arXiv:2602.01268v1 [cs.CV] 1 Feb 2026

OASIS-DC：通过稀疏融合单目伪深度的输出级对齐实现可泛化的深度补全

Jaehyeon Cho[1] 和 Jhonghyun An[1]

摘要—近期的单目基础模型在零样本深度估计方面表现出色，但其输出本质上是相对深度，而非度量深度，因此难以直接应用于机器人技术和自动驾驶。我们利用相对深度能够保留全局布局和边界这一特性：通过稀疏测距对其进行校准，将其转化为伪度量深度先验。基于这一先验，我们设计了一种细化网络，在先验可靠时遵循先验，在必要时偏离先验，从而仅凭极少量标注样本即可实现准确的度量深度预测。当缺少经过筛选的验证数据时，该系统尤其有效，能够在少样本设置下保持稳定的尺度和清晰的边缘。这些发现表明，将基础模型先验与稀疏锚点相结合，是在真实世界标注稀缺的情况下实现稳健、可部署深度补全的一种实用途径。

I. 引言

深度补全——从稀疏测距中推断稠密度量深度、并可选择由 RGB 提供引导的任务，是实现机器人技术和自动驾驶中稳健感知与规划的关键。尽管 KITTI 基准 [1], [2] 上的研究进展迅速，且强大的 RGB–激光雷达融合方法 [3], [4], [5], 不断涌现，现有流程仍默认能够获取大量标注数据，并投入大量精力筛选验证数据。然而，在实际部署场景中，运行条件（城市、路线、传感器设备）变化的速度往往快于数据标注速度；模型必须仅凭极少量样本实现泛化，并尽量减少验证工作。与此同时，近期的基础模型引导方法往往将深度补全与单目基础深度估计器（MDE）的高维特征耦合起来，例如在 MDE 骨干网络内部进行提示或特征提取 [6], [7]。这种特征级耦合会增加运行时内存占用和延迟，并使部署变得复杂。

我们通过一种先验引导框架来应对这些挑战，该框架在输出层级上运行：我们不使用基础模型的特征，而仅采用具备泛化能力的单目深度估计器（MDE）（例如 ZoeDepth [8] 或 Depth Anything [9]）的稠密深度输出，并将其与稀疏距离锚点（例如激光雷达）对齐，从而构建校准后的伪深度先验。具体而言，我们利用可用的稀疏点校正 MDE 的预测，通过带硬约束的泊松形式获得度量准确且保留边缘的伪深度图，并将该先验（以及锚点掩码）输入轻量级细化网络，使其仅学习残差。模型旨在保持与先验一致，同时通过残差校正来处理不匹配区域（细薄结构、无纹理区域、混合深度边界）。这种输出层级的组合利用了基础单目深度估计模型的外观泛化能力，同时由稀疏锚点确定绝对尺度和局部几何结构，缩小假设空间并在不增加特征级耦合开销的情况下保持小样本稳定性。从实证结果来看，这种设计带来了紧凑且易于学习的核心，有利于部署。

[1]车辆智能感知实验室（VIPLAB），加泉大学，城南市，大韩民国。电子邮箱：jjh000503@gachon.ac.kr，jhonghyun@gachon.ac.kr

我们主要在KITTI 深度补全 [1]上进行评估——这是我们面向自动驾驶的目标部署领域——并额外在NYUv 2 [23]上验证室内泛化能力。为同时评估可比性与部署真实性，我们采用两种互补设置：一种标准少样本评估，以及一种严格、面向部署的机制，其中训练与评估共享相同的少样本预算。完整的数据集划分与协议详见§IV。该评估旨在探究在缺乏大规模、精心整理的验证集时，是否仍能获得可靠的泛化能力。

贡献。（1）我们提出一种非学习的、MDE引导的伪深度构建方法，将基础单目深度估计模型与稀疏锚点对齐，并通过带硬约束的泊松形式重建度量准确、边缘保持的先验——无需任何可训练参数。（2）在该伪先验基础上，我们提出一种轻量级先验引导深度补全网络，该网络接收伪深度图（及锚点掩码），并在少样本监督下仅学习残差细化，从而在数据极度稀缺时获得清晰边界与稳健性能——同时避免与基础单目深度估计模型进行特征级耦合，以实现实际可部署性。

II. 相关工作

A. 传统深度补全。

稀疏距离测量的图像引导深度补全，已通过 RGB–激光雷达融合以及在编码器–解码器框架中的亲和度学习与传播取得进展 [10], [11], [12]。尽管这些方法在 KITTI 深度补全等精心构建的基准上达到了较高精度 [1], [2], ，但通常假设拥有充足的逐像素监督和固定的稀疏模式，因此当传感器特性、采集路线或环境发生变化时，鲁棒性会受到限制。在实践中，依赖大规模训练集/验证集划分，以及针对特定激光雷达密度/模式进行调优，往往会损害跨传感器和跨域泛化能力。

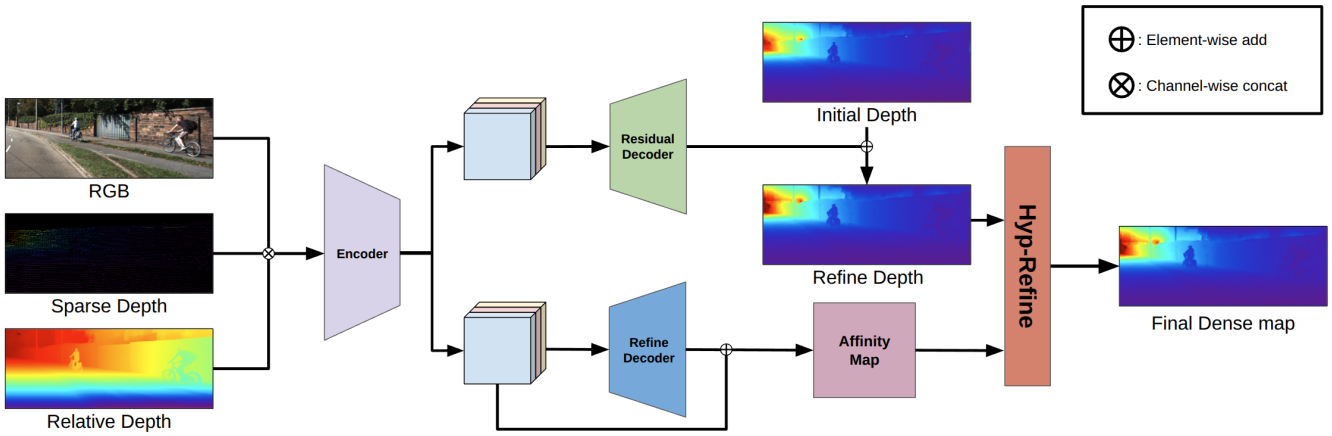

图 1. **流程概览**。 RGB、稀疏深度和相对（单目）深度经拼接后进行编码。残差解码器预测以稀疏点为锚定的初始深度，而细化分支则估计亲和力图，用于感知边缘和结构的传播。两路特征经过融合后，输入双曲细化模块（Hyp-Refine），生成最终稠密图。符号： ⊕ 逐元素相加， ⊗ 逐通道拼接。

B. 少样本深度补全与部署约束。

在已部署的机器人技术/自动驾驶（AD）场景中，运行条件（城市、路线、设备）变化的速度快于标注整理的速度；模型必须仅凭极少量标注样本，并在尽量减少验证的情况下实现泛化。排行榜式协议通常依赖完整的训练划分，以及在 KITTI 上进行大量验证集整理 [1], [2]。作为补充，我们强调面向部署的少样本评估，以检验严格数据稀缺条件下的泛化能力——例如，将训练和评估都限制在相同的少样本子集内，并报告模型在精简、固定的验证集上的表现（KITTI 提供官方1,000帧验证集划分） [1]。此类协议可直接检验：在无需大量整理验证集的情况下，能否获得稳定的度量尺度和清晰的边界。

C. 基础模型引导的补全。

近期研究利用基础单目深度估计模型，减少数据需求并提升跨传感器泛化能力。DepthPrompting[6] 引入深度提示模块，对稀疏深度进行编码，并将其与图像特征融合以构建逐像素亲和力；随后将提示嵌入预训练的单目深度估计器，以缓解传感器量程／模式偏差，并通过轻量级偏置调优，引导预测趋向绝对尺度深度。UniDC [7] 定义了通用深度补全，并提出一种简单的基线方法：（i）从基础单目深度估计模型中提取深度感知特征；（ii）基于高分辨率基础特征构建逐像素亲和力，以对齐任意稀疏测量值；（iii）将学习到的特征嵌入双曲空间，以捕捉层次化三维结构，从而提升跨传感器的零样本／少样本适应能力 [6], [7] 。这两类方法均利用了 ZoeDepth 和 Depth Anything 等基础单目深度估计模型提供的广泛外观先验 [8], [9], 并展现了将其与稀疏锚点结合的价值。尽管效果显著，这种与基础单目深度估计器骨干网络的特征层面 耦合 [6],[7] 往往会增加运行时内存占用和延迟，并使部署更加复杂。相比之下，我们在输出层级进行操作：我们仅使用冻结单目深度估计器的稠密深度输出，构建由稀疏点锚定的校准后伪深度先验。

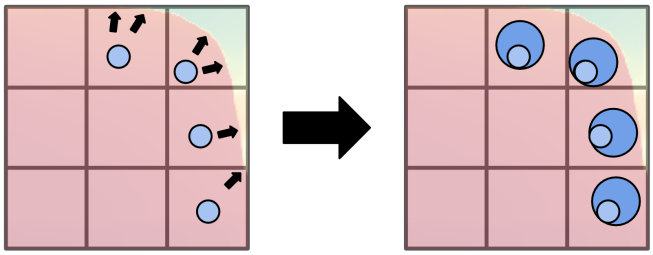

图 2. **梯度引导稠密化**。对齐后的先验形成平滑的梯度场（背景），同时稀疏度量锚点（蓝色）和图像边界保持固定。沿先验的梯度方向传播深度值以填充未知单元，同时保留结构并避免跨边缘渗漏；锚点不会移动——只补全缺失的像素。

III. 方法

A. 伪深度图构建

为在数据有限的条件下稳定训练和模型选择，我们通过融合稠密单目先验 $E \in \mathbb{R}^{H\times W}$ 与稀疏激光雷达深度 $D \in \mathbb{R}^{H\times W}$，构建伪深度图 P 。

a) 集合与边界数据： 令 Ω 表示图像域，其边界为 $\partial\Omega$。我们首先定义已知像素集合 $\mathcal{K}$ 及其补集 $\mathcal{U}$:

$$\begin{aligned}\mathcal{K} &= \{(i,j) \in \Omega \mid D_{ij} > 0\} \cup \partial\Omega, \\ \mathcal{U} &= \Omega \setminus \mathcal{K}.\end{aligned} \tag{1}$$

基于这些集合，我们定义用于硬约束的狄利克雷场 $v_{\text{known}}(i,j)$:

$$v_{\text{known}}(i,j) = \begin{cases} D_{ij}, & (i,j) \in \{D > 0\}, \\ E_{ij}, & (i,j) \in \partial\Omega. \end{cases} \tag{2}$$

b) 梯度域引导： 遵循梯度域融合 [13],，我们将未知解 x 与引导梯度 ∇E 对齐，同时精确满足测量值和边界值：

$$\min_{x} \int_{\Omega} \|\nabla x - \nabla E\|^2 \, dp \quad \text{s.t. } x|_{\mathcal{K}} = v_{\text{known}}. \tag{3}$$

欧拉–拉格朗日条件在未知集 $\mathcal{U}$上导出一个泊松方程：

$$A_{\mathcal{U}\mathcal{U}}\, x_{\mathcal{U}} = \big(LE - L v_{\text{known}}\big)_{\mathcal{U}}. \tag{4}$$

$$P = x = x_{\mathcal{U}} + v_{\text{known}}. \tag{5}$$

方程（3）–（5）定义了一个具有唯一解的严格凸问题。

$$P(p) = x(p) = \begin{cases} v_{\text{known}}(p), & p \in \mathcal{K}, \\ x_{\mathcal{U}}(p), & p \in \mathcal{U}. \end{cases} \tag{6}$$

c) 离散化与求解器。： 在像素网格上，我们使用（负）离散拉普拉斯矩阵 L。为保持离散一致性，我们将右端项设为 LE（ E的离散拉普拉斯），并将线性系统限制在未知索引集 $\mathcal{U}$上，得到式（5）；随后按式（6）进行重建。这里， $A_{\mathcal{U}\mathcal{U}}$ 是限制在 $\mathcal{U}$上的 L 。该系统为对称正定（SPD），因此可通过共轭梯度（CG） [14]高效求解。我们通过对单个 3×3 卷积进行掩码处理，以无矩阵方式实现 $A_{\mathcal{U}\mathcal{U}}$ ，该卷积用于计算 L；每次共轭梯度迭代需要 O（HW）计算量，且仅需图像大小的辅助内存。

B. 单目先验的残差校正

a) 动机。： 在训练数据有限且验证数据稀缺的情况下，纯数据驱动预测器往往表现出高方差和不稳定收敛，尤其是在细薄结构和深度不连续处。因此，我们采用先验保持型残差更新：可靠但并不完美的单目先验提供全局度量框架，而低容量解码器仅校正局部误差。这样既缩小了有效假设空间、提高了样本效率，又不牺牲场景尺度一致性。

b) 输入表示与编码器： 我们通过对齐空间并拼接以下通道来构建输入张量 X ：

$$X = \big[I,\ P,\ E,\ M_L\big]. \tag{7}$$

其中， I 是RGB图像； P 是通过泊松融合单目估计与激光雷达锚点得到的伪深度先验； E 是稠密单目先验； M_L 是二值激光雷达观测掩码。拼接之前，我们使用数据集最大深度将伪深度和单目先验都归一化到单位范围：

$$P \leftarrow \text{clip}\Big(\tfrac{P}{d_{\max}}, 0, 1\Big), \quad E \leftarrow \text{clip}\Big(\tfrac{E}{d_{\max}}, 0, 1\Big).$$

略微滥用记号后，我们仍将归一化后的图表示为 P 和 E。所有损失均在该归一化空间中计算。在推理时，预测值ˆ通过 $D = d_{\max}\, D$ 重新缩放回米，并在需要时限制为$[0, d_{\max}]$。视觉编码器 Φ 生成多模态特征$F = \Phi(X)$。我们有意限制编码器容量，以减轻过拟合，并在少样本情况下稳定优化。

c) 残差解码器与保先验更新： 浅层残差解码器预测逐像素校正 $R = f_{\text{dec}}(F)$。通过将残差加到先验上，得到校正后的初始值：

$$D^0(p) = P(p) + R(p). \tag{8}$$

（在实现中，深度会被截断至 $[0, d_{\max}]$ 。）按照设计，P 保留全局布局和绝对尺度，而 R则仅用于校正边缘、细薄结构和纹理贫乏区域附近的局部偏差。稠密单目先验 E 通过视觉编码器和跳跃连接注入，使解码器能够优先在不连续处分配残差容量，从而避免过度平滑。

C. 基于亲和度的细化

a) 范围与输入： 给定第 III-B 节中的初始化 $D^0 \in \mathbb{R}^{H \times W}$ ，我们的目标是利用几何感知的像素亲和度以及逐像素传感器锚定规则对其进行细化；该规则既遵循激光雷达观测，又能对误投影保持稳健。令 F 表示图像分辨率下的编码器特征、原始传感器（激光雷达）深度 D_s 以及对应的观测掩码 $M_L \in \{0,1\}^{H \times W}$。我们采用一组核尺寸 $\mathcal{K} = \{3, 5, 7\}$，并执行 T次传播迭代。单个曲率参数 $\kappa > 0$（在 $k \in \mathcal{K}$之间共享）定义了用于双曲距离评估的庞加莱球。

b) 符号： 对于像素 p， $\mathcal{N}_k(p)$ 表示其 $k \times k$ 邻域； $q \in \mathcal{N}_k(p)$ 对其邻居进行索引。核门控$\sigma_k(p) \in [0,1]$ 满足 $\sum_{k \in \mathcal{K}} \sigma_k(p) = 1$。所有行随机亲和权重 $A_k(p,\ q)$ 均满足 $\sum_{q \in \mathcal{N}_k(p)} A_k(p,\ q) = 1$。

c) 双曲成对亲和度：我们通过原点处的指数映射，将每个像素的特征嵌入到庞加莱球 [15], [16] 中：

$$h_p = \exp_0^{\kappa}(W_f F(p)). \tag{9}$$

其中 W_f 是一个 1×1 投影。令 $d_\kappa(\cdot,\ \cdot)$ 表示双曲距离。对于每个核 $k \in \mathcal{K}$ 和邻居 $q \in \mathcal{N}_k(p)$，我们定义未归一化权重

$$\begin{aligned} \tilde{A}_k(p,q) &= \exp\!\left(-\frac{d_\kappa(h_p, h_q)}{\tau_k}\right) \mathbf{1}\{q \in \mathcal{N}_k(p)\}, \\ A_k(p,q) &= \frac{\tilde{A}_k(p,q)}{\sum_{q' \in \mathcal{N}_k(p)} \tilde{A}_k(p,q')}. \end{aligned} \tag{10}$$

温度参数为 $\tau_k > 0$ （每个核一个）。像素级核门控由一个轻量级头部生成：

$$\sigma_k(p) = \frac{\exp\big(g_k(F(p))\big)}{\sum_{k' \in \mathcal{K}} \exp\big(g_{k'}(F(p))\big)}, \quad k \in \mathcal{K}. \tag{11}$$

直观而言， 3×3 强调边缘和细薄结构， 5×5聚合中程上下文，而 7×7 则稳定长程一致性，同时门控机制会针对每个像素自适应调整这些作用。

表 I KITTI 深度补全基准。最佳结果以**粗体**显示，次佳结果加下划线。协议：1/10/100样本仅从**训练**集划分中抽取；在**官方1,000 帧验证集**划分上进行评估（除非另有说明，否则不提交测试服务器）。

方法	1 次样本		10 次样本		100 次样本		单序列训练	
	均方根误差 (m)	平均绝对误差 (m)	均方根误差 (m)	平均绝对误差 (m)	均方根误差 (m)	平均绝对误差 (m)	均方根误差 (m)	平均绝对误差 (m)
CSPN [11]	9.2748	3.5921	2.0222	0.7825	1.4510	0.5184	2.6289	0.8355
S2D [17]	8.8479	5.6022	5.0500	3.1469	4.2799	2.6633	4.7950	2.5610
NLSPN [10]	7.2899	4.7422	4.0070	2.2588	2.4979	1.1710	4.0290	1.7881
DySPN [12]	2.6350	0.8870	2.2701	0.9150	1.8777	0.6188	2.8530	0.7980
CompletionFormer [18]	4.7212	2.3789	3.1601	1.4740	2.6122	1.3299	4.5588	1.9603
BPNet [19]	5.4000	1.0740	1.8799	0.5559	1.3001	0.3910	2.1322	0.6420
DepthPrompting [6]	2.9840	1.1430	2.3988	1.1290	1.8249	0.6240	2.9468	0.9869
Ours	**1.4190**	**0.5073**	**1.2830**	**0.4001**	**1.2455**	**0.3548**	**1.5782**	**0.5540**

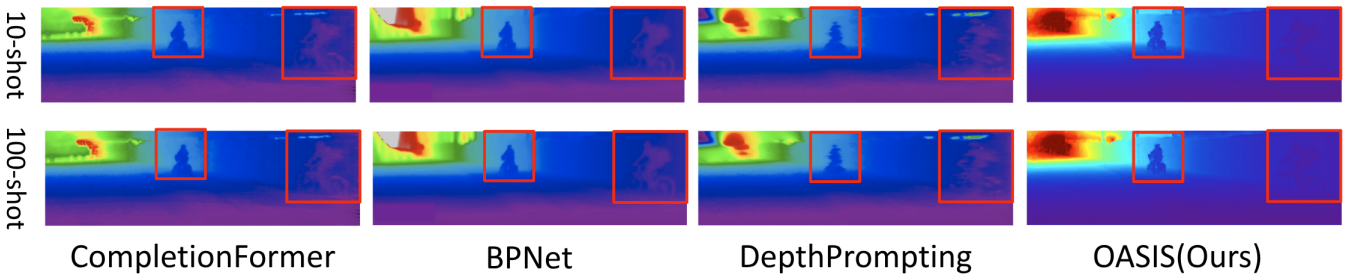

图 3. **少样本监督下 KITTI-DC 的定性比较。** 上：10 次样本；下：100 次样本。 **各列依次为：** RGB、CompletionFormer、BPNet、DepthPrompting 和 OASIS-DC（本文方法）。少样本设置仅从训练集划分中抽取样本；可视化示例根据官方的 1,000 帧验证集进行评估。红框标出了具有挑战性的区域（细薄结构、远场）。我们的结果保留了道路-墙壁边界和精细细节，同时减少了跨边缘渗漏。

d) 中心锚定多核传播。： 为防止漂移和过度平滑，每次局部更新都会以 D^0为中心重新定位。在第 t 次迭代中，我们通过将中心值替换为 D^0 来构建一个中心锚定补丁：

$$\tilde{D}^{(t)}(q) = \begin{cases} D^0(p), & q = p, \\ D^{(t)}(q), & q \neq p. \end{cases} \tag{12}$$

随后进行逐核传播和门控混合

$$D_k^{(t+1)}(p) = \sum_{q \in \mathcal{N}_k(p)} A_k(p,q)\, \tilde{D}^{(t)}(q), \tag{13}$$

$$D_{\text{mix}}^{(t+1)}(p) = \sum_{k \in \mathcal{K}} \sigma_k(p)\, D_k^{(t+1)}(p). \tag{14}$$

e) 可学习传感器锚定： 我们仅在有传感器观测的位置强制逐像素一致性，同时允许在遮挡和误投影附近进行软校正。根据特征预测锚定图 $\alpha(p) \in [0,1]$ ：

$$\alpha(p) = \sigma\big(W_\alpha F(p)\big). \tag{15}$$

其中 $\sigma(\cdot)$为逻辑函数。在观测像素处（$M_L(p)$ =1），将式（14）中的混合深度与传感器原始深度 D_s融合：

$$D^{(t+1)}(p) = \big(1 - \alpha(p)\, M_L(p)\big)\, D_{\text{mix}}^{(t+1)}(p) + \alpha(p)\, M_L(p)\, D_s(p). \tag{16}$$

并在 $M_L(p) = 0$时， $D^{(t+1)}(p) = D^{(t+1)}_{\text{融合}}(p)$。经过 T 次迭代后，我们输出 $D_{\text{final}} = D^{(T)}$ ，并在必要时将深度截断至$[0, d_{\max}]$ 。

IV. 实验

A. 数据集与评估。

我们在两个互补的基准上进行评估，涵盖户外驾驶和室内场景，并按照其官方协议报告结果。

a) KITTI 深度补全（*KITTI-DC*）。： 这是一个大规模户外驾驶数据集，包含同步的 RGB 图像以及来自 Velodyne HDL–64E 的激光雷达测量值。按照官方划分，我们使用约 86K 个训练样本、7K 个验证样本和 1K 个测试样本。图像以 1216×352 的分辨率提供。原始稀疏激光雷达深度覆盖约 6% 的图像像素，而该基准的参考深度通过多帧累积和清理生成，密度约为 20%。评估。除非另有说明，我们在官方验证集划分上报告定量结果；仅在明确说明时，才使用提交至测试服务器的结果。少样本子集（1-/10-/100 次样本和 1 序列）取自训练集划分，验证集/测试标签从不用于训练。

b) NYUv2.: 这是一个经典的室内 RGB–D 数据集，使用 Microsoft Kinect 传感器在办公室、厨房和起居空间等多种场景中采集。我们采用分辨率为 640×480 的标准标注子集，并遵循

表 II NYUv2 深度补全基准。最佳结果以**粗体**显示，次佳结果带下划线。

方法	1 次样本		10 次样本		100 次样本		单序列训练	
	均方根误差（m）	平均绝对误差（m）	均方根误差（m）	平均绝对误差（m）	均方根误差（m）	平均绝对误差（m）	均方根误差 (m)	平均绝对误差 (m)
CSPN [11]	1.4827	1.2058	0.3166	0.1961	0.2854	0.1307	0.3166	0.1961
NLSPN [10]	1.9358	1.6132	1.5995	0.8261	0.5501	0.4150	0.8881	0.6421
DySPN [12]	1.5474	1.2851	0.4102	0.2817	0.3079	0.1706	0.2584	0.1320
CompletionFormer [18]	1.8218	1.5539	1.1583	1.0162	0.9914	0.8164	0.6779	0.5356
CostDCNet [20]	1.2298	0.9754	0.2363	0.1288	0.1770	0.0836	0.2066	0.0954
BPNet [19]	0.3573	0.2077	0.2392	0.1120	0.1757	0.0793	0.2220	0.1040
DepthPrompting [6]	0.3583	0.2067	0.2195	0.1006	0.2101	0.1008	0.2335	0.1191
UniDC [7]	**0.2099**	**0.1075**	**0.1657**	**0.0794**	**0.1473**	**0.0669**	**0.1632**	**0.0745**
Ours	0.2105	0.1105	0.1670	0.0838	0.1484	0.0706	0.1644	0.0787

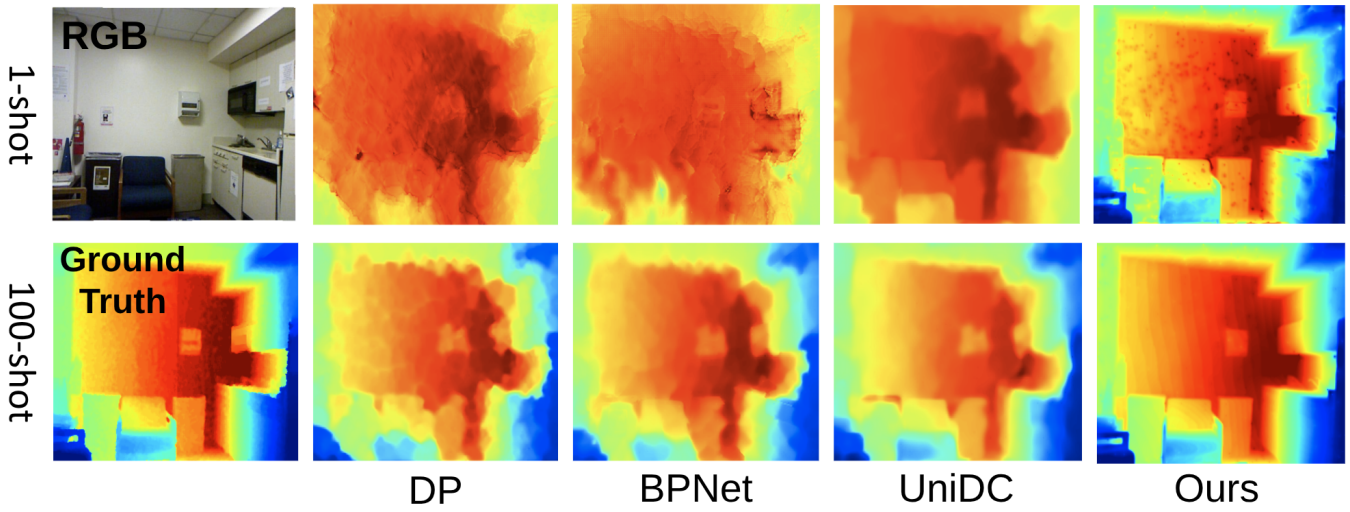

图 4。**NYUv2 定性比较**。第一行：1 次样本（左：RGB）；第二行：100 次样本（左：真值）。其余各列分别展示 DP（DepthPrompting）、BPNet、UniDC 和我们的方法。所提出的方法能够生成更洁净的平面表面和更清晰的不连续处（例如柜体边缘），同时在不同样本数量下抑制噪声和纹理复制。

传统的官方划分进行训练和评估。对于深度补全，我们通过对稠密深度图进行下采样，合成具有类似激光雷达稀疏度的稀疏输入；所有方法使用相同的稀疏掩码，以确保公平。评估。与 KITTI-DC 一样，除非另有说明，我们均在官方验证集划分上报告结果，且训练期间不使用验证/测试标注；少样本设置与 KITTI 的设置一致。

c) 指标： 我们报告均方根误差（米）和平均绝对误差（米）。两者均为越低越好。所有指标均根据各数据集的官方协议，在有效真实值像素上计算，且无效/未知像素不计入平均值。

B. 实现细节。

所有实验均在单张 NVIDIA RTX A5000（24GB）图形处理器上执行。我们在标准深度补全基准（例如 KITTI-DC [1], [2] 和 NYUv2 [23]）的官方训练划分上采用严格的少样本协议：从每个数据集中构建*1/10/100*样本 子集，以及由一段连续序列组成的*1*序列子集。对于每种样本数量和序列设置，采样均使用 r 个相互独立的随机种子重复进行；除非另有说明，我们报告这些 r 次试验的均值（并在篇幅允许时报告标准差）。训练仅使用少样本子集，验证/测试标注绝不用于训练。作为基础单目深度估计器，我们使用*Depth Anything v 2* [21]。

优化严格在少样本模式下进行，迭代次数按样本子集大小进行调整：1 次样本使用 ~ 100 次迭代，较大的少样本子集（例如 100 次样本）以及单序列情况则使用 3,000 次迭代。除非另有说明，所有其他超参数在不同样本数之间保持不变，以确保可比性。**损失函数**。我们使用两个掩码目标函数进行监督。令 D_{gt} 表示真实深度，并且

$$M = \mathbb{1}[D_{gt} > 0], \qquad n = \max\left(1, \sum M\right). \tag{17}$$

复合 L1+L2 损失为

$$e = (D_{\text{pred}} - D_{gt}) \odot M, \qquad \mathcal{L}_{\text{L1+L2}} = \frac{1}{n} \sum \left(|e| + e^2\right). \tag{18}$$

尺度不变对数损失作用于正预测值D^{+} 相对 （我们在 $(0,1]$ 使用 ε 截断）中使用归一化输入：

$$d = \big(\log(D^{+}_{\text{rel}} + \varepsilon) - \log(D_{gt} + \varepsilon)\big) \odot M,$$
$$\mathcal{L}_{\text{SI-Log}} = \frac{1}{n}\sum d^2 \; - \; \Big(\tfrac{1}{n}\sum d\Big)^2. \tag{19}$$

如果样本中没有 D_{gt}，则将相应项设为零。除非另有说明，否则我们将两项损失相加。

C. 定量结果。

a) KITTI.: 表 I 中的改进源于我们模型将问题分解为度量对齐和局部残差校正的方式。稀疏激光雷达锚点固定了全局低频结构（绝对尺度和大面积平滑表面），而在 1/10/100样本区间内，仅凭少量标注样本无法可靠地估计这些结构。在此校准后的伪深度基础上，残差分支将有限的容量集中用于处理不连续处周围的高频误差。这一设计直接针对穿透失效模式（深度跨越物体边界泄漏）：残差经过学习，仅在梯度与测量值不一致的位置偏离先验，从而使边界更清晰并抑制跨边缘渗漏。消融实验证实了这一机制：仅使用残差无法消除度量漂移；仅进行尺度校准虽能减少偏差，却无法修复边缘局部伪影；两者结合则能在不同样本数下持续带来提升。从 1 样本到 1 个序列，这些提升始终存在，说明主要驱动因素是模型的归纳偏置，而非更多监督：先验约束假设空间，残差则充当有针对性且数据高效的校正器，而不是完整的预测器。

b) NYUv2：如表 II 所示，我们的方法始终位居第二，仅以微弱差距落后于 UniDC，同时优于其余基线方法。这一差距可由两个因素解释：(i) NYUv2 上较弱的锚定——Kinect 真值与合成稀疏掩码在不连续处提供的约束不如 KITTI 上的真实激光雷达可靠——从而限制了度量对齐对先验的正则化强度；以及 (ii) 为少样本稳定性而采用的有意保守的残差容量，相对于 UniDC 更大的预测头，可能对室内精细细节拟合不足。即便如此，图 4 中的定性比较仍显示出比其他网络更锐利的边缘和更干净的平面，这与表 II 中的汇总指标一致。

D. 消融实验。

a) 不完美的真值：在 NYUv2 上，我们经过尺度校准的先验结合残差边缘细化，在各次拍摄中取得了第二好的均方根误差／平均绝对误差（表 II），略逊于 UniDC，但优于其他基线方法。图 5 解释了这一差距：Kinect 真值常会遗漏细薄结构，并过度平滑不连续处；我们的残差头能够锐化边界并恢复小物体，因此尽管几何结构明显更清晰，边界带中的像素误差仍可能增加。在消融实验（仅先验／仅残差／完整模型）中，只有完整模型能够同时稳定全局尺度

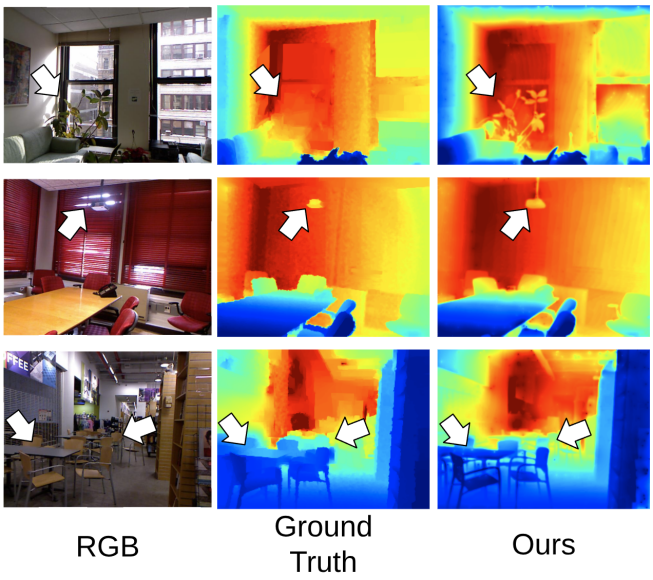

图 5. **不完美的 NYUv2 真值。** 箭头标出真值伪影（空洞／平滑）；指标使用有效真值掩码。

表 III KITTI 深度补全基准（消融实验）。残差／尺度标志表示模块使用情况：(✓,–) 仅残差头；(–,✓) 仅伪先验（尺度）；(✓,✓) 两者均使用；(–,–) 两者均不使用。

KITTI 深度补全基准						
		1 次样本		10 次样本		100 次样本
残差尺度	均方根误差	平均绝对误差	均方根误差	平均绝对误差	均方根误差	平均绝对误差
		2.8584	2.1554	2.2840	1.5562	2.2371	1.5109
✓		4.4062	2.7533	3.5208	1.9879	3.4485	1.9300
	✓	2.8939	2.1404	2.3140	1.5469	2.2666	1.5009
✓	✓	**1.4190**	**0.5073**	**1.2830**	**0.4001**	**1.2455**	**0.3548**

并抑制跨边缘渗透，从而在定量得分与边缘保真度之间实现最可靠的权衡。

b) 网络设计的消融实验： 表 III 和表 V 共同表明，我们的残差头工作在极低容量区间（0.219M 个可学习参数，推理耗时 0.013 秒；表 V）。在这种设置下，锚点 （通过稀疏深度进行尺度校准）至关重要：没有锚点时，仅残差变体会欠拟合，并放大单目先验偏差；在 1/10/100样本各区间内，其误差甚至远高于两者均不使用的配置。相比之下，停用两个模块时，仅伪深度先验就会施加强烈的归纳偏置，为性能设定“下限”（表 III 第一行），即使可训练头极小，也能避免灾难性漂移。最佳结果出现在锚点与残差结合时（最后一行）：锚点稳定全局度量尺度和低频结构，而残差则将有限容量用于高频边缘校正（穿透抑制、边界锐化）。这一规律在不同样本数区间内始终一致，进一步说明

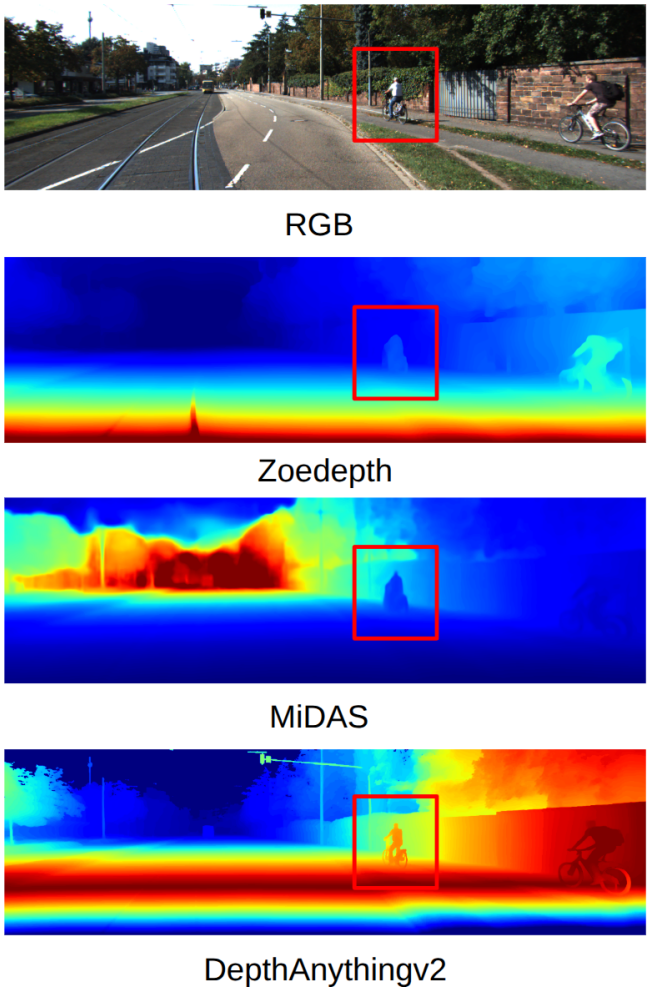

图 6. **Depth Anything v2 在保留边界方面表现出色**。KITTI 场景的定性比较。在红框内，与 ZoeDepth 和 MiDaS 相比，Depth Anything v2 能更可靠地保留精细的物体边界和细薄结构，从而呈现更清晰的轮廓，并减少色彩渗漏伪影。

对于紧凑的可学习核心（表 V），锚点信号不仅有帮助，而且是可靠进行深度补全所必需的；残差随后提供有针对性的细化，而不必承担从少量样本中学习尺度的重任。

c) 伪深度先验消融：表 IV 通过仅替换冻结的基础单目深度估计先验（Depth Anything v2 [21], MiDaS [22], ZoeDepth [8]），并采用相同的测试时稠密化流程，评估我们无需训练的伪深度生成流程，且不进行学习式细化。尽管无需任何训练，该方法在两个基准上都取得了有竞争力的绝对误差，表明仅靠伪深度流程就足以满足实际部署需求，尤其适用于标注数据或训练预算受限的情况。

在 KITTI 上，[1], [2], Zoe 先验取得了最佳结果（均方根误差／平均绝对误差为 = 1.6264/0.4135），而 Depth Anything v2 先验的结果为 1.7481/0.4525——相较于 Zoe，均方根误差高出约 ∼7.5%，平均绝对误差高出约 ∼9.4%。我们提醒

表 IV KITTI 和 NYUv2 上的伪深度先验消融。伪深度图通过对冻结的基础单目深度估计模型进行测试时稠密化获得，且不进行任何学习式细化。

	KITTI		NYUv2	
方法	RMSE	MAE	RMSE	MAE
我们的_Depth Anything v2	1.7481	0.4525	**0.2115**	**0.1135**
我们的_MiDaS	1.7117	0.4360	0.2741	0.1417
我们的_Zoe	**1.6264**	**0.4135**	0.3032	0.1223

表 V
模型计算成本。

模型	总计参数量	可学习参数	推理时间（秒）	GPU 内存 (MiB)
BPNet [19]	89.874M	89.874M	0.072	4792
LRRU [28]	**20.843M**	20.843M	0.038	3650
CompletionFormer [18]	83.574M	83.574M	0.060	4206
Ours	94.550M	**0.219M**	**0.013**	**1904**

读者注意，KITTI 提供的是稀疏激光雷达监督，基于采样点的指标往往更偏向平滑的预测，或与激光雷达采样点重合的预测；在我们的定性检查中，Depth Anything v2 对物体边界和细薄结构的保留更加清晰，但这并未在稀疏指标中得到充分体现。与此一致，在稠密 室内 NYUv2 基准上，[23],Depth Anything v2 的误差最低（0.2115/0.1135），均方根误差／平均绝对误差分别比 MiDaS 低 ≈22.8%/19.9%，比 ZoeDepth 低 ≈30.2%/7.2%。这些结果表明，我们无需训练生成的伪深度图已经足够稳健，可用于工业场景；而稠密真实值评估能更好地反映 Depth Anything v2 先验所展现的边缘保真度。

五、结论

我们提出了一种先验引导的少样本深度补全框架，将基础单目深度估计模型与稀疏锚点融合，形成校准后的伪深度先验，并训练细化网络，使其在遵循该先验的同时校正局部不匹配。此互补设计缩小了有效假设空间，使模型能够从极少样本中稳定学习，并保留度量尺度和边缘细节。我们在 KITTI 上的评估遵循两种互补协议：用于比较的标准少样本设置，以及面向实际部署的严格方案。后者在 1/10/100 次样本和 1 个序列子集上使用五个随机种子（除 1 次样本外，均采用 70/30 的训练/验证划分），最终在官方的 1,000 张图像验证集划分上报告结果；该验证集从未用于训练或超参数调优 [1], [2]。通过消融实验（无先验/无稀疏锚点/完整先验）分离伪先验与稀疏锚点各自作用的结果表明，移除先验会显著损害泛化能力，而移除锚点则会加剧尺度漂移。未来工作包括加强对单目先验的感知不确定性校准、处理动态场景

，以及将方法扩展到不同激光雷达稀疏度和跨数据集泛化。

参考文献

[1] J. Uhrig、N. Schneider、L. Schneider、U. Franke、T. Brox 和 A. Geiger，“稀疏度不变卷积神经网络”，载于 3DV 会议论文集，2017年。[2] A. Geiger、P. Lenz 和 R. Urtasun，“我们准备好自动驾驶了吗？KITTI视觉基准套件”，载于 CVPR 会议论文集，2012年。[3] J. Tang、F.-P. Tian、W. Feng、J. Li 和 P. Tan，“用于深度补全的学习引导卷积网络”，IEEE图像处理汇刊，第30卷，第1116–1129页，2021年。[4] J. Qiu、Z. Cui、Y. Zhang、X. Zhang、S. Liu、B. Zeng 和 M. Pollefeys，“DeepLiDAR：利用稀疏激光雷达数据和单幅彩色图像，通过深度表面法线引导预测室外场景深度”，载于 CVPR 会议论文集，2019年。[5] M. Hu、S. Wang、B. Li、S. Ning、L. Fan 和 X. Gong，“PENet：迈向精确高效的图像引导深度补全”，载于 ICRA 会议论文集，2021年。[6] J.-H. Park、C. Jeong、J. Lee 和 H.-G. Jeon，“用于传感器无关深度估计的深度提示”，载于 CVPR 会议论文集，2024年。[7] J.-H. Park 和 H.-G. Jeon，“一种简单而通用的深度补全框架”，载于 NeurIPS 会议论文集，2024年。[8] S. F. Bhat、R. Birkl、D. Wofk、P. Wonka 和 M. M¨uller，“ZoeDepth：结合相对深度与度量深度实现零样本迁移”，arXiv:2302.12288，2023年。[9] L. Yang、B. Kang、Z. Huang、X. Xu、J. Feng 和 H. Zhao，“Depth Anything：释放大规模无标注数据的力量”，载于 CVPR 会议论文集，2024年。[10] J. Park、K. Joo、Z. Hu、C.-K. Liu 和 I. S. Kweon，“用于深度补全的非局部空间传播网络”，载于 ECCV 会议论文集，2020年。[11] X. Cheng、P. Wang 和 R. Yang，“通过卷积空间传播网络学习亲和度进行深度估计”，载于 ECCV 会议论文集，2018年。[12] Y. Lin、T. Cheng、Q. Zhong、W. Zhou 和 H. Yang，“用于深度补全的动态空间传播网络”，载于 AAAI 会议论文集，2022年。[13] P. P´erez、M. Gangnet 和 A. Blake，“泊松图像编辑”，载于 SIGGRAPH 会议论文集，2003年。[14] M. R. Hestenes 和 E. Stiefel，“求解线性系统的共轭梯度方法”，《美国国家标准局研究杂志》，第49卷，第6期，第409–436页，1952年。[15] O.-E. Ganea、G. B´ecigneul 和 T. Hofmann，“双曲神经网络”，载于 NeurIPS 会议论文集，2018年。[16] M. Nickel 和 D. Kiela，“用于学习层次表示的 Poincar´e 嵌入”，载于 NIPS 会议论文集，2017年。[17] F. Ma 和 S. Karaman，“稀疏到稠密：根据稀疏深度样本和单幅图像预测深度”，载于 ICRA 会议论文集，2018年。[18] Y. Zhang、X. Guo、M. Poggi、Z. Zhu、G. Huang 和 S. Mattoccia，“CompletionFormer：结合卷积与视觉变换器的深度补全”，载于 CVPR 会议论文集，2023年。[19] J. Tang、F.-P. Tian、B. An、J. Li 和 P. Tan，“用于深度补全的双边传播网络”，载于 CVPR 会议论文集，2024年。[20] J. Kam、J. Kim、S. Kim、J. Park 和 S. Lee，“CostDCNet：基于代价体的单幅 RGB-D 图像深度补全”，载于 ECCV 会议论文集，2022年。[21] L. Yang、B. Kang、Z. Huang、Z. Zhao、X. Xu、J. Feng 和 H. Zhao，“Depth Anything v2”，《神经信息处理系统进展》（NeurIPS），第37卷，第21875–21911页，2024年。[22] R. Ranftl、K. Lasinger、D. Hafner、K. Schindler 和 V. Koltun，“迈向稳健的单目深度估计：混合数据集以实现零样本跨数据集迁移”，IEEE模式分析与机器智能汇刊，第44卷，第3期，2022年。[23] N. Silberman、D. Hoiem、P. Kohli 和 R. Fergus，“基于 RGB-D 图像的室内场景分割与支撑推断”，载于 ECCV 会议论文集，2012年。[24] A. Wong 和 S. Soatto，“通过校准反投影层实现无监督深度补全”，载于 ICCV 会议论文集，2021年。[25] B. Ke、A. Obukhov、S. Huang、N. Metzger、R. C. Daudt 和 K. Schindler，“将基于扩散的图像生成器用于单目深度估计”，载于 CVPR 会议论文集，2024年。[26] M. Gui 等，“DepthFM：基于流匹配的快速单目深度估计”，arXiv:2403.13788，2024年。[27] H. Lee、K. S. Kim、B.-K. Kwon 和 T.-H. Oh，“通过测试时对齐与仿射不变深度先验实现零样本深度补全”，arXiv:2502.06338，2025年。[28] Y. Wang、B. Li、G. Zhang、Q. Liu、T. Gao 和 Y. Dai，“LRRU：用于深度补全的长短程循环更新网络”，载于 *ICCV* 会议论文集，2023年。[29] A. Bansal、H.-M. Chu、A. Schwarzschild、S. Sengupta、M. Goldblum、J. Geiping 和 T. Goldstein，“扩散模型的通用引导”，载于 *ICLR* 会议论文集，2024年。[30] J. Song、C. Meng 和 S. Ermon，“去噪扩散隐式模型”，载于 *ICLR* 会议论文集，2021年。[31] H. Chung、J. Kim、M. T. McCann、M. L. Klasky 和 J. C. Ye，“一般有噪逆问题的扩散后验采样”，载于*ICLR* 会议论文集，2023年。[32] B. Song、S. M. Kwon、Z. Zhang、X. Hu、Q. Qu 和 L. Shen，“通过硬数据一致性利用潜在扩散模型求解逆问题”，载于 *ICLR* 会议论文集，2024年。[33] L. Bartolomei、M. Poggi、A. Conti、F. Tosi 和 S. Mattoccia，“从立体匹配视角重新审视深度补全，以实现跨域泛化”，载于 *3DV* 会议论文集，2024年。[34] Y. Saad，稀疏线性系统的迭代方法，第2版，SIAM，2003年。[35] D. Eigen、C. Puhrsch 和 R. Fergus，“使用多尺度深度网络进行单幅图像深度图预测”，载于*NeurIPS* 会议论文集，2014年。[36] R. C. Gonzalez 和 R. E. Woods，数字图像处理，第4版，Pearson，2018年。[37] P. Soille，形态学图像分析：原理与应用，第2版，Springer，2003年。