NeRF: Representing scenes as neural radiance fields for view synthesis

Journal: *Communications of the ACM* (2022)

Authors:
Ben Mildenhall (Google Research)
Pratul P. Srinivasan (Google Research)
Matthew Tancik (University of California, Berkeley)
Jonathan T. Barron (Google Research)
Ravi Ramamoorthi* (University of California, San Diego)
Ren Ng (University of California, Berkeley)

Citation: This paper presents a superior method for synthesizing photorealistic views of complex scenes using a continuous neural radiance field representation, outperforming previous approaches.

Summary:

NeRFs: the search for the best 3D representation

Abstract: Neural Radiance Fields or NeRFs have become the representation of choice for problems in view synthesis or image-based rendering, as well as in many other applications across computer graphics and vision, and beyond. At their core, NeRFs describe a new representation of 3D scenes or 3D geometry. Instead of meshes, disparity maps, multiplane images or even voxel grids, they represent the scene as a continuous volume, with volumetric parameters like view-dependent radiance and volume density obtained by querying a neural network. The NeRF representation has now been widely used, with thousands of papers extending or building on it every year, multiple authors and websites providing overviews and surveys, and numerous industrial applications and startup companies. In this article, we briefly review the NeRF representation, and describe the three decades-long quest to find the best 3D representation for view synthesis and related problems, culminating in the NeRF papers. We then describe new developments in terms of NeRF representations and make some observations and insights regarding the future of 3D representations.

1. Introduction and Basic NeRF Algorithm

This article is written in response to the Frontiers of Science Award generously granted in 2023 to the papers on "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis" [41, 42] at ECCV 2020 and CACM 2022. NeRFs were introduced as a method that achieved the highest quality visual and quantitative results to date for the task of view synthesis or image-based rendering. This problem is easy to state, although notoriously difficult to solve well. Given a number of input images of a 3D scene of interest, we seek to optimize some kind of internal representation, which will then allow us to synthesize novel views. Achieving this enables one to take a few hand-held images of an object or scene and be able to sense it in 3D instead of as a flat 2D image. There are numerous applications in immersive virtual and augmented reality. As such, view synthesis enables promoting our standard 2D images or photographs to full 3D representations, also enabling newer applications like digital twins and the metaverse. For example, NeRFs have already been used [29, 49] in Google's Maps and StreetView to create immersive renderings from source photographs of cities, buildings and streets, and within the Luma AI mobile phone app to take a few images of an object of interest and produce a compelling fly-through.[1]

NeRFs represent a 3D scene using a fully-connected neural network, known as a multi-layer perceptron or MLP. As opposed to the commonly used convolutional neural network or CNN that has made possible many advances in artificial intelligence (AI) in recent years [28], this representation is non-convolutional and in recent work is often not even a particularly-deep network, enabling simpler and more efficient algorithms that can be more easily trained and better suit the problem at hand. In particular, the input is a single continuous 5D coordinate, involving the spatial position (x, y, z) and the viewing direction (θ, ϕ) and the output is the volume density σ (essentially the absorption or extinction coefficient in the volume, since there is no scattering) and the view-dependent emitted radiance or loosely color $\boldsymbol{c}$. Views are synthesized by querying 5D coordinates along camera rays and using standard volume rendering techniques well known in computer graphics and other fields to project these output colors and densities to an image. A key aspect of deep learning or optimization is to be able to differentiate the relevant functions to run a gradient-descent optimizer. Volume rendering is naturally differentiable (no discontinuities), so the only input needed is the set of images with known camera poses (usually obtained from a system such as COLMAP [58]). Optimization is done simply by comparing the reconstructed value to the "ground truth" input images. In many respects this is a significantly simpler pipeline than previous methods. We show the

[1]The student co-first authors of the original papers now work at Google (Srinivasan, Mildenhall) and Luma AI (Tancik), as of this writing on July 2023. The author of this article is not affiliated with either company.

FIGURE 1. The canonical view synthesis problem. Given a set of input images (here 100 views of the synthetic drums scene randomly captured on the surrounding hemisphere), we first optimize a NeRF representation, then use it with volume rendering to synthesize two novel views, including view-dependent specular effects and reflections. For the past three decades, a quest in view synthesis or image-based rendering has been to find the "best" intermediate representation for 3D scenes that will enable the highest-quality view synthesis; this figure could generically represent many previous papers in the area.

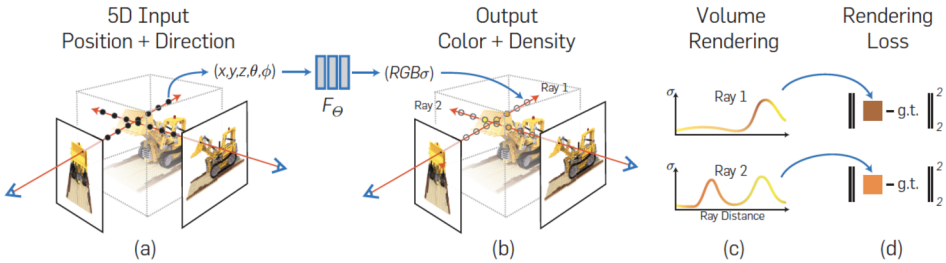

FIGURE 2. Optimizing our neural radiance field representation is done simply by computing a simple loss or error between the reconstructed image computed using volume rendering and held-out views. Medium parameters are determined by a multi-layer perceptron (MLP) that takes 5D coordinates as input and outputs view-dependent emissive radiance (rgb) and volume density (σ).

setup of the problem and the optimization performed in Figs. 1 and 2 (figures taken from [42]).

More formally, the color $C(\boldsymbol{r})$ of camera ray $\boldsymbol{r}(t) = \boldsymbol{o} + t\boldsymbol{d}$ with origin (camera) $\boldsymbol{o}$ and direction $\boldsymbol{d}$ (t is the distance parameter) with near and far bounds t_n and t_f in the volume is given by [42],

$$C(\boldsymbol{r}) = \int_{t_n}^{t_f} T(t)\sigma(\boldsymbol{r}(t))\boldsymbol{c}(\boldsymbol{r}(t), \boldsymbol{d})\, dt, \tag{1}$$

FIGURE 3. Real results obtained by the original NeRF algorithm, showing a close resemblance to ground truth. The ground truth is on the left of the insets, with the NeRF algorithm on the right.

where $T(t)$ is the attenuation of the emissive radiance $\boldsymbol{c}$ at t from traveling through the medium, σ is the volume density and $\boldsymbol{c}$ is the view-dependent emissive color.[2] The attenuation or accumulated transmittance along the ray corresponds to the optical path length, and is given simply by,

$$T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\boldsymbol{r}(s))\, ds\right). \tag{2}$$

We refer to the original papers for details on how to discretize and solve these equations numerically to obtain the computed color $\hat{C}(\boldsymbol{r})$. The optimization simply minimizes the error between the rendered and true pixel colors:

$$\mathcal{L} = \sum_{\boldsymbol{r}\in R} \left\| \hat{C}(\boldsymbol{r}) - C(\boldsymbol{r}) \right\|_2, \tag{3}$$

where R is the set of rays, and we simply compare ground truth (acquired images) $C(\boldsymbol{r})$ and our reconstructed colors $\hat{C}(\boldsymbol{r})$.

While this is a simple formulation, it is important to note the physical limitations. The theory of volumetric scattering or radiative transfer [7] is well known in astrophysics and computer graphics, among many other fields. We have taken only a subset of it, omitting scattering entirely, but keeping absorption and emission. Moreover, most real scenes are not actually emissive and in many cases not really volumes (rather surfaces). However, the volumetric representation is critical, as we will discuss later. Even though the scene is not emissive and actually obtained by an interplay of lighting/illumination and the reflectance properties of the objects, if we fix the lighting, we can "bake in" this illumination and effectively treat the scene as emitting light. However, it is to be understood that these are mathematical conveniences rather than a correct physical model.

[2]It is possible to bake/pre-multiply σ into the color $\boldsymbol{c}$ but this does not provide as elegant a formulation consistent with the volume rendering equations and may be harder to optimize; it may also makes it harder to separate volumetric "occupancy" from color.

Finally, the original papers introduced a positional encoding, which is crucial for an MLP to recover high-frequency functions and produce high-quality results, a topic which we studied in more depth in a companion work [67]. More specifically, the inputs are mapped to a higher-dimensional space, in essence Fourier functions at a variety of higher frequencies, making it easier for the network to learn these higher frequencies. Consider a single input p in the real numbers $\mathcal{R}$ (the same mapping can be applied separately to Cartesian components of spatial position and direction). We lift this input to a higher-dimensional space $\mathcal{R}^{2L}$ with the encoding,

$$\gamma(p) = \big(\sin(2^0\pi p), \cos(2^0\pi p), \ldots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p)\big). \tag{4}$$

Putting this all together, some results for view synthesis on real scenes for the original NeRF paper are shown in Fig. 3, where the quality closely matches the ground truth (input images).

The original NeRF algorithm had many limitations, in terms of excessive training and rendering times, optimizations for one scene/object at a time, requirements of tens to a hundred views, limitations to static bounded scenes, only view synthesis without lighting changes or recovering material parameters, and many more. At least initial work on addressing all of these limitations has proceeded with more than 3,000 papers from research groups around the globe having built on the original NeRF method to date as of July 2023, addressing alomst every conceivable application in graphics, vision and beyond. NeRFs have been used in applications relating to robotics [25], tomography [56], and astronomy [30]. A number of websites[3] have been written and maintained by independent authors, and there are independent blogs and articles summarizing and discussing the large number of NeRF papers at recent vision conferences.[4] Almost every subsequent vision, learning, and graphics conference has given at least one of their best paper awards[5] to articles based on NeRF and extensions, for example, Giraffe at CVPR 21 [46], MipNeRF and Coco3D at ICCV 21 [2, 55], RefNeRF at CVPR 22 [73], our own work on Neural Implicit Evolution at ECCV 22 [39], InstantNGP from SIGGRAPH 22 [43], Dynamic image-based rendering from CVPR 23 [33], 3D Gaussian Splatting for Real-Time Radiance Field Rendering from SIGGRAPH 23 [24], and DreamFusion at ICLR 23 [50]. This is likely not an exhaustive list.

[3] https://dellaert.github.io/NeRF/ and https://neuralradiancefields.io/ among many others.

[4] ICCV 21: https://dellaert.github.io/NeRF21/, CVPR 22: https://dellaert.github.io/NeRF22/, NeurIPS 22: https://markboss.me/post/nerf_at_neurips22/, ECCV 22: https://markboss.me/post/nerf_at_eccv22/, CVPR 23: https://markboss.me/post/nerf_at_cvpr23/. The reader who seeks to follow the recent NeRF literature can follow the links and excellent summaries on these pages.

[5] A subset of the original NeRF authors are the drivers or some of this work [2, 39, 50, 73] but the majority of these and related papers come from a diverse set of research groups around the world.

The subsequent InstantNeRF method from NVIDIA [43] has addressed many shortcomings in terms of rendering and training time, often enabling training in under a minute and rendering at real-time frame rates, leading to recognition by Time Magazine as one of the best inventions of 2022.[6] This has enabled a number of applications, including creation of NeRFs by lay users; some artist creations are showcased in an Instant NeRF AI Gallery website.[7] The New York Times has used NeRFs for capturing portraits,[8] and Mark Zuckerberg, CEO of Meta/Facebook has mentioned NeRFs and Inverse Rendering as key technologies for the metaverse.[9] Given NeRF's popularity, a number of software frameworks have been developed; one open source example is NeRFStudio [68]. A number of surveys on NeRF have been published [13, 70] which we refer readers to, and even the popular press has highlighted the method [27]. Our goal is not to duplicate these writeups, and indeed it is impossible to do justice to all the excellent follow-on work in a few pages.

The remainder of this article provides historical context on the three-decades long quest for the "right" 3D scene representation for view synthesis or image-based rendering. We then discuss new developments in core NeRF representations, and relate them to earlier work on light transport reprsentation and real-time rendering. Our goal is not to be exhaustive, but identify some of the core challenges and solutions in the NeRF algorithm.

2. Scene Representations for Image-Based Rendering

Our story begins at least 30 years ago, pre-dating the research careers (and for students indeed their birth) of any of the authors of the NeRF paper. In SIGGRAPH 93, Chen and Williams [10] first introduced the notion of view interpolation for image synthesis.[10] This paper has the nearly unique distinction of being included in both volumes of seminal graphics papers [74, 77] produced by SIGGRAPH for the 25th and 50th anniversary conferences. While Chen and Williams used synthetic z-buffer images, this paper is widely taken as defining the view synthesis problem, and leading to the immensely popular field of image-based rendering (IBR) in computer graphics and vision in the 1990s.

At that time, computer graphics rendering or image synthesis had already achieved striking realistic imagery but was limited by the quality of input models for geometry, lighting and materials. Image-based rendering

[6] https://time.com/collection/best-inventions-2022/6225489/nvidia-instant-nerf/

[7] https://www.nvidia.com/en-us/research/ai-art-gallery/instant-nerf/

[8] https://rd.nytimes.com/projects/creating-workflows-for-nerf-portraiture

[9] https://youtu.be/hvfV-iGwYX8?t=4400

[10] Lance Williams received SIGGRAPH's 2001 Coons Lifetime Achievement Award, the community's highest honor, primarily for pioneering the image-based approach to graphics or "brute force in image space" including shadow mapping [75] and mip-mapping [76].

or view synthesis offered an alternative of using by definition photorealistic input images and combining them to create new images. Within computer vision, this was an exciting new problem direction, as opposed to the traditional problems of 3D reconstruction or recognition. Indeed, much of the early promise in image-based rendering was to bypass error-prone 3D reconstruction entirely. Tomaso Poggio's group pioneered work on facial analysis and synthesis in an image-based way (for example [15]), but often does not get the credit richly deserved since these publications were not in the mainstream graphics community.

Much like the explosion of work on NeRFs, there was an explosion of work on view synthesis and image-based rendering, including classic papers on plenoptic modeling [38], light field rendering [31], the lumigraph [18], layered-depth images [60], and thousands of others we cannot cover in this article. Indeed, the seminal graphics papers volume from the 50$^{\text{th}}$ SIGGRAPH conference in 2023 [74] includes at least a dozen IBR papers from the mid-1990s to the mid-2000s (listed in chronological order) [10, 38, 31, 18, 12, 81, 11, 78, 5, 45, 65]. I highly recommend re-reading these classics. The SIGGRAPH technical awards program and committee did a remarkable job keeping up with these advances, with major computer graphics technical awards deservedly given to at least one author of essentially all of the above-mentioned papers.

Subsequent developments have especially been inspired by the light field[11] and lumigraph papers [18, 31] which show how resampling and combining rays can achieve view synthesis, and more importantly provide the foundations for reconstructing the light field from sparse images. Indeed, my own interest in the project that eventually led to NeRF started with depth reconstruction and view interpolation from light field cameras [23, 69], using Lytro cameras from co-author Ren Ng's startup. Layered-depth images or LDIs [60] are one precursor to Multi-Plane Images or MPIs widely used prior to the NeRF method (including in our work on local light field fusion [40]).

A critical question in these and follow-on works has been the right 3D scene representation for image-based rendering. Some early papers like the original ight field work [31] argued that we did not need a scene representation; one was simply interpolating images. In this way, the error-prone 3D reconstruction step could completely be skipped. However, in a seminal paper at SIGGRAPH 2000, Chai et al. [21] formally demonstrated a plenoptic sampling theory, showing that depth information can enable much sparser acquisition, and in fact there is a tradeoff between depth accuracy and sampling rate (an analysis we exploited in our own subsequent work [40]). The lumigraph paper [18] had already explored constructing an approximate geometric representation.

[11]Some seminal early work on the plenoptic function or light field comes from Adelson [1], although the concept of light fields and light field cameras goes back over a century, pre-dating computing [20, 34].

Ultimately, it became clear that despite the name "image-based rendering", the key element was the intermediate 3D scene representation.[12] In essence, most algorithms follow the pipeline of Fig. 1, where one takes the input views, optimizes an intermediate 3D scene representation, and then renders the final views. Inspired by traditional computer graphics and vision, it is natural for the scene representation to be approximate 3D geometry such as a rough mesh, or perhaps a disparity map. But these representations can be very challenging for high-quality view synthesis. For example, consider the canonical voxel-coloring approach by Seitz and Dyer [**59**]. Once a mistake is made, the visibility relationships in the shapes are impacted going forward and the final recovered 3D structure is inaccurate with incorrect visibility. Moreover, if geometry is incorrect, then view-dependent effects like specularities are often incorrect. Nevertheless, the early image-based rendering work achieved a certain level of success with increasingly sophisticated algorithms over the next two decades; representative papers are [**8**] that uses more conventional depth synthesis and warping, and [**48**] that acknowledges the limitations of depth and 3D reconstruction, aiming for a notion of a "soft" or "probabilistic" 3D reconstruction that does not commit to a hard 3D shape. In effect, along a given ray, we are not considering a single surface or intersection points, but a probability distribution of intersections. This is a very important insight, which NeRFs in some sense take to its natural conclusion.

In 2012, Krizhevsky et al. [**28**] introduced massive gains in classification problems with deep convolutional networks, leading to a tsunami of work in deep learning for computer vision. For a while, it seemed that image synthesis would not be touched by these developments. However, fully convolutional networks and U-nets enabled creation of high-resolution imagery [**36**] and results were quickly applied to stereo reconstruction [**16**]. Our own work considered view synthesis from light-field cameras, obtaining a fast algorithm to obtain the full light field from only four corner views [**23**]. This paper (coauthored with my postdoc Nima Kalantari and Ph.D. student Ting-Chun Wang) can reasonably be considered the first effective deep-learning algorithm for view synthesis, and was a critical marker for the research that followed. Nevertheless, the representation proposed still largely used disparities and warping in the classical sense.

Our paper on Neural Radiance Fields presents a dramatic break from this tradition of 3D geometric representations. First, the representation of the scene is a volume rather than a surface. Second, it is a continuous volume rather than a discrete voxel grid. It is likely that this core insight of a continuous volumetric representation is the crucial lasting insight from the NeRF paper, although there are also a number of contributions in the specific learning algorithms. Of course, research doesn't happen in a vacuum, and there were a number of important precursors. Volume rendering [**14**] has a

[12]Readers are encouraged to read the ACM Dissertation Award honorable mention theses of the NeRF authors; in particular Pratul Srinivasan's thesis [**66**] discusses his quest for the best representation with deep learning for view synthesis.

long history in graphics, and complex materials are increasingly represented in a volumetric fashion [22]. Indeed, I had a large NSF grant starting in 2010 to propose a fundamentally new volume-based material representation; the NeRF paper a decade later has truly made volumes a first-class primitive for scene representation. We were inspired by the neural volumes work [35] that uses a discrete voxel grid with a few other limitations, predicted using a deep 3D CNN (a similar idea is proposed in Deep Voxels [61]), but NeRF generalizes this idea in many ways and in particular to a *continuous* rather than discrete volumetric representation. As an aside, instead of simply storing volume densities, one can store reflectance to recover light and view-dependent properties, a line of work my group pursued concurrently [3, 4].

In terms of learning algorithms also, our NeRF paper breaks new ground. Instead of using a CNN, the dominant paradigm at the time, one optimizes a simple MLP; the original work required only 5MB for the weights of this fully-connected network to represent the entire scene. Again, the idea is not entirely new, and builds on pioneering work on deep signed distance functions [47] and scene representation networks [62]. Finally, NeRF introduces the notion of positional encoding for high-frequency features, analyzed in more detail in a companion paper [67], and built upon in a wide range of applications by many authors. This leverages work in the learning community on spectral bias in neural networks [52]. It should be noted, as discussed in the next section, that a number of new NeRF representations have modified the learning approach to use only shallow MLPs without positional encoding, and in the most recent work [24] not even using a neural representation. However, the core concept of a continuous volumetric representation, introduced in NeRF, has pervaded essentially all of the subsequent work.

3. Representations for Neural Radiance Fields

The original NeRF paper has spawned an explosion of new representations, which we briefly discuss here (but this is far from an exhaustive list). Though rarely cited, this parallels the development two decades ago of compact representations for light transport and precomputed radiance transfer (PRT) [64], in which the author of this article played a major role [53]. For example, spherical harmonics and reflection-based reparameterizations are commonly used [73]. An early method in the PRT literature was to cluster scenes [63] or to break images into blocks [44]. Similarly, KiloNeRF [54] breaks the scene into regions, each fit with small MLPs. However, there has so far been little analysis or understanding of optimal block and cluster sizes, or the tradeoffs involved, work which has been pursued in some depth in the PRT literature [37]. One strong recommendation we make for future authors is to be aware of the PRT literature, and cite it/build on it appropriately. There is also potential in the other direction, of bringing NeRF representations into classical graphics problems; we have made a first step towards this using NeRF-like ideas for PRT with glossy global illumination [51].

Many recent methods are based on the notion of feature-fields. This makes a step back towards discrete volumetric representations, wherein a coarse volumetric grid of abstract features is stored explicitly. One can then interpolate this grid at any point along the ray traversal to obtain features, which are subsequently fed into a small MLP to decode the color and density values, after which rendering proceeds like in the original NeRF algorithm. While this does take considerably more memory than a standard NeRF (but much less than a full-resolution voxelized volumetric model), it is easier to train and substantially faster to evaluate, given the smaller MLPs required. One challenge then is the appropriate representation and compression of the feature field itself. Two recent representative works in this area are Instant Neural Graphics Primitives [43] which uses a multiresolution hash-grid and TensorRF [9] which does a tensor deocmposition of the feature grid. Note that TensorRF has strong similarities to clustered tensor approximation [72] in PRT, and we encourage future authors to more carefully investigate the potential for clustering and analyze the number of terms needed in this representation. It is also possible to combine TensorRF for factorization and InstantNGP for representation of the lower-dimensional components with multiresolution hash-grids; this idea has been used for example for dynamic full-body human models [19] and for static PRT in our own work [51]. In any event, this explosion of new representations, factorizations and compressions indicates that mathematical representations and signal-processing are key even in the deep learning-NeRF era.

There have also been a few other novel representations proposed. EG3D [6] essentially projects a 3D point to lower-dimensional 2D planes. Since the 2D planes are lower-dimensional, they can represent much higher-resolution information explicitly as a simple texture-map, and CNN or transformer-based algorithms can even predict these representations directly [71]. This can be extended to higher dimensions using K-planes [17], which can even operate without any neural layers or MLP (earlier work on plenoxels developed an adaptive octree voxel representation, again without neural components [80]).

The breakthrough in NeRFs was achieved by leveraging older computer graphics notions of volume rendering. In the past year, further breakthroughs have been achieved through an even older graphics technique of point-based rendering, where points are used as a display primitive [32, 57]. Point-nerf [79] introduced the notion of point-based NeRF representations as easier to train. A radical new approach was proposed by Kerbl et al. [24] using 3D gaussian splats for their point primitives. This is currently the state-of-the-art method in terms of visual quality, rendering time and performance, but does have higher memory requirements. In a sense, this work comes full-circle in the memory-speed tradeoff with a fully explicit volumetric representation. Most remarkably, the technique is entirely classical, with no neural primitives, and the continuous volume essentially represented adaptively with gaussian blobs or splats, which serve as the point samples. Within the past

decade, this is perhaps the first time a major neural algorithm has been superseded by a non-neural classical equivalent. Again, research doesn't happen in a vacuum, and a similar Gaussian splat representation for surfaces and volumes in differentiable rendering for somewhat different shape from silhoutte and related problems was proposed a year earlier by Keselman and Hebert [**26**].

Which brings us back to a key point made in the last section: While NeRFs are often thought about as another major victory for AI-based techniques in making remarkable advances on decades-old problems, the key insight may actually simply be in the idea of a continuous volumetric representation,[13] instead of surfaces or discrete voxel grids, rather than any particular learning representation or algorithm. In fact, there may be no need for any neural component at all, and what we really care about is just the radiance field.

We have devoted a few paragraphs to a subject that has involved hundreds or thousands of papers over the past three years. The field of NeRF representations is rapidly evolving, with many chapters still to be written in the years to come. As such, NeRFs can be seen as one milestone, but by no means the final word on the "best" 3D scene representation for image-based rendering.

4. Conclusion: NeRFs as the new Geometric Primitive

A representation of 3D geometry is critical within computer graphics. Indeed, the canonical graphics pipeline involves 3D scene geometric models, along with lights, materials, animations and a rendering algorithm to create images. Geometry is also manipulated by smoothing it, simulating it, meshing it, or using it as a representation for perception, recognition or editing. To date, the most successful representation of 3D geometry in computer graphics is perhaps simply the triangle mesh. But the history of computer graphics has had many rich alternatives, including spline patches, subdivision surfaces, implicit surfaces, and points.

In conclusion, we argue that NeRFs are the new geometric primitive, based on volumes rather than surfaces, and based on neural rather than explicit representations (although we have seen some work in the last paragraph can achieve state-of-the-art performance without any neural primitives). As such, NeRFs and neural implicit functions can often be a plug-in replacement for any problem that requires a 3D geometric representation. This has led to their immense popularity, since the ideas can be applied to almost any domain. It remains an open question as to whether operations in the computer graphics pipeline like physical simulation and geometry processing that have

[13] It can be argued that the point-based representation in Gaussian splats [**24**] is not strictly a volume and does not actually require ray marching, but the alpha compositing of splats obeys the same equations and we view it an adaptive representation of the continuous volume.

historically been applied on finite element models, meshes and tetrahedra should be extended to the NeRF domain, or whether it is simpler to extract an explicit mesh representation from a NeRF. Regardless, the view of NeRFs as the new 3D geometric primitive is a valuable one, and enables their potential usage anywhere one seeks a 3D (or higher-dimensional) geometric model. This gives us an immense set of applications to focus on, and we expect research on both core radiance field representations and NeRF applications to continue for many years to come.

Acknowledgements:

First, I of course want to thank my co-authors on the NeRF papers, my Ph.D. student Pratul Srinivasan, Ben Mildenhall, Matt Tancik, Jon Barron, and Ren Ng. We also acknowledge Kevin Cao, Guowei Yang and Nithin Raghavan for comments and discussions. I wish to acknowledge as well the early students in the light field projects, including Michael Tao, Ting-Chun Wang and my postdoc Nima Kalantari, who first showcased the use of deep-learning based view synthesis for light fields. I am grateful to all of my colleagues and students within the UC San Diego Center for Visual Computing. I wish to thank the many funders of our research over the years; the original papers acknowledge ONR grants N000141712687, N000141912293, N000142012529 and most recently N00014-23-1-2526 and NSF Chase-CI grants 2100237 and 2120019, the Ronald L. Graham Chair, as well as NSF (Pratul Srinivasan, Matt Tancik) and Hertz Fellowships (Ben Mildenhall), compute credits through the BAIR commons program, and Blend Swap users for models. I also acknowledge many gifts and industrial contracts over the years to our light field research program from Google, Adobe, Samsung, Qualcomm, Sony, and many other partners. In particular, I would like to acknowledge program manager Behzad Kamgar-Parsi at ONR for a now close to two-decades long support of my research endeavors in physics-based computer vision, including the light field and NeRF projects.

It is with a deep gratitude that we thank the organizers of the International Congress on Basic Science for recognizing Neural Radiance Fields with an inaugural Frontiers of Science Award, and for their support of basic scientific research in so many fields. NeRFs represent the culmination, or perhaps a waypoint and milestone, in a three-decades long quest to find the best scene representation for view synthesis and image-based rendering. Perhaps the greatest thanks go to the many researchers who initiated this research effort and sustained it over the past 30 years, and the many thousands of authors who have built on the original NeRF paper in the past three years. The Frontiers of Science Award really belongs to them all.

References

[1] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, MIT Press, 1991.

[2] J. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. Srinivasan. Mip-NeRF: A multiscale reprsentation for anti-aliasing neural radiance fields. In *International Conference on Computer Vision (ICCV 21)*, pages 5855–5864, 2021.

[3] S. Bi, Z. Xu, P. Srinivasan, B. Mildenhall, K. Sunkavalli, M. Hasan, Y. Geoffroy, D. Kriegman, and R. Ramamoorthi. Neural reflectance fields for appearance acquisition. Technical Report, arXiv:2008.03824, 2020.

[4] S. Bi, Z. Xu, K. Sunkavalli, M. Hasan, Y. Hold-Geoffroy, D. Kriegman, and R. Ramamoorthi. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *European Conference on Computer Vision (ECCV)*, 2020.

[5] C. Buehler, M. Bosse, L. McMilland, S. Gortler, and M. Cohen. Unstructured lumigraph rendering. In *SIGGRAPH 01*, pages 425–432, 2001.

[6] E. Chan, C. Lin, M. Chan, K. Nagano, B. Pan, S. Mello, O. Gallo, L. Guibas, J. Tremblay, S. Khamis, T. Karras, and G. Wetzstein. Efficient geometry-aware 3D generative adversarial networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 16102–16112, 2022.

[7] S. Chandrasekhar. *Radiative Transfer*. Courier Corporation, 1960.

[8] G. Chaurasia, S. Duchene, O. Sorkine-Hornung, and G. Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Transactions on Graphics*, 32(3):30:1–30:12, 2013.

[9] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su. TensoRF: Tensorial radiance fields. In *European Conference on Computer Vision (ECCV)*, 2022.

[10] S. Chen and L. Williams. View interpolation for image synthesis. In *SIGGRAPH 93*, pages 279–288, 1993.

[11] P. Debevec, T. Hawkins, C. Tchou, H. P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *SIGGRAPH 00*, pages 145–156, 2000.

[12] P. Debevec, C. Taylor, and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *SIGGRAPH 96*, pages 11–20, 1996.

[13] F. Dellaert and Y. Lin. Neural volume rendering: NeRF and beyond. Technical Report, arXiv:2103.13178, 2021.

[14] R. Drebin, L. Carpenter, and P. Hanrahan. Volume rendering. In *SIGGRAPH 88*, pages 65–74, 1988.

[15] T. Ezzat and T. Poggio. Facial analysis and synthesis using image-based models. In *FG 96*, pages 116–121, 1996.

[16] J. Flynn, I. Neulander, J. Philbin, and N. Snavely. Deepstereo: Learning to predict new views from the world's imagery. In *IEEE CVPR*, pages 5515–5524, 2016.

[17] S. Fridovich-Keil, G. Meanti, F. Warburg, B. Recht, and A. Kanazawa. K-planes: Explicit radiance fields in space, time, appearance. In *Computer Vision and Pattern Recognition (CVPR)*, 2023.

[18] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The Lumigraph. In *SIGGRAPH 96*, pages 43–54, 1996.

[19] M. Isik, M. Runz, M. Georgopoulos, T. Khakhulin, J. Starck, L. Agapito, and M. Niessner. HumanRF: High-fidelity neural radaince fields for humans in motion. *ACM Transactions on Graphics (SIGGRAPH 23)*, 42(4), 2023.

[20] F. Ives. Parallax stereograms and process of making same. US Patent 725567, 1903.

[21] X. Tong J. Chai and H. Shum. Plenoptic sampling. In *SIGGRAPH 00*, pages 307–318, 2000.

[22] W. Jakob, A. Arbree, J. Moon, K. Bala, and S. Marschner. A radiative transfer framework for rendering materials with anisotropic structures. *ACM Transactions on Graphics*, 29(4), 2010.

[23] N. Khademi Kalantari, T. Wang, and R. Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics (SIGGRAPH Asia 16)*, 35(6):193:1–193:10, 2016.

[24] B. Kerbl, G. Kopanas, T. Leimkuhler, and G. Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (SIGGRAPH 2023)*, 42(4), 2023.

[25] J. Kerr, L. Fu, H. Huang, Y. Avigal, M. Tancik, J. Ichnowski, A. Kanazawa, and K. Goldberg. Evo-NeRF: Evolving NeRF for sequential robot grasping of transparent objects. In *Robot Learning*, 2022.

[26] L. Keselman and M. Hebert. Approximate differentiable rendering with algebraic surfaces. In *European Conference on Computer Vision (ECCV)*, volume 32, pages 596–614, 2022.

[27] W. Knight. A new trick lets artificial intelligence see in 3d. Wired Magazine: https://www.wired.com/story/new-way-ai-see-3d/, 2022.

[28] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[29] Luma Labs. Luma AI app. https://lumalabs.ai/, 2023.

[30] A. Levis, P. Srinivasan, A. Chael, R. Ng, and K. Bouman. Gravitationally lensed black hole emission tomography. In *Computer Vision and Pattern Recognition (CVPR)*, 2022.

[31] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH 96*, pages 31–42, 1996.

[32] M. Levoy and T. Whitted. The use of points as a display primitive. Technical report, University of North Carolina at Chapel Hill, 1985.

[33] Z. Li, Q. Wang, F. Cole, R. Tucker, and N. Snavely. DynIBAR: Neural dynamic image-based rendering. In *CVPR 23*, 2023.

[34] G. Lippmann. Epreuves reversibles photographies integrales. *Academie des Sciences*, pages 446–451, 1908.

[35] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, and Y. Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *ACM Transactions on Graphics*, 38(4):65:1–65:14, 2019.

[36] J. Long, E. Shelhammer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *IEEE CVPR*, pages 3431–3440, 2015.

[37] D. Mahajan, I. Kemelmacher-Shlizerman, R. Ramamoorthi, and P. Belhumeur. A theory of locally low dimensional light transport. *ACM Transactions on Graphics (Proc. SIGGRAPH 07)*, 26(3):62, 2007.

[38] L. McMillan and G. Bishop. Plenoptic modeling: an image-based rendering system. In *SIGGRAPH 95*, pages 39–46, 1995.

[39] I. Mehta, M. Chandraker, and R. Ramamoorthi. A level set theory for neural implicit evolution under explicit flows. In *ECCV 22*, 2022.

[40] B. Mildenhall, P. Srinivasan, R. Ortiz-Cayon, N. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (SIGGRAPH 2019)*, 38(4):29:1–29:14, 2019.

[41] B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, R. Ramamoorthi, and R. Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, pages I–405–I–421, 2020.

[42] B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, R. Ramamoorthi, and R. Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2022.

[43] T. Muller, A. Evans, C. Schied, and A. Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (SIGGRAPH 22)*, 41(4), 2022.

[44] S. Nayar, P. Belhumeur, and T. Boult. Lighting-sensitive displays. *ACM Transactions on Graphics*, 23(4):963–979, 2004.

[45] S. Nayar, G. Krishnan, M. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics (SIGGRAPH 2006)*, 25(3), 2006.

[46] M. Niemeyer and A. Geiger. GIRAFFE: Representing scenes as compositional generative neural feature fields. In *Computer Vision and Pattern Recognition (CVPR)*, pages 11453–11464, 2021.

[47] J. Park, P. Florence, J. Straub, and R. Newcombe. DeepSDF: Learning continuous signed distance functions for shape representation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 165–174, 2019.

[48] E. Penner and L. Zhang. Soft 3D reconstruction for view synthesis. *ACM Transactions on Graphics*, 36(6):235:1–235:11, 2017.

[49] C. Phillips. New ways maps is getting more immersive and sustainable. https://blog.google/products/maps/sustainable-immersive-maps-announcements/, 2023.

[50] B. Poole, A. Jain, J. Barron, and B. Mildenhall. DreamFusion: Text-to-3D using 2d diffusion. In *International Conference on Learning Research*, 2023.

[51] N. Raghavan, Y. Xiao, K. Lin, T. Sun, S. Bi, Z. Xu, T. Li, and R. Ramamoorthi. Neural free-viewpoint relighting for glossy indirect illumination. *Computer Graphics Forum (EGSR 23)*, 2023.

[52] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville. On the spectral bias of neural networks. In *International Conferencd on Machine Learning (ICML)*, 2018.

[53] R. Ramamoorthi. Precomputation-based rendering. *Foundations and Trends in Computer Graphics and Vision*, 3(4):281–369, 2009.

[54] C. Reiser, S. Peng, Y. Liao, and A. Geiger. KiloNeRF: Speeding up neural radiance fields with thousands of tiny MLPs. In *International Conference on Computer Vision (ICCV)*, pages 14315–14325, 2021.

[55] J. Reizenstein, R. Shapovalov, P. Henzler, L. Sbordone, P. Labatut, and D. Novotny. Common objects in 3D: Large-scale learning and evaluation of real-life 3d category reconstruction. In *International Conference on Computer Vision (ICCV)*, pages 10901–10911, 2021.

[56] D. Ruckert, Y. Wang, R. Li, R. Idoughi, and W. Heidrich. NeAT: Neural adaptive tomography. *ACM Transactions on Graphics*, 41(4):55:1–55:13, 2022.

[57] S. Rusinkiewicz and M. Levoy. QSplat: A multiresolution point rendering system for large meshes. In *SIGGRAPH 00*, pages 343–352, 2000.

[58] J. Schonberger, E. Zheng, M. Pollefeys, and J. Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[59] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proc. CVPR*, 1997.

[60] J. Shade, S. Gortler, L. He, and R. Szeliski. Layered depth images. In *SIGGRAPH*, pages 231–242, 1998.

[61] V. Sitzmann, J. Thies, F. Heide, M. Niessner, G. Wetzstein, and M. Zollhofer. Deepvoxels: Learning persistent 3D feature embeddings. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2437–2446, 2019.

[62] V. Sitzmann, M. Zollhofer, and G. Wetzstein. Scene representation networks: Continuous 3D-structure-aware neural scene representations. In *NeurIPS*, pages 1119–1130, 2019.

[63] P. Sloan, J. Hall, J. Hart, and J. Snyder. Clustered principal components for precomputed radiance transfer. *ACM Transactions on Graphics (Proc. SIGGRAPH 03)*, 22(3):382–391, 2002.

[64] P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics (Proc. SIGGRAPH 02)*, 21(3):527–536, 2002.

[65] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics (SIGGRAPH 2006)*, 25(3):835–846, 2006.

[66] P. Srinvasan. *Scene Representations for View Synthesis with Deep Learning*. PhD thesis, University of California Berkeley, December 2020.

[67] M. Tancik, P. Srinivasan, B. Mildenhall, S. FridovichKeil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *NeurIPS*, 2020.

[68] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, T. Wang, A. Kristoffersen, J. Austin, K. Salahi, A. Ahuja, D. McAllister, J. Kerr, and A. Kanazawa. Nerfstudio: A modular framework for neural radiance field development. In *SIGGRAPH 23 (conference papers)*, pages 72:1–72:12, 2022.

[69] M. Tao, P. Srinivasan, S. Hadap, S. Rusinkiewicz, J. Malik, and R. Ramamoorthi. Shape estimation from shading, defocus and correspondence using light-field angular coherence. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 39(3):546–560, 2017.

[70] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, Y. Wang, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, T. Simon, C. Theobalt, M. Niessner, J. Barron, G. Wetzstein, M. Zollhoefer, and V. Golyanik. Advances in neural rendering. *Computer Graphics Forum (EuroGraphics 22 STAR)*, 41(2):703–735, 2022.

[71] A. Trevithick, M. Chan, M. Stengel, E. Chan, C. Liu, Z. Yu, S. Khamis, M. Chandraker, R. Ramamoorthi, and K. Nagano. Real-time radiance fields for single-image portrait view synthesis. *ACM Transactions on Graphics*, 42(4):1–15, 2023.

[72] Y. Tsai and Z. Shih. All-frequency precomputed radiance transfer using spherical radial basis functions and clustered tensor approximation. *ACM Transactions on Graphics (Proc. SIGGRAPH 06)*, 25(3):967–976, 2006.

[73] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. Barron, and P. Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In *Computer Vision and Pattern Recognition (CVPR)*, 2022.

[74] M. Whitton, editor. *Seminal Graphics Papers: Pushing the Boundaries*, volume 2. Association for Computing Machinery (ACM), August 2023.

[75] L. Williams. Casting curved shadows on curved surfaces. In *SIGGRAPH 78*, pages 270–274, 1978.

[76] L. Williams. Pyramidal parametrics. In *SIGGRAPH 83*, pages 1–11, 1983.

[77] R. Wolfe, editor. *Seminal Graphics: Pioneering Efforts that Shaped the Field*, volume 1. Association for Computing Machinery (ACM), July 1998.

[78] D. Wood, D. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. Salesin, and W. Stuetzle. Surface light fields for 3D photography. In *SIGGRAPH 00*, pages 287–296, 2000.

[79] Q. Xu, Z. Xu, J. Philip, S. Bi, Z. Shu, K. Sunkavalli, and U. Neumann. Point-nerf: Point-based neural radiance fields. In *Computer Vision and Pattern recognition (CVPR)*, pages 5438–5448, 2022.

[80] A. Yu, S. Fridovich-Keil, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa. Plenoxels: Radiance fields without neural networks. In *Computer Vision and Pattern Recognition (CVPR)*, 2022.

[81] D. Zongker, D. Werner, B. Curless, and D. Salesin. Environment matting and compositing. In *SIGGRAPH 99*, pages 205–214, 1999.

Ravi Ramamoorthi, University of California San Diego.
The author is also currently affiliated with NVIDIA
Email address: ravir@cs.ucsd.edu