need for explicit modeling of geometry, material, lighting, and light transport. We can query the MLP for the volume density at densely-sampled points between the location and every light source to estimate the attenuation of light before it reaches that location. Such novel view synthesis pattern is employed, corresponding to the MAP hint.

$$L = \sum_{r \in \mathbb{R}} \|\widehat{C}(r) - C(r)\|_2^2. \tag{2}$$

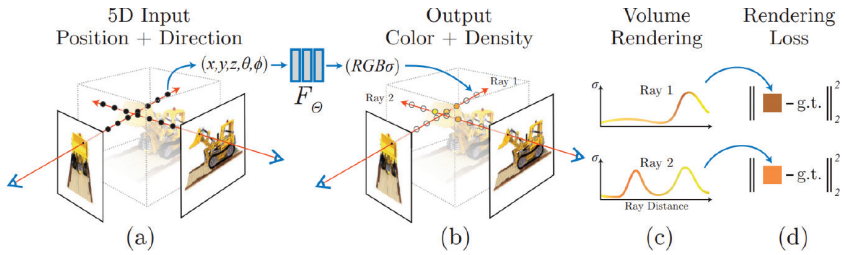

Figure 2: Overview of neural radiance field scene representation and differentiable rendering procedure. This picture is from [55].

In order to calculate the color of any ray passing through the scene, corresponding to a pixel in a target image, classical ray definition and volumetric rendering principles are used as in (3).

$$\begin{aligned} r(t) &= o + td \\ C(r) &= \int_{t_n}^{t_f} T(t)\delta(r(t)) \cdot c(r(t), d)dt \\ T(t) &= exp(-\int_{t_n}^{t_f} \sigma(r(s))ds). \end{aligned} \tag{3}$$

The procedure includes the density function(σ) at a point, the reflected radiance scattered(c)from a point in a direction, and the transmittance function (T). Transmittance function T also known as accumulated density, directly follows Beer's law, which is the classical solution that relates the attenuation of light to the properties of the space through which the light is traveling. The strict form of this process is continuous integral.

Such volumetric rendering procedures under Cartesian coordinates and embedded volumetric parameters in Euclidean space greatly aid the NeRF.

NeRF makes a great progress towards a graphics pipeline based on real world imagery, and can be directly used in novel viewpoint synthesis. However, the original NeRF has two major limitations: (1) since 3D content is encoded into the weights of an MLP, the trained network can only represent the learned structure, and is difficult to generalize across novel geometries; (2) the training process and inference process is very time-consuming. Many follow-up works try to address the limitations, introduce NeRFs that have better capacity, and adopt NeRFs to more applications.

We think the original NeRF has five main characters: volumetric rendering, novel view synthesis, factorizable embedded space, multi-view consistency and weighted importance sampling. In the following five sections, we review recent important NeRF-based works according to the character they make further progress in.

3 Volumetric Rendering

The volumetric rendering procedure under Cartesian coordinates is the most important character in the NeRF definition. Many related works make further progress in this character.

3.1 Faster Training NeRFs

How to effectively construct the reasonable and efficient calculation process of continuous points rendering is important to NeRFs' deployment. The gradual improvement of the calculation process of image pixels and spatial light, based on different scales or other factors, is the basis for promoting more accurate volumetric models and more fidelity rendering results. Since the initial design of NeRF, such a question has gained attention from researchers, a lot of effort has been put into these areas. The original NeRF is known by the need of too much training time and inefficiency in rendering new views. Many follow-up works have shown significant speed up in the training of NeRF. Yu *et al.* [106] render 800×800 images at more than 150 FPS, 3000 times faster than conventional NeRFs, by pre-tabulating the NeRF into a PlenOctree. They showed that PlenOctrees can also lead to equal or better quality. Other works include Sparse Neural Radiance Grid (SNeRG) [27], and the most recent and most famous breakthrough, Instant NGP [57]. The main contribution of Instant NeRF is the introduction of Multi-Resolution Hash Grid Coding, to help organize the volumetric rendering procedure. Based on such an input parameter feature space representation method, training completion time is reduced to seconds from previous hours.

Sun *et al.* [81] presented a super-fast convergence approach that directly optimizes the voxel grid, and reduce training time of NeRF from many hours to 15 minutes. The main contribution of Instant NGP [57], from Nivida, is the introduction of Multi-Resolution Hash Grid Coding, to help organize the volumetric rendering procedure, as shown in Figure 3. TensoRF [7] factorize the 4D scene tensor into multiple compact low-rank tensor components by applying traditional CP decomposition. As a result, TensoRF achieves fast reconstruction (<30 min) with better rendering quality and a smaller model size (<4 MB). Point-NeRF [102] combines the feature vector of 2D plane segments and the related point set from view depth fusion, to form the initial neural point cloud (each point has a space position, confidence, and reprojection of image features), with multi-view consistency. Then the neural point cloud helps to construct the NeRF's MLP based on the image feature vectors in the spatial point neighbors. Point-NeRF models a volumetric radiance field with a neural point cloud. This enables highly efficient reconstruction with only 20–40 min per-scene optimization, while original NeRF requires more than 20 hours.

Figure 3: With a realtime SLAM implementation estimates camera poses, Instant NGP can provide training and rendering live feedback. This picture is from [57].

Associated breakthroughs included a series of papers aimed at continuous innovation, from the original NeRF to Mip-NeRF [2] and the more recent Mip-NeRF-360 [3]. Under these continuous improvements, ray-color related spatial MLP outputs evolve from ray casting to linear cone casting and finally to no-linear cone casting. Different from such optimization in light tracking maturity, Shafiei *et al.* [75] proposed neural learning of the transmittance function. Great efficiency can be gained from such innovation, especially under complex environmental conditions.

3.2 Faster Inference NeRFs

Synthesizing high-resolution novel view from NeRF often requires time-consuming optical ray marching. There are many works that focus on acceleration of inference. NSVF [46] consists of a set of voxel-bounded implicit fields organized in a sparse voxel octree. NSVF is 10 times faster than the original NeRF [55] while achieving higher quality results. Lindell *et al.* [45] introduced automatic integration, a new framework that instantiate the computational graph for training and reassemble the graph to obtain a network. They improved render times by greater than 10× with a tradeoff of reduced image quality. Decomposed radiance fields [68] increase the inference efficiency of neural rendering via spatial Voronoi decomposition, which is compatible with the Painter's algorithm and makes inference pipeline GPUfriendly. KiloNeRF [69] demonstrate that real-time rendering is possible by utilizing thousands of tiny MLPs instead of one single large MLP. Each individual MLP only needs to represent parts of the scene. FastNeRF [23] that has a core of graphics-inspired factorization is the first NeRF-based system capable of rendering high fidelity photorealistic images at 200 Hz on a high-end consumer GPU. Mixture of Volumetric Primitives (MVP) [51] combines the completeness of volumetric representations with the efficiency of primitive-based rendering to achieve quality and runtime performance. Light Field Networks (LFNs) [77] require only a single network evaluation to render a ray, as it leverage meta-learning to learn a prior. LFNs [77] represent both geometry and appearance of the underlying 3D scene in a 360-degree, four-dimensional light field parameterized

via a neural network. LFNs just need to do single evaluation for each ray. This results in dramatic reductions in time and memory complexity, and enables real-time rendering.

3.3 Sparse NeRFs

Some researchers train or inference with their NeRFs using sparse input views or even a single view.

Instant Neural Radiance Fields (Instant NGP) [57] allow training from an incremental stream of images and camera poses. AutoRF [56] is a new approach for learning neural 3D object representations where each object in the training set is observed by only a single view. To address this challenging problem, AutoRf learns a normalized, object-centric representation whose embedding describes and disentangles shape, appearance, and pose. Light Field Neural Rendering [80] enforces geometric constraints during training and inference, the scene geometry is implicitly learned from a sparse set of views. Light Field Neural Rendering performs well on datasets that with larger margins on scenes with severe view-dependent variations. Neural Point Light Fields [61] encode a local light field on a point cloud by learning realistic radiance fields with only a single radiance sample per ray. Neural Point Light Fields are functions of the ray direction and local point feature neighborhood, which allows us to interpolate the light field conditioned training images without densely captured input views. Lin *et al.* [44] propose to leverage both the global and local features to form an expressive 3D representation. The global features are learned from a vision transformer, while the local features are extracted from a 2D convolutional network. They reduce the inputs to a single unposed image. Unlike traditional MPI that uses a set of simple RGBα planes in other NeRFs, NeX [95] propose a hybrid implicit-explicit modeling strategy. NeX models view-dependent effects by instead parameterizing each pixel as a linear combination of basis functions learned from a neural network. Törf [1] replace data-driven priors with measurements from a time-of-flight (ToF) camera. Törf improves novel-view synthesis for few-view scenes and especially for dynamic scenes. Among them, Törf works as not only the seminal breakthrough of incorporating active sensors within the NeRF theoretical framework, but also the active exploitation of multi-sensor fusion's advantage under NeRF. With updating corresponding parts in NeRF for the ToF camera and including the novel view synthesis optimization, some long-standing problems of the ToF's sensing results have been greatly improved, such as the false results exceeding an unambiguous range, resistance to sensor noise, and multiple single-scattering events along a ray. Furthermore, with the final collocated radiance fields, multi-sensor systems can capture scene geometry from a single view, allowing for higher-fidelity novel-view synthesis of dynamic scenes.

Other subsequent studies, such as the Sem2NeRF [10] and SinNeRF [100] methods, investigated the relationship between 2D semantics and NeRF construction in a single view input context. Chen *et al.* [10] introduced a new task, Semantic-to-NeRF translation (Sem2NeRF), that aims to reconstruct a 3D scene conditioned on one single-view semantic mask as input. Sem2NeRF addresses the task by encoding the semantic mask into the latent code that controls the 3D scene representation of a pre-trained decoder. Single View NeRF (SinNeRF) [100] uses only a single view as input, and propagate geometry pseudo labels and semantic pseudo labels to guide the progressive training process. At the same time, 2D-3D feature projection capabilities, such as the Features Line of Sight Projection (FLoSP) introduced in the recent work MonoScene [6], will strengthen the link between semantics and NeRF in a unified 3D pattern. MonoScene [6] proposes a 3D Semantic Scene Completion (SSC) framework, where the dense geometry and semantics of a scene are inferred from a single monocular RGB image.

3.4 Depth-supervised NeRFs

Concurrent to that, Roessle *et al.* [71] presented a method for novel view synthesis using neural radiance fields (NeRF) that leverages dense depth priors. Depth-supervised NeRF [14] add a loss to encourage the distribution of a ray's terminating depth matches a given 3D keypoint, incorporating depth uncertainty, to render better images with fewer training views.

3.5 NeRFs for Big Scene Rendering

NeRF laid the foundation of space-based content combination and space representation combination and decomposition. Related research works can be referred to Professor Yu's paper regarding controllable scene content composition [109], and also Google's latest breakthrough, Block-NeRF [83], regarding rendering city-scale scenes spanning multiple blocks. Professor Yu's paper [109] generates photo-realistic and editable free-viewpoint videos for dynamic scenes using a layered neural representation. Block-NeRF [83] decouples rendering time from scene size, and enables rendering to scale to arbitrarily large environments. In particular, KiloNeRF [69] discussed the feasibility of accelerating innovation by replacing the original MLP (NeRF space representation) with many micro MLPs in sub spaces. NeRF++ [111] analyzed NeRF's success in avoiding shape-radiance ambiguity, and works with 360 capture of large-scale unbounded 3D scenes. BungeeNeRF [97] performs NeRF in city-scale scene, with views ranging from an overview of a city to complex architectural details, as shown in Figure 4. To address this issue, BungeeNeRF fitted distant views with a shallow base block, new blocks are appended to accommodate the emerging details in the increasingly closer views.

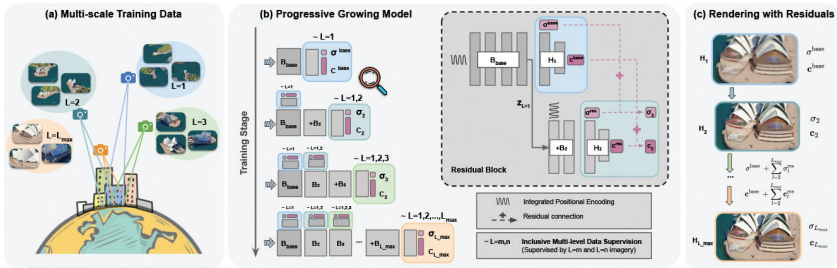

Figure 4: Overview of BungeeNeRF. This picture is from [97].

4 Novel View Synthesis

As mentioned in the last section, the neural representation, with the novel view synthesis training pattern, could improve the storage efficiency and get high-fidelity prediction through MAP. Since the original NeRF is only limited to static scenes, a lot of researchers' works extended the application scenarios to time interpolation, viewpoint interpolation, and mixed interpolation based on scenario video records with the help of the novel view synthesis paradigm. Such works include an early Meta's research work [96], and later Neural Scene Flow Fields (NSFF) [42].

The basic principle of such a transformation is straightforward: extend the original hidden space under 3D Cartesian coordinates in relation to a static scene to a 4D space-time irradiance field. However, since the input of the dynamic part should be carefully differentiated, the reconfiguration of the loss function gains quite some attention in the above studies, as cited in the papers.

Original NeRF can only inference novel view of training objects and scenes. Many researchers worked to extend NeRF to unseen objects and scenes. General Radiance Field (GRF) [86] learn local features for each pixel in 2D images and project these features to 3D points. Experiments demonstrate that GRF can generate high-quality and realistic novel views for novel objects. Tancik *et al.* [84] showed that simply modifying a coordinatebased neural representation's initial weight values can result in better generalization when only partial observations of a given signal are available. Wang *et al.* [91] introduced image-based rendering network (IBRNet) that includes a multilayer perceptron and a ray transformer that estimates radiance and volume density at continuous 5D locations. IBRNet outperforms recent novel view synthesis methods that also seek to generalize to novel scenes. Shape-conditioned Radiance Fields (ShaRF) [70] builds a geometric scaffold for an object and then uses this for estimating the radiance field. ShaRF is able to generalize to images outside of the training domain. CodeNeRF [30] learns to disentangle shape and texture by learning separate embeddings. Unseen objects can be reconstructed

from a single image, and then rendered from new viewpoints or their shape and texture edited by varying the latent codes. Neuray [50] constructs the radiance field that focus on visible image features to improve rendering quality, and uses a consistency loss to refine the visibility when finetuning on a specific scene.

5 Factorizable Embedded Space

Controllable rendering procedures are the ultimate goal of scene representation under interactive scenarios. And NeRF's MLP embeds the algorithm that reflects image computing in terms of local structure, material, and lighting, and operates directly in the representation space. Below, the three directions of controlling the embedded factors in the rendering procedure attract a lot of exploration.

5.1 Relightable NeRFs

The first is how to factorize the hidden space embedded in NeRF. Relightable NeRFs aim to get good performance with challenging input images. Recent NeRFactor [112], NeRV [78], and NeRD [5] work is noteworthy. They adopted a different rendering model, the absorb-reflect model, which requires explicit modeling of geometries (surface normal, lighting, material, and light transport). Among them, NeRFactor [112] factorsizes the appearance of a scene into 3D neural fields of surface normals, light visibility, albedo, and reflectance with an ingenious three-phases training design. Neural Reflectance and Visibility Fields (NeRV) [78] takes a set of images of a scene illuminated by unconstrained known lighting as input, and produces a 3D representation that can be rendered from novel viewpoints under arbitrary lighting conditions. NeRD [5] uses physically-based rendering to decompose the scene into spatially varying BRDF material properties. The input images can be captured under different illumination conditions. And NeRF-W uses the generative latent optimization framework (GLO) to optimize the appearance of each input image into the shared appearance embedding vector during the entire input photo data set. This decouples the external view of a photo from the illumination change and makes the training of scene representation very flexible and robust, even under the scenario of a changing illumination environment. Hidden variables in higher dimensions are for the purpose of strongly constraining the non-rigid deformation of an object under time-varying observation. NeRF for Outdoor Scene Relighting (NeRF-OSR) [72] is the first neural radiance fields approach for outdoor scene relighting.

Capturing the geometry and material properties of an object is essential for several computer vision and graphics applications. Feng *et al.* [20] created the first facial albedo evaluation benchmark TRUST where subjects are balanced in

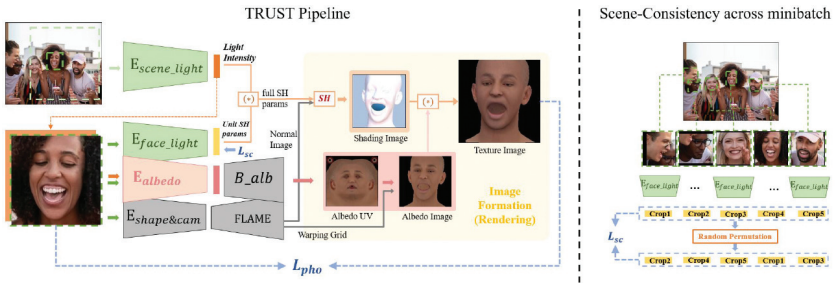

Figure 5: Overview of TRUST, which is built on the idea that the scene image can be exploited as a cue to disambiguate light and albedo, resulting in more accurate predictions. This picture is from [20].

terms of skin color, as shown in Figure 5. It's an initial step towards unbiased estimation of facial albedo from images in the wild. NeRF-OSR [72] showed that the second-order SH lighting model is capable of producing plausible relightings. NeRF-OSR allows simultaneous editing of illumination and camera viewpoint using only a collection of outdoor photos shot in uncontrolled settings.

5.2 Deformable NeRFs

And the second one is how to extend hidden space to higher dimensions for the purpose of more abundant meanings and expressions. Such research focused on deformable dynamic object modeling, which corresponds to the representation and construction of volume animation models and the synthesis of related free angles of view. Oritinal NeRF assume a static scene without moving objects. Many researchers relaxed this assumption and proposed deformable neural rendering system that is applicable to dynamic scenes. Hidden variables in higher dimensions are for the purpose of strongly constraining the non-rigid deformation of an object under time-varying observation. Related works in the recent literature cover Hyper-NeRF [64], D-NeRF [67], and HumanNeRF [113]. Dynamic NeRF (D-NeRF) [67] is the first end-to-end work on deformable NeRF, its key idea is to decompose learning in two modules. Figure 6 is an overview of D-NeRF. The first model learns a spatial mapping between each point of the scene at time t and a canonical scene configuration. The second module regresses the scene radiance emitted in each direction and volume density given the tuple.

Park *et al.* introduced their deformable NeRFs [63] that models non-rigidly deforming scenes. Their key to obtain high quality results is the as-rigid-as-possible deformation prior, and coarse-to-fine deformation regularization. The basic principle of such a transformation is straightforward: extend the original hidden space under 3D Cartesian coordinates in relation to a static

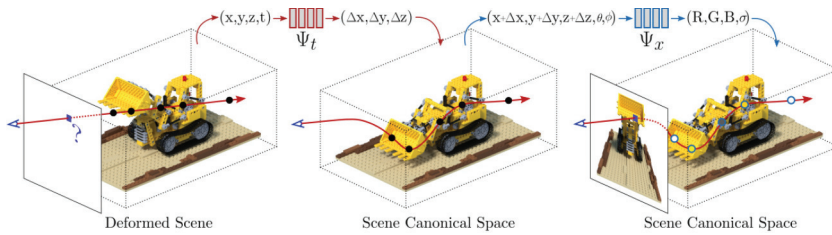

Figure 6: D-NeRF uses a deformation network to map scene deformations to, and uses a canonical network to regress volume density and color. This picture is from [67].

scene to a 4D space-time irradiance field. However, since the input of the dynamic part should be carefully differentiated, the reconfiguration of the loss function gains quite some attention in the above studies, as cited in the papers. RawNeRF [54] can reconstruct scenes from extremely noisy images captured in near-darkness. RawNeRF represents a step toward robust, high quality capture of real world environments. By the way, similarly, RAWNeRF incorporated the optical camera imaging model into the principal procedure and used image raw data (noisy linear raw images) as input. As an interesting result, the freely novel view synthesis can include controllable factors like exposure, tonemapping, and focus.

Recently, some human related NeRF methods are also proposed. H-NeRF [101] is a novel neural network that exploits the power of volumetric radiance fields to learn complex human structure and appearance, by relying on statistical implicit human pose and shape signed distance functions for accurate geometric reconstruction. H-NeRF can be used for the photo-realistic rendering and the temporal reconstruction of humans in motion. Neural Human Performer [36] combines a temporal and a multi-view Transformer that integrates multi-time and multi-view observations, and is constructed based on a parametric body model that can synthesize free-viewpoint videos for arbitrary human performers from sparse camera views. MPS-NeRF [22] is a novel-view and novel-pose human synthesis approach which can be generalizable for unseen persons with sparse multiview images as input. Its main innovation is to leverage a canonical-space NeRF and a volume deformation scheme derived by human body parametric model to achieve better generalizability.

Human related NeRFs are important in NeRF research. Neural Actor (NA) [47] is a new method for high-quality synthesis of humans from arbitrary viewpoints and under arbitrary controllable poses. NA utilizes a coarse parametric body model as a 3D proxy to unwarp the 3D space surrounding the posed body mesh into a canonical pose space. Then NA constructs a neural radiance field in the canonical pose space to learn pose-induced geometric deformations as well as both pose-induced and view-induced appearance effects. Animatable Neural Radiance Fields [65] introduce neural blend weight

fields to address the challenge of reconstructing an animatable human model from a multi-view video. Chen *et al.* [9] introduce their animatable NeRF for detailed human avatar creation from monocular videos. Their approach extends NeRF to the dynamic scenes with human movements via introducing explicit pose-guided deformation while learning the scene representation network. HumanNeRF [113] enables pausing the video at any frame and rendering the subject from arbitrary new camera viewpoints or even a full 360-degree camera path for that particular frame and body pose. Neural Novel Actor can learn a generalized animatable neural human representation from a sparse set of multi-view imagery of multiple persons. Neural Novel Actor [92] designs human representation with disentangled geometry and appearance, and leverage the features at both the spatial points and the surface points of SMPL to infer pose- and person-dependent geometry and appearance. HDHumans [26] is the first method for HD human character synthesis that jointly produces an accurate and temporally coherent 3D deforming surface and highly photo-realistic images of arbitrary novel views and of motions not seen at training time. HDHumans use deforming character template to guide NeRF, and leverage the dense point clouds resulting from NeRF to further improve the deforming surface via 3D-to-3D supervision.

5.3 NeRFs for Scene Editing

The third one is how to control the hidden space for definitive goals. Classical works like EditNeRF [49], Neural Scene Graphs [62], and Compositional Generative Neural Feature Fields (GIRAFFE) [59] are examples. Neural Scene Graphs [62] is the first approach that tackles the challenge of representing dynamic, multi-object scenes. Neural Scene Graphs encodes object transformations and radiance, then learn implicitly encoded scenes. EditNeRF [49] enables users to edit colors and shapes of objects, using a disentangled representation. All of these papers build on the early pioneering work of Generative Radiance Fields (GRAF) [74], which pioneered the integration of Generative Adversarial Networks (GAN) and NeRF in training and volumetric rendering for controllable high-resolution image synthesis.

FENeRF outperforms state-of-the-art methods in various face editing tasks. Instead of encoding the entire scene as a whole, Yang *et al.* [103] presented a novel neural scene rendering system to produce realistic rendering with editing capability for a clustered and real-world scene, as shown in Figure 7. Face Editing in Neural Radiance Fields (FENeRF) [82], which not only generated consistent views and locally edited images but also improved image fidelity. Face Editing in Neural Radiance Fields [82] is the first locally editable 3D-aware face generator FENeRF based on implicit scene representation. FENeRF can jointly render the boundary-aligned image and semantic mask and use the semantic mask to edit the 3D volume via GAN inversion. Semantic labeling is

Figure 7: Layered neural representation introduced by Yang *et al.* [103] can generate photo-realistic and editable free-viewpoint videos for dynamic scenes. The left two columns are two rendering results of different viewpoints without editing, the middle column is the edited results in a novel viewpoint, and the right column is the corresponding 3D illustration. This picture is from [103].

highly correlated with geometry and radiance reconstruction, as first argued in the work Semantic-NeRF [114]. Zhi *et al.* [114] extend NeRF to jointly encode semantics with appearance and geometry. They demonstrate its advantageous properties in efficient scene labelling tool, novel semantic view synthesis and many other applications.

6 Multi-view Consistent

Multi-view consistency is the basement of MVS. The consistent information of multiple views can be used explicitly to improve the robustness and accuracy of NeRF training. The typical research includes MVSNeRF [8] and NerfingMVS [94]. MVSNeRF [8] enables high quality radiance field reconstruction from only three input views and can achieve realistic view synthesis results from the reconstruction. MVSNeRF is a generalizable radiance field that works well across diverse datasets. NerfingMVS [94] firstly adapts a monocular depth network over the target scene by finetuning on its sparse SfM+MVS reconstruction. Then utilize the learning-based priors to guide the optimization process of NeRF to get depth images. And a per-pixel confidence map is used to further improve the depth quality. In comparison to these pixel-domain studies, DietNeRF [29] recently proposed the concept of high-level semantic consistency. With a single-view 2D representation as input, DietNeRF exploited this transferable prior knowledge to solve optimization issues and to cope with partial observability.

Furthermore, constructing a uniform geometry prior index between 3D features and 2D features, with multi-view consistency, is also an important di-

rection for improving the NeRF training procedure. Examples of related works include PixelNeRF [107], Point-NeRF [102], and SRF [11]. PixelNeRF [107] introduces an architecture that conditions a NeRF on image inputs in a fully convolutional manner. PixelNeRF [107] is trained across multiple scenes to learn a scene prior, then predict a continuous neural scene representation conditioned on one or few input images. This allows the network to be trained across multiple scenes to learn a scene prior, enabling it to perform novel view synthesis in a feed-forward manner from a sparse set of views (as few as one). Stereo Radiance Field (SRF) [11] is trained end-to-end and generalizes to new scenes. SRF project each 3D point to multiple views, extract features, and process them in pairs. SRF only need sparse views at test time.

7 Weighted Importance Sampling

Learning from the traditional Monte Carlo rendering experience, the adoption of weighted importance sampling (the sampled probability density function similar to the shape of the integral function) in the volumetric rendering procedure can reduce the sampling error and accelerate the convergence speed. With such a hint, related research, including DONeRF [58], DS-NeRF [14], and MINE [40], used the explicit depth information to improve the related training of NeRF and generation of visualized content. The typical one, DONeRF [58], significantly reduces the number of samples required for each view ray calculation in view rendering when samples are collected around the space surface of the scene. Inference costs can be reduced by up to 48x compared to the original NeRF design. Similar efficiency can also be seen in Point-NeRF [102], the method mentioned above, since the explicit surface indicator existed. Deng *et al.* [14] introduced Depth-supervised NeRF that takes advantage of depth supervision. This method trains 2–3 times faster and get better results from fewer training views, compared with previous works. MINE [40], used the explicit depth information to improve the related training of NeRF and generation of visualized content.

8 Application Innovations of NeRFs

The original NeRF can be used directly in novel view synthesis. In this section, we review the recent works that inspire new applications of NeRFs.

8.1 NeRFs for Pose Estimation

Inverting Neural Radiance Fields (iNeRF) [105] applies analysis by synthesis with NeRF for 6DoF pose estimation. Starting from an initial pose estimate,

the authors use gradient descent to minimize the residual between pixels rendered from an already-trained NeRF and pixels in an observed image. Articulated Neural Radiance Field (A-NeRF) [79] learns a generative neural body model from unlabelled monocular videos. A-NeRF could refine the initial 3D articulated skeleton pose estimate. NeRF– [93] showed that camera parameters can be jointly optimised through a photometric reconstruction. Bundle-Adjusting Neural Radiance Field (BARF) [43] was proposed for training NeRF from imperfect (or even unknown) camera poses. BARF can learn the 3D scene representations from scratch as well as resolve large camera pose misalignment. Self-Calibrating Neural Radiance Fields (SCNeRF) [31] is a camera self-calibration algorithm for generic cameras with arbitrary non-linear distortions. It can jointly learn the geometry of the scene and the accurate camera parameters without any calibration objects.

8.2 NeRFs for SLAM

Implicit Mapping and Positioning (iMAP) is the first work that poses dense SLAM as real-time continual learning. iMAP showed that an MLP can be trained from scratch as the only scene representation with a hand-held RGB-D camera. NICE-SLAM [115] combines neural implicit decoders with hierarchical grid-based representations. NICE-SLAM demonstrate that tiny MLPs + multi-res feature grids can guarantee fine-detailed mapping, high tracking accuracy, faster speed and much less computation.

8.3 NeRFs for Reconstruction

The typical research includes MVSNeRF [8] and NerfingMVS [94]. Oechsle *et al.* [60] thought that implicit surface models and radiance fields can be formulated by enabling both surface and volume rendering using the same model. Their model outperforms previous works in terms of reconstruction quality. Neural Implicit Surfaces (NeuS) [90] reconstructs objects and scenes with high fidelity from 2D image inputs. NeuS propose to represent a surface as the zero-level set of a signed distance function (SDF) and develop a new volume rendering method to train a neural SDF representation. Yariv *et al.* [104] thought that geometry extracted using an arbitrary level set of the density function could lead to low fidelity reconstruction. They modeled the volume density as a function of the geometry, thus improved geometry representation and reconstruction in neural volume rendering. NeuRIS [89] can build highquality reconstruction of indoor scenes. The key idea of NeuRIS is to integrate estimated normal of indoor scenes as a prior in a neural rendering framework, and reconstruct large texture-less shapes in an adaptive manner. MonoSDF [108] demonstrate that depth and normal cues predicted by general-

purpose monocular estimators can significantly improve reconstruction quality and optimization time.

8.4 NeRFs for Downstream Tasks

Distilled Feature Field (DFF) [33] distills the knowledge of off-the-shelf, supervised and self-supervised 2D image feature extractors such as CLIP-LSeg or DINO into a 3D feature field optimized in parallel to the radiance field. Given a user-specified query of various modalities such as text, an image patch, or a point-and-click selection, 3D feature fields semantically decompose 3D space without the need for re-training. Neural-Sim [24] is the first fully differentiable synthetic data pipeline that uses NeRFs in a closed-loop with a target application's loss function. And Neural-Sim is successfully used in real-world object detection tasks. GraspNeRF [13] is the first multiview RGB-based 6-DoF grasp detection network that leverages the generalizable NeRF to achieve material-agnostic object grasping in clutter. And GraspNeRF performs zero-shot NeRF construction with sparse RGB inputs and reliably detect 6-DoF grasps, both in realtime. Neural Semantic Fields (NeSFs) [88] is a novel method for simultaneous 3D scene reconstruction and semantic segmentation from posed 2D images. At inference time, NeSF construct a dense semantic segmentation field that can be queried directly in 3D or used to render 2D semantic maps from novel camera poses. Panoptic Neural Fields (PNF) [35] is an object aware neural scene representation that decomposes a scene into a set of objects (things) and background (stuff). Each object is represented by an oriented 3D bounding box and a multi-layer perceptron (MLP) that takes position, direction, and time and outputs density and radiance.

8.5 Generative NeRFs

Also, the work, GAN-based Neural Radiance Field without Posed Camera(GNeRF) [53], proved the resistance to noise and disturbance during the representation field construction by first being guided by high-level semantic consistence. 3D Generative Neural Radiance Field (GNeRF) models, which extract implicit 3D representations from 2D images, have recently been shown to produce realistic images representing rigid/semi-rigid objects, such as human faces or cars. NeRF-VAE [34] is a 3D scene generative model that incorporates geometric structure via Neural Radiance Fields (NeRF) and differentiable volume rendering. NeRF-VAE shares structure across scenes and is able to infer the structure of a novel scene using amortized inference. 3D-aware Semantic-Guided Generative model (3D-SGAN) [110] use a generative NeRF to implicitly represent the 3D human body, and render the 3D representation into 2D segmentation masks. And the masks are mapped into the final images using a VAE-conditioned texture generator. GIRAFFE [59] represents scenes

as compositional generative neural feature fields. The authors disentangle individual objects from the background as well as their shape and appearance to yield fast and controllable image synthesis. StyleNeRF [25] is a 3D-aware generative model for photo-realistic high resolution image synthesis with high multi-view consistency. StyleNeRF integrates NeRF into a style-based generator to tackle the aforementioned challenges, i.e., improving rendering efficiency and 3D consistency for high-resolution image generation. Generative radiance manifolds (Gram) [15] is proposed for 3d-aware image generation, by regulating point sampling and radiance learning on 2D manifolds for the radiance generator. Gram takes a large step towards generating 3D-aware virtual contents for real applications.

8.6 NeRFs for Robotics

Li *et al.* [41] proposed to learn viewpoint-invariant 3D-aware scene representations from visual observations using an autoencoding framework augmented with a neural radiance field rendering module and time contrastive learning. The learned 3D representations perform well on the model-based visuomotor control tasks. Ichnowski *et al.* [28] propose using NeRFs to detect, localize, and infer the geometry of transparent objects with sufficient accuracy to find and grasp them securely. They recover the geometry of transparent objects through a combination of additional lights and thresholding to find transparent points that are visible from some view directions. Lee *et al.* [37] leveraged NeRF based implicit representations to tackle active robotic 3D reconstruction of an object.

8.7 NeRFs for Medical Application

Medical Neural Radiance Fields (MedNeRF) [12] is proposed to learn reconstruct CT projections from a few or even a single-view X-ray. This model is trained on chest and knee datasets, and demonstrate qualitative and quantitative rendering results. Li *et al.* [39] apply the NeRF algorithm for the reconstruction of 3D US spine data and to evaluate the spinal curvature measurement from the reconstructed results. NeRFs are also used in Sparse-view computed tomography (CT) [32].

8.8 NeRFs for Reinforcement Learning

It is a long-standing problem to find effective representations for training reinforcement learning (RL) agents. Driess *et al.* [17] demonstrates that learning state representations with supervision from NeRFs can improve the performance of RL compared to other learned representations or even low-dimensional, hand-engineered state information. They propose to train an

encoder that maps multiple image observations to a latent space describing the objects in the scene.

8.9 NeRFs for Some Other Scientific Applications

Morphable Facial Neural Radiance Field (MoFaNeRF) [116] takes the coded facial shape, expression and appearance along with space coordinate and view direction as input, and can be used for photo-realistic image synthesis. MoFaNeRF is the first facial parametric model based on neural radiance field, and can make the face morphable in a large-scale solution space. Ref-NeRF [87] is introduced to accurately capture and reproduce the appearance of glossy surfaces. Ref-NeRF replaces NeRF's parameterization of view-dependent outgoing radiance with a representation of reflected radiance and structures this function using a collection of spatially-varying scene properties. Black Hole NeRF [38] is a novel tomography approach that leverages gravitational lensing to recover the continuous 3D emission field near a black hole. This work takes the first steps in showing how future measurements from the Event Horizon Telescope could be used to recover evolving 3D emission around the supermassive black hole in our Galactic center. Figure-Ground Neural Radiance Fields (FiG-NeRF) [98] can be used to separate foreground objects from their varying backgrounds, and modeling object categories in 3D whilst.

9 Future

After reviewing certain NeRFs' recent achievements, in this section, future opportunities with NeRF for the development of scene modeling and content rendering will be illustrated in detail. In particular, some critical innovation paradigms can be followed, like principled consistent framework redefinition and intentional embedding exploration, which will be deeply discussed.

9.1 Frame Redefinition

During the review in the previous section, most exercises of NeRFs use image sequences from passive sensors as input, although some active sensors can work as catalysts in the weighted importance sampling. Since multi-sensors are becoming more and more common in electronic systems nowadays, like the latest iPhone. If sensor enhancement and multi-sensor fusion can also be incorporated in the final context, the opportunity for NeRF to bridge the theoretical and technical domains will be greatly enhanced.

The key principal of NeRF's original definition is volumetric rendering, which strictly defines the procedure for creating an image by tracing rays through the volume and computing the radiance along each ray. In the

formula, two key sub-concepts reflect the radiance transmission and radiance generation procedures. So, strictly following the main volumetric rendering procedure and updating the related sub-procedure will beneficially extend NeRF's domain according to the classical radiance generation and transmission model gathered during long periods of sensors' development.

Furthermore, as previously stated, novel view synthesis is an important characteristic of NeRF, which optimizes the final results through the end-to-end optimization process with high fidelity input data included. All of the above principles, if followed during the framework redefinition, will eventually bridge the gap between rapidly emerging sensor-related technology and the most recent theoretical advancements in NeRFs.

Törf [1] and RAWNeRF [54] have recently demonstrated the possibilities of the above principal aligned framework definition. Among them, Törf works as not only the seminal breakthrough of incorporating active sensors within the NeRF theoretical framework, but also the active exploitation of multi-sensor fusion's advantage under NeRF.

With updating corresponding parts in NeRF for the ToF camera and including the novel view synthesis optimization, some long-standing problems of the ToF's sensing results have been greatly improved, such as the false results exceeding an unambiguous range, resistance to sensor noise, and multiple single-scattering events along a ray. Furthermore, with the final collocated radiance fields, multi-sensor systems can capture scene geometry from a single view, allowing for higher-fidelity novel-view synthesis of dynamic scenes.

By the way, similarly, RAWNeRF incorporated the optical camera imaging model into the principal procedure and used image raw data (noisy linear raw images) as input. As an interesting result, the freely novel view synthesis can include controllable factors like exposure, tonemapping, and focus.

9.2 Embedding Exploration

In addition to enlarging the coverage of NeRF's dominant domain with the above paradigm, the intentional embedding exploration can also seem to be a prospective development direction of NeRF.

First, semantic labeling is highly correlated with geometry and radiance reconstruction, as first argued in the work Semantic-NeRF [114]. Other subsequent studies, such as the Sem2NeRF [10] and SinNeRF [100] methods, investigated the relationship between 2D semantics and NeRF construction in a single view input context. At the same time, 2D-3D feature projection capabilities, such as the Features Line of Sight Projection (FLoSP) introduced in the recent work MonoScene [6], will strengthen the link between semantics and NeRF in a unified 3D pattern.

Secondly, because of the continuity of the underlying implicit representation in NeRF and also the virtues brought by the novel view synthesis pattern.

Semantic prediction combined with NeRF construction, like Semantic-NeRF, has the virtue of resulting in smooth, compact, continuous, and efficient de-noising semantic labels.

Thirdly, and most importantly, such an intention can greatly enhance the perception and understanding of the scenes with hierarchical spatial-semantic consistency for better 3D scene perception and more fidelity controllable rendering. Such an advantage can be referred to in the latest work, Face Editing in Neural Radiance Fields (FENeRF) [82], which not only generated consistent views and locally edited images but also improved image fidelity. Also, the work, GAN-based Neural Radiance Field without Posed Camera(GNeRF) [53], proved the resistance to noise and disturbance during the representation field construction by first being guided by high-level semantic consistence.

9.3 Problems to Overcome

Although NeRFs have made great progress in recent years, there exist two main problems to overcome. First, rendering results of NeRFs don't have high-enough quality, especially for input images with high resolution and rich details. Second, although there have been many works that accelerate training and inference of NeRFs, time consumption of NeRFs are still too high. This means NeRFs need too much time to adapt to or train on new scenes, and we can not get real-time rendering result with NeRFs. Other problems of NeRFs include scalability, generalizability, modeling of refractive objects, and so on.

10 Conclusion

NeRFs have raised a lot of interest in the past few years. This state-of-the-art report reflects the immense increase in research of this field. It spans a variety of use-cases that range from representation construction based on dynamic sequence input, time-spatial up-sampling based rendering, factorization for hidden space, and controllable scene composition and decomposition. NeRFs have already enabled applications that were previously intractable, especially the 6 DoF plus interactivity for immersive purposes, such as the rendering of digital avatars without any manual modeling. We believe that NeRF will have a profound impact in making complex multimedia tasks and building bonds between the digital and real worlds accessible to a much larger audience with the help of revealing the critical principal factors behind through thoroughly distillation in this report.

References

[1] B. Attal, E. Laidlaw, A. Gokaslan, C. Kim, C. Richardt, J. Tompkin, and M. O'Toole, "Törf: Time-of-flight radiance fields for dynamic scene view synthesis," *Advances in Neural Information Processing Systems*, 34, 2021, 26289–301.

[2] 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," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 5855–64.

[3] 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, 5470–9.

[4] M. Bleyer, C. Rhemann, and C. Rother, "PatchMatch stereo – Stereo matching with slanted support windows," in *British Machine Vision Conference 2011*, 2011.

[5] 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, 12684–94.

[6] A.-Q. Cao and R. de Charette, "MonoScene: Monocular 3D semantic scene completion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 3991–4001.

[7] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "TensoRF: Tensorial radiance fields," *arXiv preprint arXiv:2203.09517*, 2022.

[8] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, "Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 14124–33.

[9] J. Chen, Y. Zhang, D. Kang, X. Zhe, L. Bao, X. Jia, and H. Lu, "Animatable neural radiance fields from monocular RGB videos," *arXiv preprint arXiv:2106.13629*, 2021.

[10] Y. Chen, Q. Wu, C. Zheng, T. J. Cham, and J. Cai, "Sem2NeRF: Converting single-view semantic masks to neural radiance fields," 2022.

[11] J. Chibane, A. Bansal, V. Lazova, and G. Pons-Moll, "Stereo radiance fields (SRF): Learning view synthesis for sparse views of novel scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 7911–20.

[12] A. Corona-Figueroa, J. Frawley, S. Bond-Taylor, S. Bethapudi, H. P. Shum, and C. G. Willcocks, "MedNeRF: Medical neural radiance fields for reconstructing 3D-aware CT-Projections from a single X-ray," *arXiv preprint arXiv:2202.01020*, 2022.

[13] Q. Dai, Y. Zhu, Y. Geng, C. Ruan, J. Zhang, and H. Wang, "GraspNeRF: Multiview-based 6-DoF grasp detection for transparent and specular objects using generalizable NeRF," *arXiv preprint arXiv:2210.06575*, 2022.

[14] K. Deng, A. Liu, J.-Y. Zhu, and D. Ramanan, "Depth-supervised nerf: Fewer views and faster training for free," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 12882–91.

[15] Y. Deng, J. Yang, J. Xiang, and X. Tong, "GRAM: Generative radiance manifolds for 3D-aware image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, 10673–83.

[16] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," in *2011 IEEE International Conference on Robotics and Automation*, 2011.

[17] D. Driess, I. Schubert, P. Florence, Y. Li, and M. Toussaint, "Reinforcement learning with neural radiance fields," *arXiv preprint arXiv:2206.01634*, 2022.

[18] C. Fehn, P. Kauff, M. O. De Beeck, F. Ernst, W. Ijsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, and I. Sexton, "An evolutionary and optimised approach on 3D-TV," in *Proc. of IBC*, Vol. 2, 2002, 357–65.

[19] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Stereoscopic Displays and Virtual Reality Systems XI*, Vol. 5291, SPIE, 2004, 93–104.

[20] H. Feng, T. Bolkart, J. Tesch, M. J. Black, and V. Abrevaya, "Towards racially unbiased skin tone estimation via scene disambiguation," *arXiv preprint arXiv:2205.03962*, 2022.

[21] K. Gao, Y. Gao, H. He, D. Lu, L. Xu, and J. Li, "NeRF: Neural radiance field in 3D vision, a comprehensive review," *arXiv preprint arXiv:2210.00379*, 2022.

[22] X. Gao, J. Yang, J. Kim, S. Peng, Z. Liu, and X. Tong, "MPS-NeRF: Generalizable 3D human rendering from multiview images," *arXiv preprint arXiv:2203.16875*, 2022.

[23] S. J. Garbin, M. Kowalski, M. Johnson, J. Shotton, and J. Valentin, "FastNeRF: High-fidelity neural rendering at 200FPS," 2021.

[24] Y. Ge, H. Behl, J. Xu, S. Gunasekar, N. Joshi, Y. Song, X. Wang, L. Itti, and V. Vineet, "Neural-sim: Learning to generate training data with NeRF," *arXiv preprint arXiv:2207.11368*, 2022.

[25] J. Gu, L. Liu, P. Wang, and C. Theobalt, "Stylenerf: A style-based 3d-aware generator for high-resolution image synthesis," *arXiv preprint arXiv:2110.08985*, 2021.

[26] M. Habermann, L. Liu, W. Xu, G. Pons-Moll, M. Zollhoefer, and C. Theobalt, "HDHumans: A hybrid approach for high-fidelity digital humans," *arXiv preprint arXiv:2210.12003*, 2022.

[27] P. Hedman, P. P. Srinivasan, B. Mildenhall, J. T. Barron, and P. Debevec, "Baking neural radiance fields for real-time view synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 5875–84.

[28] J. Ichnowski, Y. Avigal, J. Kerr, and K. Goldberg, "Dex-NeRF: Using a neural radiance field to grasp transparent objects," *arXiv preprint arXiv:2110.14217*, 2021.

[29] A. Jain, M. Tancik, and P. Abbeel, "Putting NeRF on a diet: Semantically consistent few-shot view synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 5885–94.

[30] W. Jang and L. Agapito, "CodeNeRF: Disentangled neural radiance fields for object categories," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, 12949–58.

[31] Y. Jeong, S. Ahn, C. Choy, A. Anandkumar, M. Cho, and J. Park, "Self-calibrating neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 5846–54.

[32] B. Kim, H. Shim, and J. Baek, "A streak artifact reduction algorithm in sparse-view CT using a self-supervised neural representation," *Medical physics*.

[33] S. Kobayashi, E. Matsumoto, and V. Sitzmann, "Decomposing NeRF for editing via feature field distillation," *arXiv preprint arXiv:2205.15585*, 2022.

[34] A. R. Kosiorek, H. Strathmann, D. Zoran, P. Moreno, R. Schneider, S. Mokrá, and D. J. Rezende, "Nerf-vae: A geometry aware 3D scene generative model," in *International Conference on Machine Learning*, PMLR, 2021, 5742–52.

[35] A. Kundu, K. Genova, X. Yin, A. Fathi, C. Pantofaru, L. J. Guibas, A. Tagliasacchi, F. Dellaert, and T. Funkhouser, "Panoptic neural fields: A semantic object-aware neural scene representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, 12871–81.

[36] Y. Kwon, D. Kim, D. Ceylan, and H. Fuchs, "Neural human performer: Learning generalizable radiance fields for human performance rendering," *Advances in Neural Information Processing Systems*, 34, 2021, 24741–52.

[37] S. Lee, L. Chen, J. Wang, A. Liniger, S. Kumar, and F. Yu, "Uncertainty guided policy for active robotic 3D reconstruction using neural radiance fields," *IEEE Robotics and Automation Letters*, 2022.

[38] A. Levis, P. P. Srinivasan, A. A. Chael, R. Ng, and K. L. Bouman, "Gravitationally lensed black hole emission tomography," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, 19841–50.

[39] H. Li, H. Chen, W. Jing, Y. Li, and R. Zheng, "3D ultrasound spine imaging with application of neural radiance field method," in *2021 IEEE International Ultrasonics Symposium (IUS)*, 2021, 1–4, DOI: 10.1109/IUS52206.2021.9593917.

[40] J. Li, Z. Feng, Q. She, H. Ding, C. Wang, and G. H. Lee, "Mine: Towards continuous depth mpi with nerf for novel view synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 12578–88.

[41] Y. Li, S. Li, V. Sitzmann, P. Agrawal, and A. Torralba, "3d neural scene representations for visuomotor control," in *Conference on Robot Learning*, PMLR, 2022, 112–23.

[42] 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*, 2021, 6498–508.

[43] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, "Barf: Bundle-adjusting neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 5741–51.

[44] K.-E. Lin, L. Yen-Chen, W.-S. Lai, T.-Y. Lin, Y.-C. Shih, and R. Ramamoorthi, "Vision transformer for NeRF-based view synthesis from a single input image," *arXiv preprint arXiv:2207.05736*, 2022.

[45] D. B. Lindell, J. Martel, and G. Wetzstein, "AutoInt: Automatic integration for fast neural volume rendering," in, 2020.

[46] L. Liu, J. Gu, K. Zaw Lin, T.-S. Chua, and C. Theobalt, "Neural sparse voxel fields," *Advances in Neural Information Processing Systems*, 33, 2020, 15651–63.

[47] L. Liu, M. Habermann, V. Rudnev, K. Sarkar, J. Gu, and C. Theobalt, "Neural actor: Neural free-view synthesis of human actors with pose control," *ACM Transactions on Graphics (TOG)*, 40(6), 2021, 1–16.

[48] S.-L. Liu, H.-X. Guo, H. Pan, P.-S. Wang, X. Tong, and Y. Liu, "Deep implicit moving least-squares functions for 3D reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 1788–97.

[49] S. Liu, X. Zhang, Z. Zhang, R. Zhang, J.-Y. Zhu, and B. Russell, "Editing conditional radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 5773–83.

[50] Y. Liu, S. Peng, L. Liu, Q. Wang, P. Wang, C. Theobalt, X. Zhou, and W. Wang, "Neural rays for occlusion-aware image-based rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 7824–33.

[51] S. Lombardi, T. Simon, G. Schwartz, M. Zollhoefer, Y. Sheikh, and J. Saragih, "Mixture of volumetric primitives for efficient neural rendering," 2021.

[52] L. M. Meesters, W. A. IJsselsteijn, and P. J. Seuntiëns, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3), 2004, 381–91.

[53] Q. Meng, A. Chen, H. Luo, M. Wu, H. Su, L. Xu, X. He, and J. Yu, "Gnerf: Gan-based neural radiance field without posed camera," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 6351–61.

[54] 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, 16190–9.

[55] 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 *European Conference on Computer Vision*, Springer, 2020, 405–21.

[56] N. Müller, A. Simonelli, L. Porzi, S. R. Bulò, M. Nießner, and P. Kontschieder, "AutoRF: Learning 3D Object Radiance Fields from Single View Observations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 3971–80.

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

[58] T. Neff, P. Stadlbauer, M. Parger, A. Kurz, J. H. Mueller, C. R. A. Chaitanya, A. Kaplanyan, and M. Steinberger, "DONeRF: Towards real-time rendering of compact neural radiance fields using depth oracle networks," in *Computer Graphics Forum*, Vol. 40, No. 4, Wiley Online Library, 2021, 45–59.

[59] 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 (CVPR)*, June 2021, 11453–64.

[60] 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, 5589–99.

[61] J. Ost, I. Laradji, A. Newell, Y. Bahat, and F. Heide, "Neural point light fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 18419–29.

[62] J. Ost, F. Mannan, N. Thuerey, J. Knodt, and F. Heide, "Neural scene graphs for dynamic scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 2856–65.

[63] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Deformable neural radiance fields," 2020.

[64] K. Park, U. Sinha, P. Hedman, J. T. Barron, S. Bouaziz, D. B. Goldman, R. Martin-Brualla, and S. M. Seitz, "Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields," *arXiv preprint arXiv:2106.13228*, 2021.

[65] S. Peng, J. Dong, Q. Wang, S. Zhang, Q. Shuai, H. Bao, and X. Zhou, "Animatable neural radiance fields for human body modeling," 2021.

[66] 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, 9054–63.

[67] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "DNeRF: Neural radiance fields for dynamic scenes," in, 2020.

[68] D. Rebain, W. Jiang, S. Yazdani, K. Li, K. M. Yi, and A. Tagliasacchi, "DeRF: Decomposed radiance fields," 2020.

[69] C. Reiser, S. Peng, Y. Liao, and A. Geiger, "Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 14335–45.

[70] K. Rematas, R. Martin-Brualla, and V. Ferrari, "Sharf: Shape-conditioned radiance fields from a single view," *arXiv preprint arXiv:2102.08860*, 2021.

[71] B. Roessle, J. T. Barron, B. Mildenhall, P. P. Srinivasan, and M. Nießner, "Dense depth priors for neural radiance fields from sparse input views," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, 12892–901.

[72] V. Rudnev, M. Elgharib, W. Smith, L. Liu, V. Golyanik, and C. Theobalt, "NeRF for outdoor scene relighting," in *European Conference on Computer Vision*, Springer, 2022, 615–31.

[73] L. Schirmer, G. Schardong, V. da Silva, H. Lopes, T. Novello, D. Yukimura, T. Magalhaes, H. Paz, and L. Velho, "Neural networks for implicit representations of 3D scenes," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, IEEE, 2021, 17–24.

[74] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger, "Graf: Generative radiance fields for 3d-aware image synthesis," *Advances in Neural Information Processing Systems*, 33, 2020, 20154–66.

[75] M. Shafiei, S. Bi, Z. Li, A. Liaudanskas, R. Ortiz-Cayon, and R. Ramamoorthi, "Learning neural transmittance for efficient rendering of reflectance fields," *arXiv preprint arXiv:2110.13272*, 2021.

[76] K. Singla and P. Astya, "Enhancing 3D implicit shape representation by leveraging periodic activation functions," in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, IEEE, 2021, 101–6.

[77] V. Sitzmann, S. Rezchikov, W. T. Freeman, J. B. Tenenbaum, and F. Durand, "Light field networks: Neural scene representations with single-evaluation rendering," 2021.

[78] 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 *CVPR*, 2021.

[79] S.-Y. Su, F. Yu, M. Zollhöfer, and H. Rhodin, "A-NeRF: Articulated neural radiance fields for learning human shape, appearance, and pose," in *Advances in Neural Information Processing Systems*, 2021.

[80] M. Suhail, C. Esteves, L. Sigal, and A. Makadia, "Light Field Neural Rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 8269–79.

[81] C. Sun, M. Sun, and H.-T. Chen, "Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 5459–69.

[82] J. Sun, X. Wang, Y. Zhang, X. Li, Q. Zhang, Y. Liu, and J. Wang, "Fenerf: Face editing in neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 7672–82.

[83] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. P. Srinivasan, J. T. Barron, and H. Kretzschmar, "Block-NeRF: Scalable large scene neural view synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 8248–58.

[84] M. Tancik, B. Mildenhall, T. Wang, D. Schmidt, P. P. Srinivasan, J. T. Barron, and R. Ng, "Learned initializations for optimizing coordinate-based neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 2846–55.

[85] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, W. Yifan, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, *et al.*, "Advances in neural rendering," in *Computer Graphics Forum*, Vol. 41, No. 2, Wiley Online Library, 2022, 703–35.

[86] 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, 15182–92.

[87] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, "Ref-NeRF: Structured view-dependent appearance for neural radiance fields," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2022, 5481–90.

[88] S. Vora, N. Radwan, K. Greff, H. Meyer, K. Genova, M. S. Sajjadi, E. Pot, A. Tagliasacchi, and D. Duckworth, "NeSF: Neural semantic fields for generalizable semantic segmentation of 3D scenes," *arXiv preprint arXiv:2111.13260*, 2021.

[89] J. Wang, P. Wang, X. Long, C. Theobalt, T. Komura, L. Liu, and W. Wang, "NeuRIS: Neural reconstruction of indoor scenes using normal priors," *arXiv preprint arXiv:2206.13597*, 2022.

[90] 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," *arXiv preprint arXiv:2106.10689*, 2021.

[91] Q. Wang, Z. Wang, K. Genova, P. P. Srinivasan, H. Zhou, J. T. Barron, R. Martin-Brualla, N. Snavely, and T. Funkhouser, "IBRNet: Learning multi-view image-based rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, 4690–9.

[92] Y. Wang, Q. Gao, L. Liu, L. Liu, C. Theobalt, and B. Chen, "Neural novel actor: Learning a generalized animatable neural representation for human actors," *arXiv preprint arXiv:2208.11905*, 2022.

[93] Z. Wang, S. Wu, W. Xie, M. Chen, and V. A. Prisacariu, "NeRF–: Neural radiance fields without known camera parameters," *arXiv preprint arXiv:2102.07064*, 2021.

[94] Y. Wei, S. Liu, Y. Rao, W. Zhao, J. Lu, and J. Zhou, "NerfingMVS: Guided optimization of neural radiance fields for indoor multi-view stereo," in, 2021.

[95] S. Wizadwongsa, P. Phongthawee, J. Yenphraphai, and S. Suwajanakorn, "NeX: Real-time view synthesis with neural basis expansion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[96] W. Xian, J.-B. Huang, J. Kopf, and C. Kim, "Space-time neural irradiance fields for free-viewpoint video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 9421–31.

[97] Y. Xiangli, L. Xu, X. Pan, N. Zhao, A. Rao, C. Theobalt, B. Dai, and D. Lin, "BungeeNeRF: Progressive neural radiance field for extreme multi-scale scene rendering," in *The European Conference on Computer Vision (ECCV)*, 2022.

[98] C. Xie, K. Park, R. Martin-Brualla, and M. Brown, "FiG-NeRF: Figure-ground neural radiance fields for 3D object category modelling," in *International Conference on 3D Vision (3DV)*, 2021.

[99] Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, and S. Sridhar, "Neural fields in visual computing and beyond," in *Computer Graphics Forum*, Vol. 41, No. 2, Wiley Online Library, 2022, 641–76.

[100] D. Xu, Y. Jiang, P. Wang, Z. Fan, H. Shi, and Z. Wang, "SinNeRF: Training neural radiance fields on complex scenes from a single image," in, 2022.

[101] H. Xu, T. Alldieck, and C. Sminchisescu, "H-NeRF: Neural radiance fields for rendering and temporal reconstruction of humans in motion," *Advances in Neural Information Processing Systems*, 34, 2021, 14955–66.

[102] Q. Xu, Z. Xu, J. Philip, S. Bi, Z. Shu, K. Sunkavalli, and U. Neumann, "Point-NeRF: Point-based neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 5438–48.

[103] B. Yang, Y. Zhang, Y. Xu, Y. Li, H. Zhou, H. Bao, G. Zhang, and Z. Cui, "Learning object-compositional neural radiance field for editable scene rendering," in *International Conference on Computer Vision (ICCV)*, October 2021.

[104] L. Yariv, J. Gu, Y. Kasten, and Y. Lipman, "Volume rendering of neural implicit surfaces," *Advances in Neural Information Processing Systems*, 34, 2021, 4805–15.

[105] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin, "iNeRF: Inverting neural radiance fields for pose estimation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[106] A. Yu, R. Li, M. Tancik, H. Li, R. Ng, and A. Kanazawa, "PlenOctrees for real-time rendering of neural radiance fields," in, 2021.

[107] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, "pixelnerf: Neural radiance fields from one or few images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, 4578–87.

[108] Z. Yu, S. Peng, M. Niemeyer, T. Sattler, and A. Geiger, "MonoSDF: exploring monocular geometric cues for neural implicit surface reconstruction," *arXiv preprint arXiv:2206.00665*, 2022.

[109] J. Zhang, X. Liu, X. Ye, F. Zhao, Y. Zhang, M. Wu, Y. Zhang, L. Xu, and J. Yu, "Editable free-viewpoint video using a layered neural representation," *ACM Transactions on Graphics (TOG)*, 40(4), 2021, 1–18.

[110] J. Zhang, E. Sangineto, H. Tang, A. Siarohin, Z. Zhong, N. Sebe, and W. Wang, "3D-aware semantic-guided generative model for human synthesis," *arXiv preprint arXiv:2112.01422*, 2021.

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

[112] X. Zhang, P. P. Srinivasan, B. Deng, P. Debevec, W. T. Freeman, and J. T. Barron, "Nerfactor: Neural factorization of shape and reflectance under an unknown illumination," *ACM Transactions on Graphics (TOG)*, 40(6), 2021, 1–18.

[113] F. Zhao, W. Yang, J. Zhang, P. Lin, Y. Zhang, J. Yu, and L. Xu, "HumanNeRF: Efficiently generated human radiance field from sparse inputs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, 7743–53.

[114] S. Zhi, T. Laidlow, S. Leutenegger, and A. J. Davison, "In-place scene labelling and understanding with implicit scene representation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 15838–47.

[115] Z. Zhu, S. Peng, V. Larsson, W. Xu, H. Bao, Z. Cui, M. R. Oswald, and M. Pollefeys, "NICE-SLAM: Neural implicit scalable encoding for SLAM," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022.

[116] Y. Zhuang, H. Zhu, X. Sun, and X. Cao, "MoFaNeRF: Morphable facial neural radiance field," *arXiv preprint arXiv:2112.02308*, 2021.