有道翻译 输入arXiv论文检索地址，免费翻译全文→ https://fanyi.youdao.com/trans/#/home

arXiv: 2308.04413 v1 [cs。CV] 8月2023日

室外神经辐射场深度先验挖掘

陈王
cw.chenwang@outlook.com清华大学，百度研究
中国,北京

至于吴*
百度研究RAL，北京

，中国
西北工业大学，百度研究
中国,北京

沈Zhelun*
中国，北京，百度研

究中心
浙江大学，百度研究中心，中国北京

达扬吴
北京大学信息工程研究所，
中国科学院，北京

戴刘宇超(音)
西北工业
大学
西安,中国

陈良军入伙张
USA加州森尼维尔百
度研究中心

ACM参考格式:

王晨、孙嘉岱、刘丽娜、吴晨明、沈哲伦、吴大艳、戴玉超、张良军。2023。户外神经辐射场深度先验挖掘。第31届ACM多媒体国际会议论文集(MM'23)，2023年10月29日至11月3日，加拿大渥太华。ACM，USA纽约，10页。https://doi.org/10.1145/3581783.3612306

摘要

神经辐射场(nerf)在视觉和图形任务中表现出令人印象深刻的性能，例如新视图合成和沉浸式现实。然而，辐射场的形状-辐射模糊度仍然是一个挑战，特别是在稀疏视点设置中。最近的工作是将深度先验整合到户外NeRF培训中，以缓解这个问题。然而，选择深度先验的标准以及不同先验的相对优点还没有被彻底研究。此外，选择不同方法来使用深度先验的相对优劣也是一个尚未探索的问题。本文对将深度先验应用于室外神经辐射场进行了全面的研究和评估，涵盖了常见的深度感知技术和大多数应用方式。具体而言，我们使用两种具有代表性的NeRF方法，配备了四种常用的深度先验和不同的深度用法，在两个广泛使用的户外数据集上进行了广泛的实验。我们的实验结果揭示了几个有趣的发现，这些发现可能有助于从业者和研究人员使用深度先验训练他们的NeRF模型。项目页面:https://cwchenwang.github.io/outdoor-nerf-depth

CCS的概念

· 计算方法→计算机视觉;计算机图形学。

关键字

神经辐射场，深度估计，深度完成

1 介绍

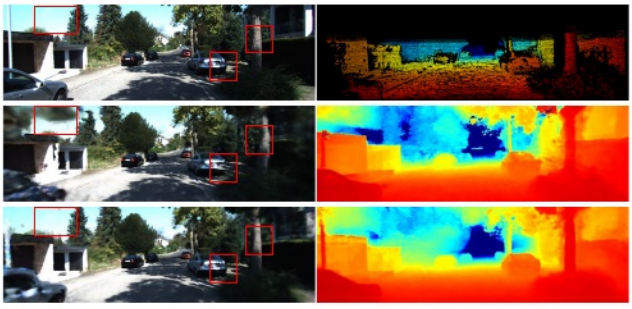

图1:从测试的角度来看，地面真值的图像和深度图可视化(上)，用纯RGB训练(中)，用单目深度估计训练(下)。即使使用单目深度(与其他深度相比，质量是最差的)，与仅使用RGB帧相比，NeRF的视图合成也可以在更少的漂浮物和更好地保存物体形状(汽车或树木)方面得到显著改善。

新视图合成，即在任意视点合成逼真的图像，是计算机视觉和多媒体领域的一项长期任务。最近，神经辐射场(neural radiance fields, NeRF)[35]及其变体已成为一种新的视图合成方法，并取得了令人印象深刻的性能。具体来说，NeRF通过一个连续函数来表示一个3D场景，该函数以一对3D位置和2D观看方向作为输入来预测RGB颜色和体积密度。这使我们能够使用标准的体绘制方程[19]来渲染图像。来自NeRF的逼真效果图

*Corresponding authors ({wuchenming, shenzhelun}@baidu.com).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM' 23, October 29-November 3, 2023, Ottawa, ON, Canada
©2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0108-5/23/10...$15.00 https://doi.org/10.1145/3581783.3612306

激发了多媒体领域的大量近期工作[54,55,60,63,69,73]。

尽管NeRF的发展令人印象深刻，但从辐射场中定义合理的底层几何结构仍然是一个未解决的问题。为了解决这个问题，一些作品使用距离场[30,56]来明确定义NeRF框架作品的几何形状。然而，这些方法需要大量来自不同视角的输入图像(通常需要360度捕获)才能成功重建一个物体。在户外场景中，存在着许多前景和背景物体，完全捕捉所有物体是一项艰巨的任务。使用几何先验，特别是深度先验，来促进室外NeRF训练是必要的，并且在以前的工作中被证明是有效的[34,42,58,61]。具体来说，原始激光雷达深度、深度完成和深度估计是常用的深度先验。由于它们的代价不同，精度也不同，因此有必要进一步研究选择深度先验的标准以及它们的相对优点。

在本文中，我们对将不同模态输入(即深度先验)融合到室外神经辐射场进行了全面的研究和评估，涵盖了神经辐射场全天研究中常见的深度感知技术和大多数应用方式。具体来说，室外场景中的深度感知可分为两类:(1)主动深度感知:利用光学器件获取深度的方法。一个主要的限制是光学设备，如光探测和测距(激光雷达)是累赘的，只能提供稀疏的测量。因此提出了深度补全，从稀疏深度图恢复稠密深度图。(2)被动深度感知:直接从图像推断深度的方法，这是一个便宜得多的选择，但牺牲了准确性。单目深度估计和双目深度估计是常用的两种方法。在实验上，我们选择了两种具有代表性的NeRF方法，并对两个流行的户外自动驾驶数据集KITTI[13]和Argoverse[7]使用不同的深度监督和不同的损失函数对其进行增强。因此，我们总结了实验结果，并有如下有趣的发现:

- 密度:即使是非常稀疏的深度监督也可以显著提高视图合成质量，通常密度越高越好。
- 质量:(1)单目深度足以满足稀疏视图输入，甚至可以达到与地面真值深度监督相当的结果;(2)深度监督是密集视图的一种选择，也就是说，如果相应的应用程序需要所使用的NeRF具有更好的几何形状，例如3D重建，则必须进行深度监督。
- 监督:室外NeRF不需要复杂的深度滤波和损失函数，直接用MSE监督裁剪天空区域就足够了。

据我们所知，我们的工作是第一次将深度先验应用于室外神经辐射场的定量和定性比较，我们相信我们的发现将有助于从业者和研究人员更全面地了解如何有效地将深度先验应用于室外nerf的训练。

2 相关工作

神经辐射场。神经辐射场(Neural radiance fields, nerf)[36]通过多层感知器(MLP)预测3D场景的逐点颜色和亮度，在新型视图合成中显示出卓越的有效性。然而，香草NeRF需要数小时的优化，并假设静态场景以及密集的视点。以下工作在不同方面对其进行了扩展，例如建模动态和可变形场景[40]，超分辨率[55]，稀疏或不完美的姿态[26]，对目标场景的泛化[68]和快速优化[12]。为了在无界室外场景中使用NeRF，NeRF++[74]引入了倒球面参数化来处理无界场景。MipNeRF-360[2]用逆深度间距重新参数化它们的场景坐标，在无界区域实现均匀间隔的光线间隔。对于大型户外场景，Block-NeRF[50]将场景分解为多个block，并单独训练NeRF。最近的研究[59,67]验证了户外NeRF在自动驾驶仿真中的适用性。

神经表示的深度监督。虽然香草NeRF只需要RGB图像进行训练，但当输入视点稀疏时，由于形状-亮度模糊，优化很容易陷入局部最小值。在这种情况下，额外的深度监督被发现是有用的。DS-NeRF[10]首次证明了深度信息在具有稀疏输入的NeRF中的有效性，使用的是来自运动结构的粗点云。密集深度先验[43]训练了一个额外的网络用于深度补全和不确定性估计，证明了密集深度监督的有效性。对于街景，稀疏的Li-DAR观测可以被纳入到NeRF的地理尝试中[42,58,62]。除了使用真实深度外，现有工作还表明，估计的单目深度也有助于改进神经3D重建和视图合成。MonoSDF[70]用现成的预测器寻找单目深度线索可以提高神经表面重建的质量和优化时间。NICER-SLAM[77]还集成了单目深度，以促进室内场景SLAM中的制图过程。NoPe-NeRF[4]和Meuleman等[34]利用单目深度估计的点云重构NeRF，并从一系列帧中联合估计相机姿态。

深度恢复。目前的深度恢复技术大致可以分为三类:(1)深度补全。深度补全的目标是从稀疏深度图中恢复稠密深度图，例如从激光雷达获得的深度图。深度补全分为无引导方法和有引导方法。非引导方法[11,53,66]旨在用深度神经网络直接补全稀疏深度图。引导方法[21,28,65]使用RGB图像完成稀疏到密集的深度映射。提出了几种策略来提高深度补全的性能，如早期融合[32,41]，后期融合[29,51]，残差深度模型els[25,27]，以及基于空间传播网络的网络[9,39]。(2)单目深度估计。单目深度估计的目标是从单幅图像中估计出深度图。早期工作主要采用手工特征[18,44]来做单目深度估计，在复杂场景中往往失败。目前，基于学习的方法已经显示出其优越性，编码器-解码器网络[5,8,23,46]是该领域最常用的ar结构。(3)双目深度估计。双目深度估计，即立体匹配的目标是从一对立体图像中估计出视差/深度图。它是一个经典的

任务，并建立了著名的四步流水线[45]。基于早期学习的方法[31, 71]主要采用卷积网络代替传统管道中的一个步骤，即特征提取。GCNet[20]是一个突破，它首先提出了一个端到端的网络来模拟典型管道的步骤。然后，通过后续的方法提出更好的特征提取[24,38]、成本体量构建[16,17]、成本聚合[6,64,72]、视差计算[52,76]，进一步优化管道。

3 DEPTH-SUPERVISED削弱

如第1节所述，先前的工作已经证明深度先验有利于NeRF训练，特别是在室外场景中，并且已经提出了多种方法将深度先验合并到NeRF框架中。表1对现有的深度监督NeRF方法进行了分类，从表中可以得出两个结论:

- 多个深度先验已应用于深度监督NeRF方法。然而，所有这些方法都只测试一种深度先验，而不与其他方法进行比较。因此，选择深度先验的标准以及不同深度先验的相对优点还没有被深入研究。此外，在所有现有的工作中，都遗漏了一种可用的深度先验，即双目深度估计。
- 现有的深度监督NeRF方法提出了多个损失函数，将深度先验合并到NeRF框架中。与前一个类似，选择不同的损失函数来使用深度先验的相对优点也是一个未被探索的问题。

因此，有必要对室外神经辐射场使用深度先验进行全面的研究和评估。具体来说，我们将在本节中对当前深度监督NeRF方法和使用的深度先验进行分类。

3.1 深度监督NeRF的分类

NeRF[35]将3D场景表示为一个连续函数，该函数将3D位置x∈R3和2D视图方向d∈R2映射到亮度颜色c∈R3和密度θ∈R。该函数通常用MLP进行参数化F_θ:(x, d)→(c，θ)。为了渲染一个图像I，我们沿着每个相机光线r(T)= o+Td整合颜色，这些光线从相机中心o拍摄，方向为d，用体渲染:

$$\mathbf{I}_\theta(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})\mathrm{d}t, \tag{1}$$

T(T)= exp(− ∫ TTσ(r(T))dT)表示累积反式-

N

密度表示光线从$_{TN}$to$_T$without到达任何粒子的概率。与颜色类似，NeRF中的深度图可以渲染如下:

$$\mathbf{D}_\theta(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))t\mathrm{d}t. \tag{2}$$

给定一组摆好姿势的图片I ={I_I|I= 0,1，⋯，通过比较两者之间的均方误差(MSE)来优化香草NeRF

rg

b

渲染图像和地面实况:LMSE =我NI| |II−I⊠I| | 22。

后续工作为NeRF培训增加了额外的深度信息。为了执行深度监督，NeRF-based方法

表1:现有深度监督NeRF方法的分类。

Methods	Depth priors	Loss Type
DS-NeRF [10]	SfM	KL
Urban-NeRF [42]	LiDAR	URF
S-NeRF [62]	Completion	L1
MonoSDF [70]	Mono	MSE
NICER-SLAM [77]	Mono	MSE
NoPe-NeRF [4]	Mono	L1
FEGR [58]	LiDAR	L1

表2:深度先验分类。(·)表示可选选择。

Depth priors	LiDAR	Camera	Cost	Density
Raw LiDAR depth	✔	✘	high	sparse
Depth completion	✔	(✔)	high	dense
Monocular depth estimation	✘	✔	low	dense
Binocular depth estimation	✘	✔	middle	dense

先采样一批$N_R$rays，再用不同的损失函数比较渲染深度和ground truth深度。现有的基于NeRF-based方法从直接和间接两个方面进行深度监督。下面我们将介绍每个类别的更多细节。

直接监督。直接监督直接使用监督损失将NeRF渲染的深度与之前的深度进行比较，包括MSE和L1:

$$\mathcal{L}^d_{\mathrm{MSE}} = \sum_i^{N_r} ||D(r_i) - \hat{D}(r_i)||^2, \tag{3}$$

$$\mathcal{L}_{\mathrm{L1}} = \sum_i^{N_r} |D(r_i) - \hat{D}(r_i)|, \tag{4}$$

其中$_{D(RI)}$和$_D$⊠$_{(RI)}$是rayRI的预测深度和ground truth$_{深度}$。L1损耗和MSE损耗都包括在我们的实验中。

间接监管。间接监督使用深度优先规则NeRF的权重，包括DS-NeRF中的KL损失[10]和Urban-NeRF中的URF损失[42]:

$$\mathcal{L}_{\mathrm{KL}} = \sum_i^{N_r}\sum_k \log w_k \exp(-\frac{(t_k - D(r_i))^2}{2\hat{\sigma}^2})\Delta t_k, \tag{5}$$

$$\mathcal{L}_{\mathrm{URF}} = \sum_{t=t_n}^{D(r_i)-\epsilon} w(t)^2 + \sum_{t=D(r_i)-\epsilon}^{D(r_i)+\epsilon} (w(t) - \mathcal{K}_\epsilon(t - D(r_i))^2, \tag{6}$$

其中$D_(R_I)$和$_D$⊠$_{(RI)}$为射线RW的预测真值深度和真实真值深度，$_{WK}$is为NeRF的渲染权值，$_{TK}$andΔT_Kare为射线的采样点和距离R_I，T(T)为距离点对应的权值，()为一个积分为1的核(即一个分布)，具有一个参数化为的有界域。由于URF损失没有开源实现，我们选择KL损失作为间接监督的代表。

3.2 深度先验的分类

如前所述，原始激光雷达深度、深度补全、单眼深度估计和双目深度估计是主要的

室外神经辐射场中使用的深度先验。当前深度先验的税收分类结果如表2所示，从这个表中我们可以得出两个观察结果:

- 原始激光雷达深度和深度完成是基于激光雷达的方法。这些方法的一个主要限制是，所采用的激光雷达是笨重的，只提供稀疏的测量。在此基础上，提出了深度补全，从稀疏深度图恢复稠密深度图。
- 单目深度估计和双目深度估计是基于相机的方法。这些方法可以直接从图像中得到一个密集的深度，这是一个便宜得多的选择，同时在精度上做出了很大的妥协。

下面我们将更详细地介绍每种方法。
原始激光雷达深度。原始激光雷达深度可以直接从所使用的光学设备，即激光雷达中获取。由于激光雷达只提供稀疏的深度测量，一些方法也选择将点云的多帧进行组合[14,33]，以获得更密集的深度图。

深度完成。表示输入的原始激光雷达深度为$D_L$and，对应的图像为I。深度补全过程可以表示为:

$$\begin{aligned} D_{guide} &= \delta(I, D_l), \\ D_{unguide} &= \delta(D_l), \end{aligned} \tag{7}$$

其中$D_{GUIDE}$and$D_{UNGUIDE}$denote分别是非引导方法和引导方法。目前，后者可以通过图像引导实现更高的精度。注意，图像信息也是NeRF的必要输入。因此，我们选择引导方法MFFNet[29]作为我们的深度补全方法。

单目深度估计。编码器-解码器网络[3,22,49]是这项任务最常用的架构。让我们将输入图像定义为$_I$。整个单目深度估计过程可以表示为:

$$D_{mono} = \varphi_d(\varphi_e(I)), \tag{8}$$

其中$\phi_E$denotes为编码器，$\phi_D$denotes为解码器。我们选择了一个具有代表性的编码器-解码器网络BTS[22]作为我们的单目深度估计方法。注意，这个方法是有监督的，输出具有正确的尺度。对于自监督或零样本预训练的单目深度估计模型，在训练过程中应该估计额外的尺度和漂移。

双目深度估计。特征提取、成本量构建、成本聚合和视差计算是目前深度立体匹配方法的典型流程。将输入的左右图像表示为$I_L$andI_R。整个双目深度估计过程可以表示为:

$$d = \eta(\delta(\partial(f_\theta(I_l), f_\theta(I_r)))), \tag{9}$$

其中$F_\theta$is为特征提取网络， is为成本量构建网络，δis为成本聚合网络，ηis为视差计算步骤。我们选择了两个具有代表性的立体匹配网络CFNet[47]和PCWNet[48]作为我们的双目深度估计方法。

4 实验与发现

在这一部分，我们介绍了本文的实验设置和结果。更多细节可在补充部分找到。

4.1 数据集

KITTI[13]和Argoverse[7]是真实室外驾驶场景的大型数据集。我们从KITTI里程计和Argoverse立体序列中选择片段来评估和比较这些方法。与NeRF中常用的以对象为中心的数据集相比，自动驾驶场景中的车辆通常只向前移动或转弯。为了减少光照变化和移动物体的影响，我们最终选择了KITTI中Seq 00、02、05、06中的5个序列(125、133、175、295、320帧)和Argoverse中的3个序列(73、72、73帧)。详情请参考补充材料。对于每个序列，我们保持每10帧作为测试集，其他用于训练。为了验证稀疏视点的影响，我们模拟了2.5 Hz的低频成像。为此，我们选择25%的KITTI训练数据，即每4个取1个。对于Argoverse数据集，我们选择50%的训练数据，因为它的数据记录频率(5 Hz)是KITTI (10 Hz)的1/2。我们还对所有训练数据(即密集视点)进行了实验。对于图像的位姿，为了避免SfM (structure from motion)位姿与真实深度尺度不一致，我们使用了KITTI odometry提供的位姿和Argoverse提供的跟踪位姿。

4.2 评价指标

4.2.1逼真度指标。我们使用新视图合成文献中的常见指标来比较测试视点下的合成视图与真实值:PSNR, SSIM[57]和LPIPS[75]。

4.2.2深度精度指标。继之前的工作[15,22]之后，我们使用ABSREL(平均绝对相对误差)和RMSE(均方根误差)作为我们的深度评价指标。

4.3 包括NeRF基线

NeRF的原始参数化只能处理有界或面向前方的场景。由于我们处理的是无界的真实场景，我们在实验中选择了以下NeRF变体。
NeRF++[74]将无界场景划分为两个体量:内部单位球体和外部体量。因此，体绘制也由两部分组成。我们使用Eq(2)的扩展版本在NeRF++中渲染深度:

$$\begin{aligned} \mathbf{D}(\mathbf{r}) = &\int_{t=0}^{t'} \sigma(\mathbf{r}(t)) t \cdot e^{-\int_{s=0}^{s} \sigma(\mathbf{r}(s))\mathrm{d}s} \mathrm{d}t \quad + \\ &e^{-\int_{s=0}^{t'} \sigma(\mathbf{r}(s))\mathrm{d}s} \cdot \int_{t=t'}^{\infty} \sigma(\mathbf{r}(t)) t \cdot e^{-\int_{s=t'}^{s} \sigma(\mathbf{r}(s))\mathrm{d}s} \mathrm{d}t, \end{aligned} \tag{10}$$

其中$_T$∈(0，T')是球体内部，$_T$∈(T'，∞)是无界区域。

MipNeRF-360[2]是对MipNeRF[1]的无界扩展，提出用契约函数对半径为2的球内的三维欧氏空间进行参数化。对于MipNeRF-360，我们可以直接使用Eq.(1)和Eq.(2)在收缩空间中渲染图像和深度图。

表3:在KITTI数据集上与选择方法的定量比较。最好和次好的结果分别以粗体和下划线的形式显示。

Methods	Depth Supervision	Dense PSNR↑	Dense SSIM↑	Dense LPIPS↓	Dense RMSE↓	Dense ABSREL↓	Sparse PSNR↑	Sparse SSIM↑	Sparse LPIPS↓	Sparse RMSE↓	Sparse ABSREL↓
MipNeRF-360 [2]	RGB-Only	**21.99**	**0.692**	**0.437**	3.090	0.088	16.93	0.589	0.498	4.662	0.144
	GT Depth	21.84	0.682	0.451	0.918	0.032	19.14	0.630	**0.474**	1.044	0.040
	Depth Completion	21.51	0.670	0.467	**0.818**	**0.026**	19.65	0.631	0.482	**1.030**	**0.032**
	Stereo Depth	21.53	0.665	0.469	1.192	0.030	**19.80**	**0.637**	0.475	1.246	0.034
	Mono Depth	21.48	0.668	0.468	2.161	0.059	19.35	0.625	0.485	1.890	0.058
NeRF++ [74]	RGB-Only	**20.29**	0.520	0.585	48.638	3.917	17.60	0.535	0.562	56.253	4.960
	GT Depth	20.08	0.574	0.563	**1.914**	**0.078**	**18.90**	**0.554**	**0.568**	**1.882**	**0.089**
	Depth Completion	20.15	**0.576**	**0.560**	2.618	0.102	**18.90**	0.553	0.569	2.022	0.094
	Stereo Depth	20.10	0.575	**0.560**	1.934	0.087	18.85	0.550	0.574	2.061	0.100
	Mono Depth	19.87	0.566	0.567	2.256	0.092	18.74	0.548	0.574	2.670	0.110
Instant-NGP [37]	RGB-Only	20.51	0.630	0.460	9.575	0.507	15.44	0.499	0.536	15.011	0.793
	GT Depth	**21.31**	**0.650**	**0.444**	**1.571**	0.052	18.53	**0.586**	**0.469**	**1.751**	0.060
	Depth Completion	20.90	0.632	0.470	1.661	**0.050**	**18.62**	0.576	0.492	1.833	**0.059**
	Stereo Depth	20.93	0.629	0.472	1.830	0.057	18.60	0.574	0.493	1.984	0.064
	Mono Depth	20.59	0.617	0.483	2.679	0.085	18.17	0.557	0.502	2.868	0.096

表4:选择序列对KITTI数据集深度先验的定量评价。评价指标的具体定义请参考之前的工作[15,22]。

Methods	$\delta < 1.25$↑	$\delta < 1.25^2$↑	$\delta < 1.25^3$↑	ABSREL↓	Sq Rel↓	RMSE↓	RMSE log↓	Density
Monocular Estimation	0.970	0.997	0.999	0.058	0.156	2.020	0.085	100%
Stereo Matching	0.996	0.998	0.999	0.016	0.035	1.080	0.040	100%
Stereo Matching_confidence	**0.999**	0.999	0.999	0.014	0.016	0.71	0.025	92.28%
Depth Completion	0.998	**1.000**	**1.000**	**0.010**	**0.015**	**0.622**	**0.020**	100%

Instant-NGP[37]提出了一种新的场景表示，它将实际场景边界到一个轴线对齐的边界框中，并使用一个小的神经网络，该网络由可训练特征向量的多分辨率哈希表增强，哈希表的值通过随机梯度下降进行优化。这些特征被进一步映射到颜色和密度。由于它仍然使用标准的体渲染，因此Instant-NGP中的深度可以类似于在香草NeRF中渲染(Eq.(2))。

4.4 评估结果和对比

在本节中，我们对Argoverse和KITTI数据集进行了实验，以验证采用不同深度先验的相对优点。下面我们将更详细地描述每个数据集的结果。

KITTI定性深度监督NeRF结果和相应的深度先验质量评价见表3和表4。如表4所示，不同深度先验之间的性能差距较大。具体来说，深度补全的精度最高，然后是双目深度估计和单目深度估计。下面我们将进一步分析密集和稀疏视图使用不同深度先验的相对优点。

(1)稀疏视图。我们首先讨论了在稀疏视角设置下的实验结果。如图所示，使用纯RGB训练的NeRF存在严重的形状-亮度模糊(图2)，因此在新视点下的视图合成质量显着下降。在这种情况下，深度信息的重要性是显而易见的，我们可以从tab中看到。3 .任何类型的深度都可以有利于并大大提高合成的视图。以MipNeRF-360为例，在任何类型的深度先验条件下，我们可以看到11.55% ~ 14.49%的逼真度度量改进(PSNR)和59.72% ~ 77.78%的深度精度度量改进(ABSREL)。此外，我们可以观察到，即使不同深度先验之间的深度质量差距很大，使用不同深度先验的真实感性能增益也很接近。也就是说，我们可以使用最便宜的深度先验(即单眼深度估计)来实现与昂贵的深度先验(例如，激光雷达收集的地面真实深度)相似的性能改进。在Instant-NGP上也可以观察到类似的情况。因此，我们可以得到我们的发现1:单目深度对于稀疏视图已经足够了。我们的第一个反直觉的发现是，使用单目深度估计可以显著提高NeRF的质量，甚至可以达到与稀疏视图下的地面真值深度监督相当的结果。一般来说，单目深度估计是一种更便宜的选择，不需要额外的设备，例如激光雷达。因此，单目深度估计在稀疏视图中是一个更好的选择，如果需要更好的深度图质量，双目深度估计也是一个选择。

(2)密集视点。然后讨论了在密集视角设置下的实验结果。如图所示，深度监督对深度精度指标也有帮助。以MipNeRF-360为例，在任何类型的深度先验下，我们都可以看到32.96% ~ 70.45%的深度精度指标改进(ABSREL)。这就是深度先验

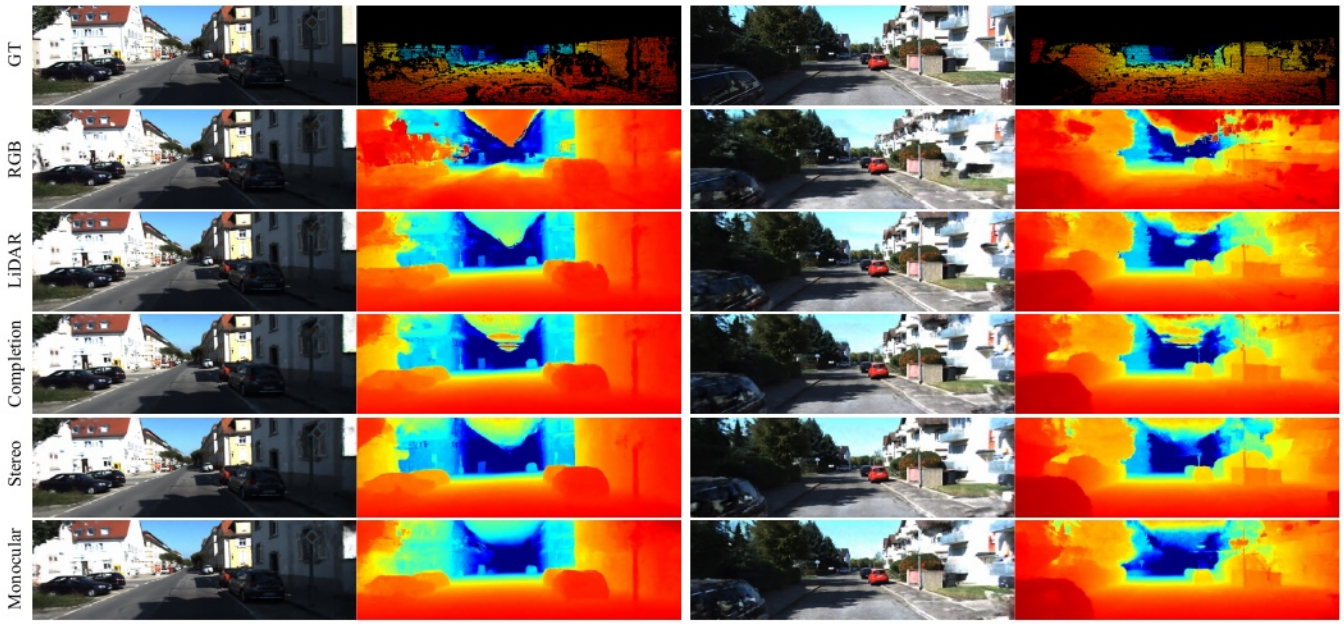

图2:MipNeRF-360稀疏视点下KITTI数据集的定性结果。与RGB训练相比，增加深度监督能显著提高训练质量。更好的查看放大和彩色。

表5:MipNeRF-360和Instant-NGP在Argoverse数据集上的评估和比较。最好和次好的结果分别以粗体和下划线的形式显示。

Methods	Depth Supervision	Dense PSNR↑	Dense SSIM↑	Dense LPIPS↓	Dense RMSE↓	Dense ABSREL↓	Sparse PSNR↑	Sparse SSIM↑	Sparse LPIPS↓	Sparse RMSE↓	Sparse ABSREL↓
MipNeRF-360 [2]	RGB-Only	**29.35**	**0.855**	**0.446**	6.113	0.120	25.81	0.829	0.468	6.971	0.139
	GT Depth	28.78	0.846	0.458	**2.251**	**0.044**	28.01	**0.840**	**0.462**	**2.443**	**0.048**
	Stereo Depth	28.32	0.837	0.470	4.271	0.064	27.72	0.833	0.471	4.310	0.067
	Mono Depth	28.58	0.841	0.466	4.601	0.093	**28.04**	0.836	0.468	4.868	0.093
Instant-NGP [37]	RGB-Only	28.07	**0.847**	**0.450**	13.478	0.493	22.18	0.816	0.494	17.439	0.593
	GT Depth	**28.92**	**0.847**	**0.450**	**1.804**	**0.045**	**27.38**	**0.834**	**0.460**	**1.881**	**0.048**
	Stereo Depth	28.32	0.839	0.460	5.613	0.090	27.10	0.828	0.467	5.843	0.097
	Mono Depth	28.31	0.838	0.466	6.083	0.122	26.97	0.829	0.471	6.643	0.132

仍然是辐射场获得合理的基础几何形状的必要条件。另一方面，在逼真度量中的性能增益并不那么值得注意(Instant-NGP)，甚至会导致某些方法(MipNeRF-360)的性能下降。我们将性能下降归因于收缩函数下深度和RGB的优化方向不一致，因为Instant-NGP和MipNeRF-360之间的主要区别之一是使用无界收缩。因此，我们可以得到我们的发现2:深度监督是密集视图的一个选项。我们第二个有趣的发现是，使用深度监督可以在密集视图中实现显著的几何改进和微不足道的逼真度量改进。因此，深度监督是密集视图中的一个选项。如果相应的应用程序需要所使用的NeRF具有更好的几何形状，例如重建和潜在的重新照明，阴影等，则仍然有必要。

我们还在Argoverse数据集上进行了实验，以进一步验证我们的说法。请注意，Argoverse数据集不提供深度补全任务，因此我们没有在实验中包含此设置。定性结果可以在表5中找到。如图所示，新视点下的视图合成质量在稀疏视图下也会显著下降，任何深度都可以大大提高合成视图。我们还观察到密集视图设置的深度精度指标有显著提高，以及微不足道的逼真度指标。KITTI数据集的情况也是如此，这进一步支持了我们的发现1和2的有效性。

4.5 烧蚀研究

虽然第4.5节中的实验结果已经证明了采用不同深度先验的相对优点，但仍然有许多基础设置可以揭示更深刻的发现，如不同深度密度、深度范围、置信度、深度损失函数等的影响。在本节中，我们进行

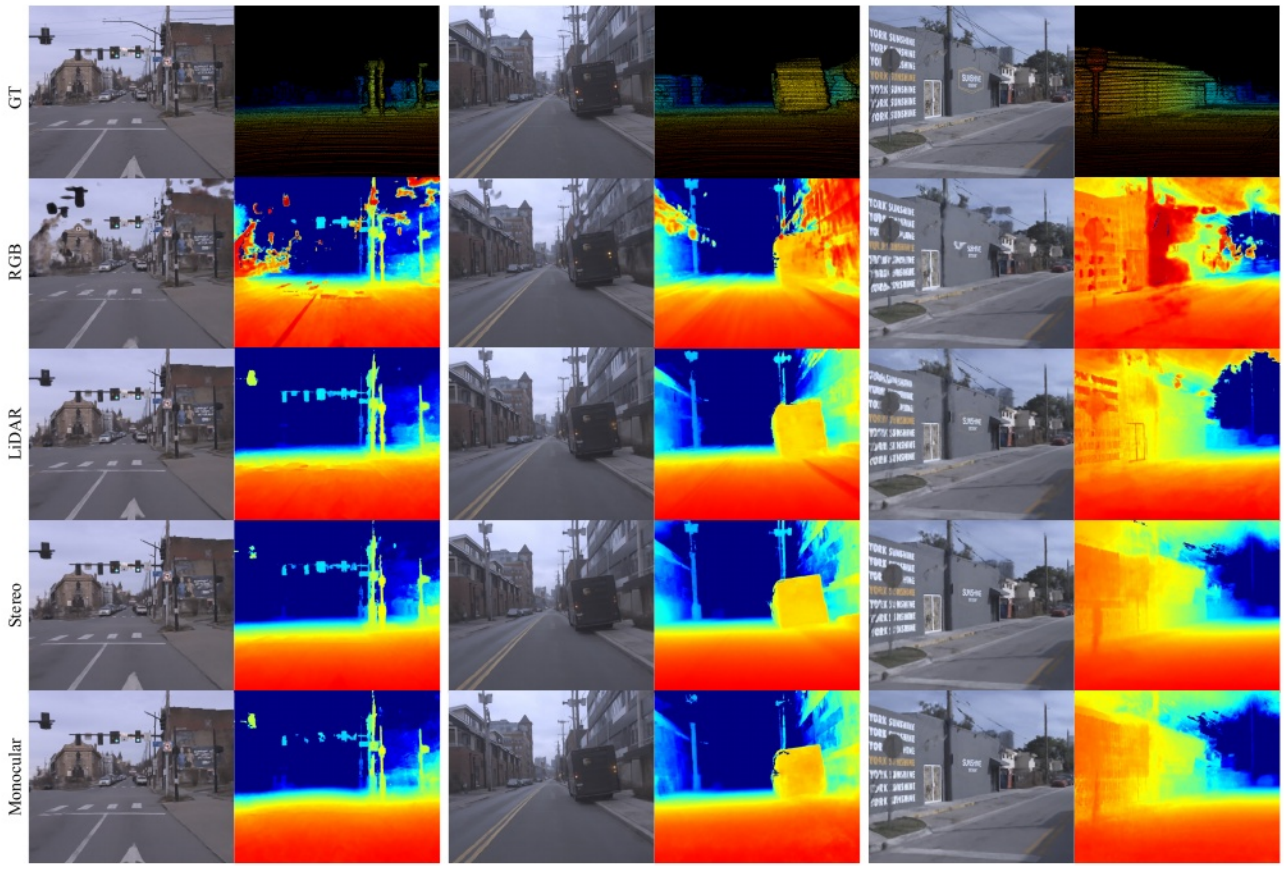

图3:在MipNeRF-360稀疏视点的Argoverse数据集上的定性结果(GT深度用3 ×3内核进行扩展，以便更好地可视化，这是非常稀疏的)。与RGB训练相比，增加深度监督能显著提高训练质量。更好的查看放大和彩色。

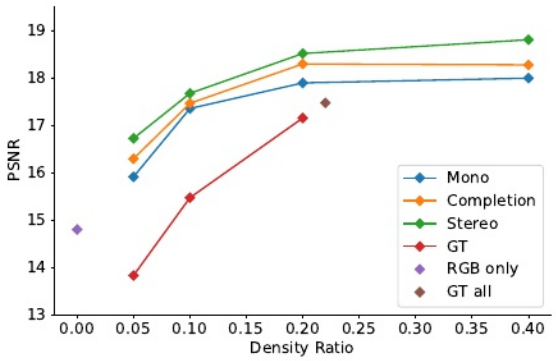

图4:不同类型深度先验下PSNR与深度监督密度比的关系。

利用MipNeRF-360和稀疏视图设置对KITTI数据集的一个序列进行多次消融实验，以进一步探索这些因素。

4.5.1密度。密度是区分激光雷达地面真值与其他深度先验值的主要因素，例如KITTI数据集上激光雷达地面真值的密度接近20%，而其他深度先验值的密度为100%。我们进行了两个消融实验，以进一步研究不同密度的影响。

GT掩蔽是有趣的，单目深度估计可以达到相当甚至更好的性能，比地面真实激光雷达在真实感指标。这可能是因为单目深度比激光雷达密度大得多，后者只能提供大约20%像素的有效深度gt。为了验证这一假设，我们尝试使用地真激光雷达的无效位置来掩盖单目深度估计结果，并确保两个深度先验具有相同的密度。结果表明，掩模后使用单目深度的真实感性能下降到接近或低于使用地面真值深度。这表明，即使不精确的密集监督在稀疏视图中对于新视图合成仍然可以比稀疏监督更有益。

深度密度为了进一步评估监督稀疏性的影响，我们还通过迭代地从原始深度先验中删除固定比例的像素来测试不同的深度密度。图4的结果显示，少量(5%)的深度监督

足以提高NeRF的性能，更密集的深度监督可以获得更可观的收益。具体来说，在20%之前提升更显著，在足够的监督(40%)之后，结果变得稳定。此外，我们还观察到，即使随机选择不到20%的单目深度监督(或深度补全和立体深度)，其结果也已经超过了使用所有地面真值(22%)，这表明在稀疏视图下，覆盖范围更广的不精确监督(如单目深度估计)优于覆盖中心区域的精确监督(如LIDAR)。GT掩蔽实验进一步支持了我们的主张，即稀疏且覆盖同一区域的单目深度估计无法击败LIDAR的地面真值监督。

综合以上实验，我们可以得到我们的发现3:密度越高越好。我们第三个有趣的发现是，即使是非常稀疏的深度监督，也可以显著提高稀疏视图中新视图合成的质量，而且一般来说，我们观察到深度越密集，我们可以获得的新视图合成的质量越好。

4.5.2深度损失型。除了MSE，我们还测试L1和KL损耗。如表6所示，L1和MSE损耗之间没有显著的性能差异。此外，KL损失明显大于MSE，这可能是因为它是一个间接损失函数，在NeRF优化中有很强的约束，如果深度估计不够准确，会产生不利影响。

4.5.3深度滤波。不同方法的深度质量差异很大。一般情况下，质量从完成到立体，最后到单目(参见表4)。为了解决许多下游任务中深度质量的不稳定呈现，存在一些方法来过滤某些深度值，并只使用其中的子集。在本研究中，我们选择了两种简单且广泛使用的深度滤波方法，即阈值裁剪和基于置信度的滤波，来调查其影响。

阈值裁剪在表5和表3中，我们直接使用不同方法的深度预测(裁剪天空区域)。一般情况下，背景区域(远位置)的精度要低于前景区域(近位置)。在这里，我们测试过滤掉比阈值更远的预测，即比阈值大的预测的影响S。在我们的实验中，我们设置Sas 40M和80M。结果如表6所示。我们可以看到，用S= 80M进行裁剪前后，图像和深度指标都没有显著差异。对于S= 40M，图像和深度指标略有下降，可能是因为它丢失了背景区域的深度信息，尽管它的准确性不如前景区域。

基于置信度的滤波置信度估计也是深度估计和深度监督NeRF中常见的操作。在这里，我们以双目深度估计为例测试这种置信度滤波器。具体来说，置信度滤波器是通过CFNet中提出的不确定性估计来实现的。对应的深度质量评价结果如表4所示。从表6中我们可以看到，类似于阈值裁剪，经过滤波后，性能没有明显变化。

表6:带有GT掩蔽、阈值裁剪、置信度过滤和损失函数的消融研究。

Experiments	Factor	Depth Type	PSNR↑	SSIM↑	LPIPS↓	RMSE↓	ABSREL↓
-	RGB-Only	-	14.80	0.475	0.551	4.569	0.153
GT Masking	-	LiDAR	17.47	**0.542**	**0.507**	**1.173**	**0.045**
	Yes	Mono	17.11	0.535	0.512	2.345	0.076
	No	Mono	**17.97**	**0.542**	0.510	2.383	0.073
Threshold Clipping	-	Mono	17.97	**0.542**	0.510	**2.383**	0.073
	40m	Mono	17.60	0.541	**0.509**	2.470	**0.071**
	80m	Mono	**18.18**	**0.542**	0.510	2.390	0.073
Confidence Filtering	No	Stereo	**18.87**	0.562	0.501	**1.349**	**0.0405**
	Yes	Stereo	18.85	**0.565**	**0.495**	1.467	0.0424
Loss Function	MSE	Mono	**17.97**	**0.542**	**0.510**	**2.383**	**0.073**
	L1	Mono	17.91	0.519	0.550	2.543	0.075
	KL	Mono	16.55	0.526	0.515	2.487	0.076

发现4:简单的损失函数和深度滤波就足够了。以上三个消融研究导致了我们的第四个发现:在室外NeRF中不需要复杂的深度滤波和损失函数，在MSE监督下直接裁剪天空区域(无限远点)就足够了。

4.6 讨论

与DS-NeRF类似[10]，我们也通过训练Instant-NGP进行了30次epoch的实验，发现额外的深度加快了收敛速度，并且在极端少镜头的情况下也是有益的。在稀疏设置下，具有深度的训练优于仅具有30个epoch的RGB训练，即使在第一个epoch也是如此。对于极度稀疏的视图，我们进一步选择1/ 8, 1 /16的输入视图在Kitti数据集中进行训练。得到的PSNR/ LPIPS为13.1/0.57 (RGB Only)， 16.4/0.52(单眼深度)为1/8,11.6/0.61 (RGB Only)， 13.0/0.57(单目深度)为1/16。使用其他深度监督甚至比使用单目深度更好。

5 结论

本文首次对室外神经辐射场应用深度先验进行了深入研究和评估，涵盖了所有常见的深度感知技术和大多数应用方式。因此，我们总结了实验结果，并有以下有趣的发现:(1)密度:即使是非常稀疏的深度监督也可以显著提高视图合成质量，通常越密集越好;(2)质量:(a)对于稀疏视图，单目深度足够，甚至可以达到与地面真值深度监督相当的效果。(b)深度监督是密集视野的一个选择，即，如果相应的应用需要所采用的NeRF具有更好的几何形状，则深度监督是必要的;(3)监督:室外NeRF不需要复杂的深度滤波和损失函数，直接裁剪天空区域进行MSE监督就足够了。我们相信这些发现可以潜在地帮助从业者和研究人员在训练他们的NeRF模型时使用深度先验。

致谢

本研究得到国家自然科学基金项目(62106258和62271410)的部分资助。

参考文献

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-nerf: A multiscale representa-tion for anti-aliasing neural radiance fields. In Proc. of the IEEE/CVF Intl. Conf. on Computer Vision (ICCV).

[2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[3] Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. 2021. Adabins: Depth estimation using adaptive bins. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). 4009–4018.

[4] Wenjing Bian, Zirui Wang, Kejie Li, Jia-Wang Bian, and Victor Adrian Prisacariu. 2022. NoPe-NeRF: Optimising Neural Radiance Field with No Pose Prior. arXiv preprint arXiv:2212.07388 (2022).

[5] Yuanzhouhan Cao, Yidong Li, Haokui Zhang, Chao Ren, and Yifan Liu. 2021. Learning Structure Affinity for Video Depth Estimation. In Proceedings of the 29th ACM International Conference on Multimedia. 190–198.

[6] Jia-Ren Chang and Yong-Sheng Chen. 2018. Pyramid stereo matching network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 5410–5418.

[7] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. 2019. Argoverse: 3d tracking and forecasting with rich maps. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). 8748–8757.

[8] Zhi Chen, Xiaoqing Ye, Liang Du, Wei Yang, Liusheng Huang, Xiao Tan, Zhenbo Shi, Fumin Shen, and Errui Ding. 2021. AggNet for Self-supervised Monocular Depth Estimation: Go An Aggressive Step Furthe. In Proceedings of the 29th ACM International Conference on Multimedia. 1526–1534.

[9] Xinjing Cheng, Peng Wang, and Ruigang Yang. 2018. Depth estimation via affinity learned with convolutional spatial propagation network. In Proc. Eur. Conf. Comput. Vis. (ECCV). 103–119.

[10] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. 2022. Depth-supervised nerf: Fewer views and faster training for free. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[11] Abdelrahman Eldesokey, Michael Felsberg, Karl Holmquist, and Michael Persson. 2020. Uncertainty-aware cnns for depth completion: Uncertainty from beginning to end. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). 12014–12023.

[12] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields without Neural Networks. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[13] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[14] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3354–3361.

[15] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J Brostow. 2019. Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE International Conference on Computer Vision. 3828–3838.

[16] Xiaodong Gu, Zhiwen Fan, Siyu Zhu, Zuozhuo Dai, Feitong Tan, and Ping Tan. 2020. Cascade cost volume for high-resolution multi-view stereo and stereo matching. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2495–2504.

[17] Xiaoyang Guo, Kai Yang, Wukui Yang, Xiaogang Wang, and Hongsheng Li. 2019. Group-wise correlation stereo network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3273–3282.

[18] Jose L Herrera, Carlos R Del-Blanco, and Narciso Garcia. 2018. Automatic depth extraction from 2D images using a cluster-based learning framework. IEEE Trans. on Image Processing (TIP) 27, 7 (2018), 3288–3299.

[19] James T Kajiya and Brian P Von Herzen. 1984. Ray tracing volume densities. ACM SIGGRAPH computer graphics 18, 3 (1984), 165–174.

[20] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. 2017. End-to-end learning of geometry and context for deep stereo regression. In IEEE International Conference on Computer Vision (ICCV). 66–75.

[21] Md Fahim Faysal Khan, Nelson Daniel Troncoso Aldas, Abhishek Kumar, Sid-dharth Advani, and Vijaykrishnan Narayanan. 2021. Sparse to dense depth completion using a generative adversarial network with intelligent sampling strategies. In Proceedings of the 29th ACM International Conference on Multimedia. 5528–5536.

[22] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. 2019. From big to small: Multi-scale local planar guidance for monocular depth estimation. arXiv preprint arXiv:1907.10326 (2019).

[23] Rui Li, Xiantuo He, Yu Zhu, Xianjun Li, Jinqiu Sun, and Yanning Zhang. 2020. Enhancing self-supervised monocular depth estimation via incorporating robust constraints. In Proceedings of the 28th ACM International Conference on Multimedia. 3108–3117.

[24] Zhengfa Liang, Yulan Guo, Yiliu Feng, Wei Chen, Linbo Qiao, Li Zhou, Jianfeng Zhang, and Hengzhu Liu. 2019. Stereo matching using multi-level cost volume and multi-scale feature constancy. In IEEE transactions on pattern analysis and machine intelligence (TPAMI).

[25] Yiyi Liao, Lichao Huang, Yue Wang, Sarath Kodagoda, Yinan Yu, and Yong Liu. 2017. Parse geometry from a line: Monocular depth estimation with partial laser observation. In Proc. IEEE Int. Conf. Robot. Automat. (ICRA). 5059–5066.

[26] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. 2021. Barf: Bundle-adjusting neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 5741–5751.

[27] Lina Liu, Yiyi Liao, Yue Wang, Andreas Geiger, and Yong Liu. 2021. Learning steering kernels for guided depth completion. IEEE Trans. on Image Processing (TIP) 30 (2021), 2850–2861.

[28] Lina Liu, Xibin Song, Xiaoyang Lyu, Junwei Diao, Mengmeng Wang, Yong Liu, and Liangjun Zhang. 2021. Fcfr-net: Feature fusion based coarse-to-fine residual learning for depth completion. In Proceedings of the AAAI conference on artificial intelligence, Vol. 35. 2136–2144.

[29] Lina Liu, Xibin Song, Jiadai Sun, Xiaoyang Lyu, Lin Li, Yong Liu, and Liangjun Zhang. 2023. MFF-Net: Towards Efficient Monocular Depth Completion with Multi-modal Feature Fusion. IEEE Robot. Automat. Lett. (RA-L) (2023).

[30] Xiaoxiao Long, Cheng Lin, Lingjie Liu, Yuan Liu, Peng Wang, Christian Theobalt, Taku Komura, and Wenping Wang. 2023. NeuralUDF: Learning Unsigned Dis-tance Fields for Multi-view Reconstruction of Surfaces with Arbitrary Topologies. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[31] Wenjie Luo, Alexander G Schwing, and Raquel Urtasun. 2016. Efficient deep learning for stereo matching. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 5695–5703.

[32] Fangchang Ma, Guilherme Venturelli Cavalheiro, and Sertac Karaman. 2019. Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. In Proc. IEEE Int. Conf. Robot. Automat. (ICRA). 3288–3295.

[33] Moritz Menze and Andreas Geiger. 2015. Object scene flow for autonomous vehicles. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3061–3070.

[34] Andreas Meuleman, Yu-Lun Liu, Chen Gao, Jia-Bin Huang, Changil Kim, Min H Kim, and Johannes Kopf. 2023. Progressively Optimized Local Radiance Fields for Robust View Synthesis. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[35] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Proc. Eur. Conf. Comput. Vis. (ECCV).

[36] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. Commun. ACM (2021).

[37] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. ACM Trans. on Graphics (TOG) (2022).

[38] Guang-Yu Nie, Ming-Ming Cheng, Yun Liu, Zhengfa Liang, Deng-Ping Fan, Yue Liu, and Yongtian Wang. 2019. Multi-level context ultra-aggregation for stereo matching. In IEEE conference on computer vision and pattern recognition (CVPR). 3283–3291.

[39] Jinsun Park, Kyungdon Joo, Zhe Hu, Chi-Kuei Liu, and In So Kweon. 2020. Non-local spatial propagation network for depth completion. In Proc. Eur. Conf. Comput. Vis. (ECCV). 120–136.

[40] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Gold-man, Steven M Seitz, and Ricardo Martin-Brualla. 2021. Nerfies: Deformable neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 5865–5874.

[41] Chao Qu, Ty Nguyen, and Camillo Taylor. 2020. Depth completion via deep basis fitting. In Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV). 71–80.

[42] Konstantinos Rematas, Andrew Liu, Pratul P Srinivasan, Jonathan T Barron, Andrea Tagliasacchi, Thomas Funkhouser, and Vittorio Ferrari. 2022. Urban radiance fields. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[43] Barbara Roessle, Jonathan T Barron, Ben Mildenhall, Pratul P Srinivasan, and Matthias Nießner. 2022. Dense depth priors for neural radiance fields from sparse input views. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 12892–12901.

[44] Ashutosh Saxena, Min Sun, and Andrew Y Ng. 2008. Make3d: Learning 3d scene structure from a single still image. IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI) 31, 5 (2008), 824–840.

[45] Daniel Scharstein and Richard Szeliski. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision (IJCV) 47, 1-3 (2002), 7–42.

[46] Guibao Shen, Yingkui Zhang, Jialu Li, Mingqiang Wei, Qiong Wang, Guangyong Chen, and Pheng-Ann Heng. 2021. Learning regularizer for monocular depth estimation with adversarial guidance. In Proceedings of the 29th ACM International Conference on Multimedia. 5222–5230.

[47] Zhelun Shen, Yuchao Dai, and Zhibo Rao. 2021. Cfnet: Cascade and fused cost volume for robust stereo matching. In Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition. 13906–13915.

[48] Zhelun Shen, Yuchao Dai, Xibin Song, Zhibo Rao, Dingfu Zhou, and Liangjun Zhang. 2022. Pcw-net: Pyramid combination and warping cost volume for stereo matching. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII. Springer, 280–297.

[49] Minsoo Song, Seokjae Lim, and Wonjun Kim. 2021. Monocular depth estimation using laplacian pyramid-based depth residuals. IEEE transactions on circuits and systems for video technology 31, 11 (2021), 4381–4393.

[50] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. 2022. Block-nerf: Scalable large scene neural view synthesis. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[51] Jie Tang, Fei-Peng Tian, Wei Feng, Jian Li, and Ping Tan. 2020. Learning guided convolutional network for depth completion. IEEE Trans. on Image Processing (TIP) 30 (2020), 1116–1129.

[52] Stepan Tulyakov, Anton Ivanov, and Francois Fleuret. 2018. Practical deep stereo (pds): Toward applications-friendly deep stereo matching. Advances in neural information processing systems 31 (2018).

[53] Jonas Uhrig, Nick Schneider, Lukas Schneider, Uwe Franke, Thomas Brox, and Andreas Geiger. 2017. Sparsity invariant cnns. In Proc. of the Intl. Conf. on 3D Vision (3DV). 11–20.

[54] Chen Wang, Angtian Wang, Junbo Li, Alan Yuille, and Cihang Xie. 2023. Benchmarking robustness in neural radiance fields. arXiv preprint arXiv:2301. 04075 (2023).

[55] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. 2022. NeRF-SR: High Quality Neural Radiance Fields using Super-sampling. In Proc. of the ACM Intl. Conf. on Multimedia. (MM).

[56] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In Proc. of the Conference on Neural Information Processing Systems (NeurIPS).

[57] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13, 4 (2004), 600–612.

[58] Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. 2023. Neural Fields meet Explicit Geometric Representation for Inverse Rendering of Urban Scenes. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).

[59] Chenming Wu, Jiadai Sun, Zhelun Shen, and Liangjun Zhang. 2023. MapNeRF: Incorporating Map Priors into Neural Radiance Fields for Driving View Simulation. In IEEE International Conference on Intelligent Robots and Systems (IROS).

[60] Zijin Wu, Xingyi Li, Juewen Peng, Hao Lu, Zhiguo Cao, and Weicai Zhong. 2022. DoF-NeRF: Depth-of-Field Meets Neural Radiance Fields. In Proc. of the ACM Intl. Conf. on Multimedia. (MM). 1718–1729.

[61] Ziyang Xie, Junge Zhang, Wenye Li, Feihu Zhang, and Li Zhang. 2023. S-NeRF: Neural Radiance Fields for Street Views. In Proc. of the Int. Conf. on Learning Representations (ICLR).

[62] Ziyang Xie, Junge Zhang, Wenye Li, Feihu Zhang, and Li Zhang. 2023. S-NeRF: Neural Radiance Fields for Street Views. arXiv preprint arXiv:2303.00749 (2023).

[63] Wenpeng Xing and Jie Chen. 2022. MVSPlenOctree: Fast and Generic Reconstruction of Radiance Fields in PlenOctree from Multi-view Stereo. In Proc. of the ACM Intl. Conf. on Multimedia. (MM). 5114–5122.

[64] Haofei Xu and Juyong Zhang. 2020. AANet: Adaptive Aggregation Network for Efficient Stereo Matching. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1959–1968.

[65] Yan Xu, Xinge Zhu, Jianping Shi, Guofeng Zhang, Hujun Bao, and Hongsheng Li. 2019. Depth completion from sparse lidar data with depth-normal constraints. In Proc. of the IEEE/CVF Intl. Conf. on Computer Vision (ICCV). 2811–2820.

[66] Gengshan Yang, Joshua Manela, Michael Happold, and Deva Ramanan. 2019. Hierarchical deep stereo matching on high-resolution images. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). 5515–5524.

[67] Ze Yang, Yun Chen, Jingkang Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. 2023. UniSim: A Neural Closed-Loop Sensor Simulator. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1389–1399.

[68] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. 2021. pixelnerf: Neural radiance fields from one or few images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 4578–4587.

[69] Xianggang Yu, Jiapeng Tang, Yipeng Qin, Chenghong Li, Xiaoguang Han, Linchao Bao, and Shuguang Cui. 2022. PVSeRF: joint pixel-, voxel-and surface-aligned radiance field for single-image novel view synthesis. In Proc. of the ACM Intl. Conf. on Multimedia. (MM). 1572–1583.

[70] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. 2022. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. arXiv preprint arXiv:2206.00665 (2022).

[71] Jure Zbontar, Yann LeCun, et al. 2016. Stereo matching by training a convolutional neural network to compare image patches. J. Mach. Learn. Res. 17, 1 (2016), 2287–2318.

[72] Feihu Zhang, Victor Prisacariu, Ruigang Yang, and Philip HS Torr. 2019. Ga-net: Guided aggregation net for end-to-end stereo matching. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 185–194.

[73] Jiahui Zhang, Fangneng Zhan, Rongliang Wu, Yingchen Yu, Wenqing Zhang, Bai Song, Xiaoqin Zhang, and Shijian Lu. 2022. VMRF: View Matching Neural Radiance Fields. In Proc. of the ACM Intl. Conf. on Multimedia. (MM). 6579–6587.

[74] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. 2020. Nerf++: Analyzing and improving neural radiance fields. arXiv preprint arXiv:2010. 07492 (2020).

[75] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). 586–595.

[76] Youmin Zhang, Yimin Chen, Xiao Bai, Jun Zhou, Kun Yu, Zhiwei Li, and Kuiyuan Yang. 2019. Adaptive Unimodal Cost Volume Filtering for Deep Stereo Matching. In arXiv preprint.

[77] Zihan Zhu, Songyou Peng, Viktor Larsson, Zhaopeng Cui, Martin R Oswald, Andreas Geiger, and Marc Pollefeys. 2023. NICER-SLAM: Neural Implicit Scene Encoding for RGB SLAM. arXiv preprint arXiv:2302.03594 (2023).