

# NeRF：3D 视觉中的神经辐射场综合调研

2152085 孙亦菲

**摘要** 本调研报告深入探讨了神经辐射场（Neural Radiance Field, NeRF）这一计算机视觉和图形学领域的重要突破。NeRF 利用深度学习方法，基于多视角二维图像合成高质量的三维场景，有效解决了传统三维重建方法在细节捕捉、光照处理和视角自由度方面的局限性。报告首先介绍了 NeRF 的基本工作原理及其算法架构，并详细分析了其如何精确学习场景中每个点的颜色和密度信息。接着，报告展示了 NeRF 在三维视觉的重建、渲染、定位、生成、理解等任务中的应用，凸显其为相关领域带来的新鲜活力和变革潜力。最后，报告对 NeRF 未来的发展趋势进行了展望，包括技术革新、应用拓展以及与其他前沿技术的融合可能性。

**关键词** 神经辐射场（NeRF）、三维重建、深度学习、视觉渲染、技术革新

## 1 引言

### 1.1 调研背景

神经辐射场（Neural Radiance Field, NeRF）是近年来计算机视觉和图形学领域的一项重要突破，它基于深度学习方法，将多视角的二维图像合成为高质量的三维场景。这一技术的提出，源于解决传统三维重建方法在细节捕捉、光照处理和视角自由度方面的局限性。通过对大量图片数据的学习，NeRF 能够生成连续的三维空间表达，实现对光线传播过程的精确模拟。这种技术不仅在理论上具有创新性，也在实际应用中展示了巨大的潜力。Mildenhall 等人在 ECCV 2020 中首次对 NeRF 进行了介绍<sup>[1]</sup>。NeRF 已经达到了 SOTA 的视觉质量，其演示令人印象深刻，而且启发了许多源自这种新方法的后作。最近，NeRF 模型已在照片编辑、3D 表面提取和大型/城市规模（city-scale）的 3D 表示和视图合成中得到了应用。

### 1.2 调研内容

本文将从 NeRF 的基本工作原理及其算法架构触发，重点分析 NeRF 如何通过深度学习模型精确学习场景中每个点的颜色和密度信息。

然后本文调研了 NeRF 技术下三维视觉的重建、渲染、定位、生成、理解等任务的发展，展示了 NeRF 为计算机视觉和图形学领域带来的新鲜活力。

最后，本文将调研 NeRF 未来的发展趋势，尤其是在技术革新、应用拓展和商业模式创新等方面的潜在变化。此外，将探讨可能的技术突破点，例如新的应用场景的探索、与其他前沿技术（如人工智能、增强现实/虚拟现实）的融合等。

## 2 研究内容

### 2.1 神经辐射场 (NeRF) 理论

NeRF 的核心原理是使用一个小型的神经网络来学习场景的连续体积表示。这个表示能够预测在任意三维位置和视角下光线的颜色和密度。在实践中,这意味着 NeRF 可以从一系列普通的二维图像中,通过神经网络的训练,生成新视角下的逼真三维场景。

NeRF 的工作过程大致可以分为两个步骤:首先,神经网络根据输入的多个视角图像学习场景的体积密度和颜色,即构建隐式神经场。然后,利用体积渲染技术,通过积分神经网络预测的颜色和密度,来合成新视角下的图像。再通过渲染结果与图片的误差进行梯度下降优化神经辐射场。

这种方法的关键在于能够精确模拟光线在三维空间中的传播,包括光线的弯曲、散射和其他复杂的物理效应。工作过程如图 1所示。

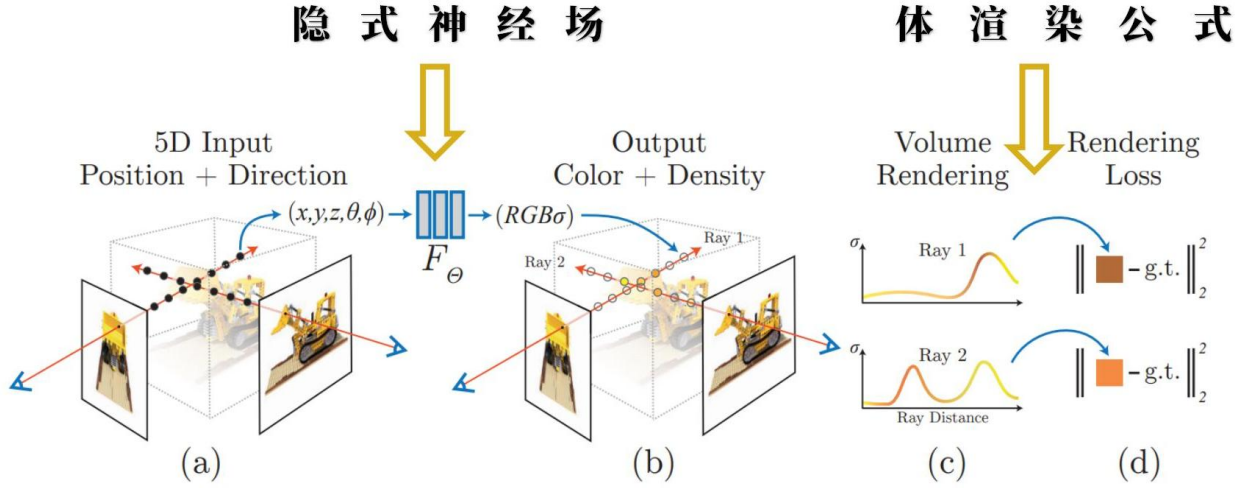


Fig. 1 隐式神经场和可微体渲染过程的概述

#### 2.1.1 隐式神经场

NeRF 中的隐式神经场,即用基于坐标的全连接网络表达了颜色场和体密度场。用通俗的话来说,这个网络能够从一系列二维图像中学习到场景的三维结构,但这个网络不直接输出三维坐标的集合,而是学习一个函数,这个函数能够映射三维空间中的任何一个点到一个颜色和密度值。这种表示方式被称为隐式,因为三维结构是通过这个映射间接表示的。这个映射是由一个神经网络学习得到的,该网络通过最小化从合成图像到实际图像的差异来进行训练。这种训练方式使得网络能够理解场景的三维结构和外观。

在数学上,这可以通过如下公式表示:

$$F(\mathbf{x}, \theta, \phi) \rightarrow (\mathbf{c}, \sigma)$$

其中  $\mathbf{x} = (x, y, z)$  是场景坐标,  $(\theta, \phi)$  表示方位角和极角,  $\mathbf{c} = (r, g, b)$  代表颜色,  $\sigma$  代表体积密度。这个 5D 函数由一个或多个多层感知机 (Multi-Layer Perceptron, MLP) 近似,有时表示为  $F_\Theta$ 。两个视角  $(\theta, \phi)$  通常由  $\mathbf{d} = (d_x, d_y, d_z)$  表示,这是一个 3D 笛卡尔单位向量。

隐式神经场的一个主要优势是它能够捕获高度复杂的几何和光学细节，而不需要传统的三维模型那样的显式结构。这使得 NeRF 非常适合于从一组有限的图像中重建复杂和高度逼真的场景。

$$F(\mathbf{x}, \theta, \phi) \rightarrow (\mathbf{c}, \sigma) \quad (1)$$

### 2.1.2 体渲染公式

NeRF 中的体渲染公式是其核心组成部分，用于从隐式神经场中合成图像。这个公式基于体积渲染的原理，通过计算光线在穿过场景时与物质的交互来生成图像。数学上，这可以表示为沿着光线路径积分颜色和密度：细言之，给定体积密度和颜色函数，以及相机位置  $\mathbf{o}$  和观察方向  $\mathbf{d}$ ，体渲染 [20] 可以获取任何相机光线  $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$  的颜色  $C(\mathbf{r})$ ，

$$C(\mathbf{r}) = \int_{t_1}^{t_2} T(t) \cdot \sigma(\mathbf{r}(t)) \cdot \mathbf{c}(\mathbf{r}(t), \mathbf{d}) \cdot dt, \quad (2)$$

其中  $T(t)$  是累积透射率 (accumulated transmittance)，表示光线从  $t_1$  传播到  $t$  而未被拦截的概率，由下式给出：

$$T(t) = \exp \left( - \int_{t_1}^t \sigma(\mathbf{r}(u)) \cdot du \right) \quad (3)$$

跟踪相机光线  $C(\mathbf{r})$  通过待合成图像的每个像素，可以渲染新视角。这个积分可以用数值计算。最初的实现 [1] 和大多数后续方法中都使用了非确定性分层采样 (deterministic stratified sampling) 方法，其中光线被分成  $N$  个等间距的箱子，并从每个箱子中均匀地抽取一个样本。其后，式子 (2) 可以近似为

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N \alpha_i T_i \mathbf{c}_i, \text{ 其中 } T_i = \exp \left( - \sum_{j=1}^{i-1} \sigma_j \delta_j \right). \quad (4)$$

$\delta_i$  是样本  $i$  到样本  $i+1$  的距离。 $(\sigma_i, \mathbf{c}_i)$  是给定射线沿样本点  $i$  估计的密度和颜色，由 NeRF MLP 计算。 $\alpha_i$  是透明度/不透明度，在样本点  $i$  使用 alpha 混合 (alpha compositing) 得到，由下式给出，

$$\alpha_i = 1 - \exp(-\sigma_i \delta_i) \quad (5)$$

## 2.2 基于 NeRF 的研究进展

自 2020 年被提出以来，NeRF 的研究百家争鸣，已经成为三维视觉领域的基本研究范式之一，推动了三维视觉的重建、渲染、定位、生成、理解等任务的发展。下面将从几个研究最为广泛的方向进行调研总结，展示 NeRF 为计算机视觉和图形学领域带来的新鲜活力。

### 2.2.1 可泛化 NeRF 重建：利用扩散模型的泛化能力

扩散模型是一种深度生成模型，它通过模拟数据的扩散过程来生成复杂的数据分布。扩散模型的泛化能力体现在其能够生成在训练数据之外的、多样化且高质量的样本。这种能力源于模型设计中的随机扩散过程和逐步精化的生成策略。在训练期间，模型学习如何从一开始的随机噪声中逐渐构造出有意义的数据结构，这一过程涉及大量的数据样本和变化，从而让模型掌握了数据的丰富特性。因此，当面对新的或未见过的数据时，扩散模型能够利用其在训练期间学到的广泛知识来生成符合数据分布的新样本，展现出优异的泛化和创造能力。

近期在可泛化的神经辐射场 (NeRF) 重建领域，特别是结合扩散模型的研究中，已经取得了显著进展。这些研究主要集中在提高 3D 场景理解和生成的质量上。例如，Deceptive-NeRF<sup>[2]</sup> 通过从稀疏输入数据重

建粗略的 NeRF，再利用这个粗略 NeRF 渲染图像并生成伪观测数据，最后利用这些增强的输入训练一个精细的 NeRF 模型，旨在通过扩散模型生成的伪观测数据来提高重建的质量。SSDNeRF<sup>[3]</sup> 则提出了一种统一的方法，使用一个表达能力强大的扩散模型直接从多视图图像中学习，简化了学习过程并可能提高学习到的 NeRF 表达的泛化性。而 Triplane 与 Gaussian Splatting 的结合<sup>[4]</sup> 则试图通过将分数蒸馏采样 (SDS) 与 3D 领域的扩散模型相结合来加速重建过程并使其更具泛化性，旨在解决单图像 3D 重建中常遇到的优化和渲染过程缓慢的问题。

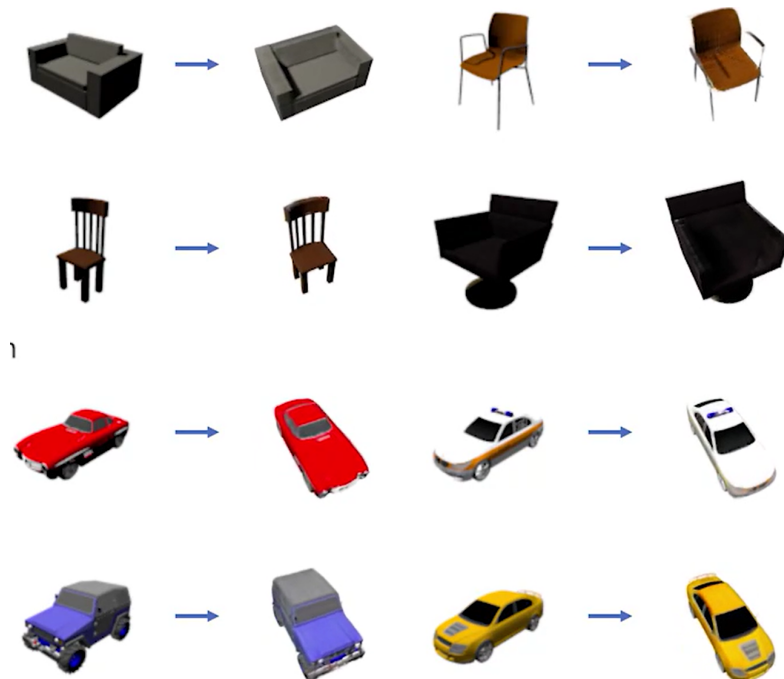


Fig. 2 SSDNeRF 单视角下的生成效果

这些进展展示了将扩散模型与 NeRF 结合的趋势，旨在提高 3D 重建的泛化性、细节和效率。这些技术能够从多样化甚至有时是稀疏的数据集中学习，每种方法都提出了独特的解决方案，旨在克服 3D 重建和表征的固有挑战。随着研究的深入，这些方法有望在未来的 3D 视觉和图形处理中发挥更大的作用。

### 2.2.2 NeRF 稀疏视点人脸重建：引入先验优化质量

在稀疏视点人脸重建领域，NeRF 技术因其在特定视点上的过拟合倾向而面临挑战，这导致新视点合成时出现伪影问题。为了克服这一难题，研究者们寻求通过引入大量人脸数据、面部关键点和模板等先验知识来优化 NeRF 的重建质量。这些方法不仅增强了模型对面部结构的理解，也为精确捕捉细节提供了必要的信息支持。

在这一背景下，诸如 LP3D<sup>[5]</sup>、HAvatar<sup>[6]</sup> 和 NeRSemble<sup>[7]</sup> 等方法被提出。LP3D 利用 EG3D 生成的人脸数据进行训练，并通过输入单一图像来推理三平面表达的 NeRF，旨在实现静态和实时的高质量人脸重建。HAvatar 则采用 3DMM 投影的三平面神经辐射场约束，通过精细调控面部特征和动态，实现更为自然的人头动态化身。NeRSemble 更进一步，通过引入 3DMM 表情参数并构建带表情语义空间变形场，致力于拟合复杂的表情动态，从而在动态面部表现上迈出了重要的一步。这些方法的提出和发展，不仅提高了稀疏视点下的重建质量，也为未来的面部动态模拟和虚拟现实技术开辟了新的可能。结合大规模人脸数据集、

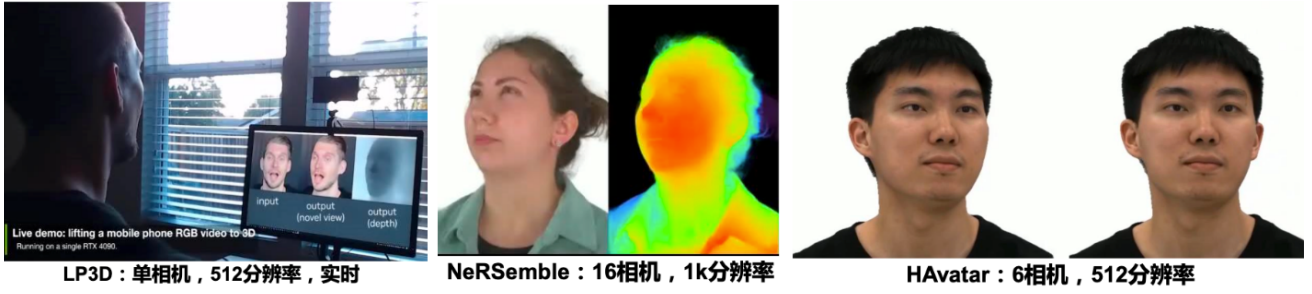


Fig. 3 人脸重建时三种方法的设备环境

关键点、模板以及如 3DMM 这样的高级建模技术，为 NeRF 框架在面对稀疏视点人脸重建时的困难提供了新的解决路径。LP3D、HAvatar 和 NeRSemble 这些方法的发展，预示着在虚拟现实、远程呈现和数字娱乐等领域，我们将能够拥有更加逼真和富有表现力的虚拟互动体验。尽管仍存在挑战，如过拟合、细节捕捉和计算效率等，但这些技术的不断进步和优化无疑将为未来的人脸重建技术带来更加广阔的应用前景。

### 2.2.3 NeRF 的其他研究方向

除了以上详细叙述的两个方面之外，NeRF 的许多其他方面同样被广泛研究。如效率速度、表征增强、动态建模、大场景建模、生成与编辑等，这些方面的进步为 NeRF 的实际应用提供了坚实的基础和广阔的前景。

## 2.3 NeRF 发展的未来展望

NeRF 技术作为一种高效的 3D 场景渲染方法，近年来在视觉领域取得了显著的进展。从最初的静态场景建模到现在对动态场景的深入探索，NeRF 正在迅速进化，拓宽其应用范围。以下是对 NeRF 未来展望的分析：

**高质量动态建模：**随着技术的不断进步，NeRF 在处理动态场景方面的能力有了显著提升。特别是在人脸和人体建模方面，研究者们通过优化算法和结构，实现了在保持实时性的同时提高渲染质量。这一趋势预示着 NeRF 将在虚拟现实、游戏、影视制作等领域有更广泛的应用，为用户提供更加逼真和丰富的交互体验。

**与大模型的结合：**将 NeRF 与生成式大模型结合，为 NeRF 打开了创造力的大门。这种结合不仅扩展了 NeRF 的应用场景，使其能够创造出原本不存在的物体和场景，而且还增强了模型的理解和表达能力。在未来，我们可以期待 NeRF 与大模型的结合将催生更多创新应用，例如自动生成或增强艺术作品、设计虚拟环境等。

**更丰富的信息嵌入：**传统的 NeRF 主要关注几何和纹理的建模。然而，随着研究的深入，将材质属性和语义信息等更多元的数据嵌入到 NeRF 中，可以显著提高模型的理解深度和渲染质量。这种丰富的信息嵌入不仅能够提升视觉效果，还能拓宽 NeRF 的应用场景，如在教育、设计和文化遗产保护中重现和解释复杂场景。

**应用到其他领域：**NeRF 技术的“破圈”行动使其在机器人、自动驾驶、医疗等领域展现了巨大的潜力。在这些领域中，NeRF 的新视点生成能力不仅能够辅助数据生成和场景理解，还能提高决策系统的精确性和效率。未来，我们可以预见 NeRF 将在这些领域中发挥更大作用，从而推动相关技术和应用的发展。

### 3 调研心得

随着我对 NeRF 技术的深入调研，我越来越认识到这项技术不仅仅是视觉计算领域的一个进步，更是一场潜在的革命。NeRF 通过其独特的方式捕捉和渲染 3D 世界，为我们提供了一个全新的视角来理解和再现现实。随着技术的不断发展和完善，我相信 NeRF 将在多个方面产生深远的影响。

传统的 3D 建模和渲染技术虽然能够提供相对准确的视觉复现，但往往缺乏细节和真实感。NeRF 通过高密度的数据采样和复杂的光照模型，能够生成极为逼真的场景，这不仅能够提升用户体验，还能够在模拟、训练和其他需要高度真实感的应用中发挥重要作用。

总之，NeRF 技术的未来充满了无限可能。随着算法的不断优化、计算资源的提升和跨学科合作的深入，NeRF 将在提高现实世界复现的质量、拓展创新应用和推动多领域技术融合方面发挥关键作用。作为一名初步学习和研究计算机视觉相关领域的学生，我有幸见证到 NeRF 技术带来的变革，不仅仅是在视觉领域，更在广泛的行业和应用中。

### 参考文献

- [1] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[A]. 2020. arXiv: 2003.08934.
- [2] LIU X, CHEN J, HONG KAO S, et al. Deceptive-nerf: Enhancing nerf reconstruction using pseudo-observations from diffusion models[A]. 2023. arXiv: 2305.15171.
- [3] RANADE S, LASSNER C, LI K, et al. Ssdnerf: Semantic soft decomposition of neural radiance fields[A]. 2022. arXiv: 2212.03406.
- [4] ZOU Z X, YU Z, GUO Y C, et al. Triplane meets gaussian splatting: Fast and generalizable single-view 3d reconstruction with transformers[A]. 2023. arXiv: 2312.09147.
- [5] TREVITHICK A, CHAN M, STENGEL M, et al. Real-time radiance fields for single-image portrait view synthesis [C]//ACM Transactions on Graphics (SIGGRAPH). 2023.
- [6] ZHAO X, WANG L, SUN J, et al. Havatar: High-fidelity head avatar via facial model conditioned neural radiance field[A]. 2023. arXiv: 2309.17128.
- [7] KIRSCHSTEIN T, QIAN S, GIEBENHAIN S, et al. Nersemble: Multi-view radiance field reconstruction of human heads[J/OL]. ACM Transactions on Graphics, 2023, 42(4): 1–14. <http://dx.doi.org/10.1145/3592455>.