DOI: 10.3969/j.issn.2097-583X.2026.04.024　　文章编号：2097-583X(2026)04-0688-10

非标准场景下的单目深度估计研究综述
——从环境退化到空间畸变

陈松楠[1a,2,4,5]，潘赫斐[2]，张　欣[2]，陈雪林[3]，毛海涛[1b]，董蕊芳[4]，赵汐璇[4]，阚江明[4*]

(1. 信阳师范大学 a. 计算机与信息技术学院；b. 后勤管理服务中心，河南 信阳 464000；
2. 武汉轻工大学 数学与计算机学院，湖北 武汉 430048；3. 信阳农林学院 管理学院，河南 信阳 464000；
4. 北京林业大学 林业装备与自动化国家林业和草原局重点实验室，北京 100083；
5. 佛山市中科农业机器人与智慧农业创新研究院，广东 佛山 528200)

摘　要：单目深度估计是自动驾驶、机器人环境感知与虚拟现实等领域的关键视觉感知技术，为二维图像到三维物理空间的几何映射提供支撑。然而，传统方法多依赖理想成像条件与标准透视投影假设，在非标准场景下性能显著退化。首先，系统梳理非标准场景下单目深度估计的前沿研究进展，将核心技术挑战凝练为环境退化与空间畸变两大维度。针对雨、雾、夜间等复杂环境引发的图像退化问题，从生成式增强与基础模型先验、自监督跨域特征适应、物理显式约束与多模态融合3个层面，归纳了提升模型鲁棒性的优化策略。针对全景视场等距柱状投影(equirectangular projection，ERP)造成的球面空间畸变问题，沿监督学习、自监督学习、半监督与知识蒸馏三大技术路径，剖析了适配球面几何特性的深度估计新范式。其次，指出基于基础模型先验的跨域迁移是解决上述两类场景中数据稀缺与分布偏移难题的共性技术趋势；最后，从基础大模型驱动的统一先验迁移、环境与空间复合感知、多模态异构融合以及评测基准标准化等4个维度，对构建通用单目深度估计方法进行了展望。

关键词：非标准场景；单目深度估计；环境退化；空间畸变；跨域迁移

中图分类号：TP399　　**文献标志码：**A　　**开放科学(资源服务)标识码(OSID)：**

A survey of monocular depth estimation in non-standard scenarios: From environmental degradation to spatial distortion

Chen Songnan[1a,2,4,5], Pan Hefei[2], Zhang Xin[2], Chen Xuelin[3], Mao Haitao[1b], Dong Ruifang[4], Zhao Xixuan[4], Kan Jiangming[4*]

(1. a. School of Computer Science and Information Technology; b. Logistics Management Service Center, Xinyang Normal University, Xinyang 464000, China;
2. School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan 430048, China;
3. School of Management, Xinyang Agriculture and Forestry University, Xinyang 464000, China;
4. Key Lab of National Forestry and Grassland Administration on Forestry Equipment and Automation, Beijing Forestry University, Beijing 100083, China;
5. Foshan-Zhongke Innovation Research Institute of Intelligent Agriculture and Robotics, Foshan 528200, China)

Abstract: Monocular depth estimation was regarded as a key visual perception technology in fields such as

收稿日期：2026-04-22；**修回日期：**2026-05-21；*. **通信联系人，**E-mail：kanjm@bjfu.edu.cn

基金项目：国家自然科学基金项目(32071680)；广东省基础与应用基础研究基金联合基金项目(2023A1515110776)；湖北省教育厅科学技术研究项目(D20241702)；河南省高等学校重点科研项目(26B520044)

作者简介：陈松楠(1988—)，男，河南信阳人，讲师，博士，硕士生导师，主要从事计算机视觉的研究；阚江明(1976—)，男，湖北咸宁人，教授，博士，博士生导师，主要从事林业装备与信息化的研究。

引用格式：陈松楠，潘赫斐，张欣，等. 非标准场景下的单目深度估计研究综述——从环境退化到空间畸变[J]. 信阳师范大学学报(自然科学版)，2026，39(4)：688-697.

Chen Songnan, Pan Hefei, Zhang Xin, et al. A survey of monocular depth estimation in non-standard scenarios: From environmental degradation to spatial distortion[J]. Journal of Xinyang Normal University (Natural Science Edition), 2026, 39(4): 688-697.

autonomous driving, robotic environment perception and virtual reality, by which the geometric mapping from two-dimensional images to three-dimensional physical space was supported. However, traditional methods mostly relied on ideal imaging conditions and standard perspective projection assumptions, and their performance was significantly degraded in non-standard scenarios. First, the frontiers of research on monocular depth estimation in non-standard scenarios were systematically reviewed, and the core technical challenges were distilled into two major aspects: environmental degradation and spatial distortion. To address the image degradation caused by complex environments such as rain, fog, and nighttime, optimization strategies for improving model robustness were summarized from three aspects: generative augmentation and foundation model priors, self-supervised cross-domain feature adaptation, and explicit physical constraints with multimodal fusion. To address the spherical spatial distortion caused by the equirectangular projection (ERP) in panoramic fields of view, novel depth estimation paradigms that adapt to spherical geometric properties were analyzed along three technical paths: supervised learning, self-supervised learning, and semi-supervised learning with knowledge distillation. Second, it was pointed out that cross-domain transfer based on foundation model priors was a common technical trend for solving the challenges of data scarcity and distribution shift in the above two types of scenarios. Finally, a prospect for constructing a general monocular depth estimation method was presented from four dimensions: unified prior transfer driven by foundation models, composite perception of environment and space, multimodal heterogeneous fusion, and standardization of evaluation benchmarks.

Key words: non-standard scenarios; monocular depth estimation; environmental degradation; spatial distortion; cross-domain transfer

0 引言

场景深度信息是连接二维视觉表征与三维物理空间的核心几何纽带,在自动驾驶辅助、机器人自主导航、无人机低空感知等任务中发挥着不可替代的作用。早期监督学习与自监督学习方法已在标准数据集上取得了显著的性能突破,但这些方法隐含两类理想化假设:其一,环境光度理想,即晴朗无干扰条件下光度一致性严格成立;其二,几何投影理想,即输入为受限视场的标准透视图像,满足卷积神经网络(convolutional neural network, CNN)的平移不变性假设。随着智能感知系统向真实非结构化环境部署,上述理想假设往往难以满足,甚至被打破。

单目深度估计在实际自然条件下的应用,面临非标准场景带来的严峻挑战,可归结为两个维度:1)环境因素导致的图像退化使物理与统计假设失效。雨、雾、夜间等复杂气象引发大气散射[1]、随机遮挡与光照剧变,严重降低成像质量;其高频噪声、积水反光与动态光源瓦解了自监督学习所依赖的光度一致性前提,迫使研究向数据增强、域适应及物理先验嵌入等方向突破。2)全方位感知需求引发的空间拓扑畸变。常规视场相机难以满足360°全景感知需求,全景图像普遍采用等距柱状投影(equirectangular projection, ERP),在高纬度(极区)区域产生严重的非线性拉伸与拓扑畸变,导致传统CNN的感受野失效,从而推动球面重投影、多面体融合以及隐式场等适配非欧几何特征的新架构研发。环境物理退化与相机空间畸变的叠加,构成单目感知技术迈向全天候、全方位实用化的核心壁垒。近年来,以视觉基础大模型为代表的通用视觉技术快速发展,应对上述双重挑战的技术路线呈现出融合演进的趋势。

基于上述背景,本文系统梳理了非标准场景单目深度估计的研究脉络,总结关键技术与发展规律,以期为该领域后续研究提供参考。

1 非标准场景下单目深度估计的核心挑战

1.1 环境退化带来的挑战

真实野外环境中的气象变化会导致严重的成像退化。以Depth Anything为代表的基础大模型[2]虽在常规场景下泛化优异,但在雨、雾、夜间等非标准场景下仍面临以下3方面制约。

1)复杂天气破坏成像物理基础。雾霾引起的大气散射效应[1]削弱了图像纹理与对比度;雨滴、雪花造成随机遮挡与高频噪声,破坏了物体结构的连续性;夜间低信噪比与强眩光进一步加剧了结构信息的缺失,导致像素级光度重构损失失效。

2)自监督核心假设在物理层面失效。Zhou等[3]和Godard等[4]提出的时序重投影自监督框架

高度依赖相邻帧间的光度一致性假设[3,5]，而复杂天气条件下的非刚性遮挡与光照突变会引发像素强度的非结构化剧烈变化，导致模型在路面反光、光源眩光等区域产生大面积的深度空洞与伪影。

3) 域偏移与叠加态分布差异。深度学习模型的泛化能力高度受限于训练数据的分布。当前主流算法在诸如KITTI等晴朗天气采集的标准视角数据集[6]上的性能已趋于饱和，但迁移至Oxford RobotCar等包含雨夜、广视角的非标准数据集时，泛化性能呈断崖式下降[7]。这表明，待感知的真实场景是环境退化与视角畸变的叠加态，亟须构建适配上述双重挑战的鲁棒特征表征。

1.2　空间畸变带来的挑战

传统单目深度估计方法均基于有限视场的透视成像，其网络架构高度依赖二维欧氏空间的平移不变性与均匀局部感受野。在360°全方位环境感知中，球面视觉信号投影为等距柱状格式时，在高纬度区域产生的非线性拉伸与拓扑畸变，会导致同一物体在不同纬度呈现出剧烈的尺度与形态变化，使得在透视图像上训练的特征提取器出现严重的对齐错误。因此，突破二维标准卷积的固有局限，设计适配球面非欧几何特性的深度解码范式，成为单目全景深度估计领域亟待解决的核心空间挑战。

2　环境退化下的单目深度估计

复杂天气引发的环境退化是制约单目深度感知系统在真实野外场景大规模部署的核心物理障碍之一。针对成像物理退化、光度一致性失效及域分布偏移等关键挑战，学术界逐步形成3类主流技术范式：1)生成式数据增强与基础模型先验，通过合成高质量退化样本与大模型预训练缓解真值稀缺；2)聚焦自监督域适应学习，在缺乏真实标注的条件下挖掘跨天气域的共享几何先验；3)物理显式约束与多模态融合，针对夜间低信噪比、高动态干扰等场景进行专项优化。

2.1　基于生成式数据增强与基础模型

真实环境退化场景面临深度真值稀缺与数据分布受限的双重瓶颈。该范式下演化出利用生成式模型主动合成高质量退化样本以扩充数据和挖掘大模型先验两条路线。

早期基于生成对抗网络(generative adversarial network, GAN)[8]的风格迁移易导致边缘细节区域出现畸变[9]；ACDepth则提出单步扩散策略[10-11]，在保持场景三维几何拓扑一致性的前提下，引入可控气象扰动，以低成本的计算开销生成逼真的雨雾退化图像，有效缓解仿真到真实场景的域偏移问题。

合成数据难以完全复现真实光学散射与随机遮挡的场景，促使学术界转向基础大模型驱动的路线(如Depth Anything方法[2])。其强语义先验能有效抵抗高频噪声干扰，却无法直接消除物理遮挡(如镜头雨滴)和预测绝对尺度的深度信息。为此，研究者引入具有强大多模态生成能力的扩散模型。例如，Ke等[12]重构了预训练潜扩散模型(latent diffusion model, LDM)的生成先验，并配合仿射不变深度损失进行微调，展现出卓越的零样本跨域迁移能力。ECoDepth以ViT(vision transformer)提取的全局图像特征作为扩散骨干架构的引导信息，当输入纹理被雨雾严重破坏时，利用潜空间的流形投影恢复大尺度的连贯物理结构[13]。此外，Iris方法引入光谱门控一致性机制[14]在频域中显式解耦物理域偏移，在迁移低频真实世界先验的同时强制保留高频几何细节，大幅提升了模型在极端天气下的鲁棒性。然而，该路径仍存在双重瓶颈。一是扩散模型(如Marigold、ECoDepth、Iris)的多步迭代马尔可夫链采样与频域转换操作的计算与显存开销较大，难以满足高频实时推理的需求。二是现有主流架构从底层均基于标准透视投影，其强依赖的全局透视先验与空间特征在全景图像极区易发生严重扭曲甚至失效。目前，学术界尚未提出能够同时应对“环境退化”与“全景空间畸变”叠加态难题的统一生成式感知机制。

2.2　基于自监督域适应学习

针对合成图像与真实退化环境间的分布偏移，自监督域适应学习旨在缺乏深度标签的前提下，挖掘跨天气域的共享几何先验，主要演化出几何结构迁移[15]与跨域蒸馏[16]两大分支。

在结构迁移方面，Yan等[15]提出的Synthetic-to-Real方法基于运动结构保持相对稳定的物理特性，以代价体作为特征承载单元，强制模型在隐空间中传递跨域共享的运动与结构先验，从而强化退化场景下的跨域几何一致性。在跨域蒸馏方面，md4all利用清晰白昼图像训练的教师网络，指导学生网络在环境退化条件下学习稳健的几何特

征[16],有效规避了光度一致性假设失效的瓶颈。此外,文献[17]通过多教师协同蒸馏策略统一相对深度与度量深度的尺度空间,提升了模型在复杂场景下的泛化能力。然而,此类方法的精度上限仍受限于教师模型在复杂表观下的预测鲁棒性,且跨域特征的精准对齐仍是当前的技术难点。

在雨滴遮挡、积水反光等极限退化场景中,传统自监督所依赖的朗伯体反射假设被彻底破坏,导致局部光度损失遭遇病态梯度爆炸。为此,Choi等[18]提出了反射感知的三元组挖掘机制,利用跨视角的相机几何关系精准定位像素级反射区域,并结合自适应掩码屏蔽失效信号,强制模型依赖全局拓扑先验进行深度推断。

尽管基于自监督域适应学习进展显著,但在非标准场景中仍面临两大核心局限。首先,如Synthetic-to-Real等方法[15]高度依赖场景的静态刚性假设,在密集交通流等高动态退化场景中,运动先验的失效极易导致感知失败。其次,无论是跨域蒸馏还是掩码机制,在屏蔽局部退化信号后,深度的推断难以保障微观几何细节的绝对精度。这表明,仅依靠统计学层面的域适应机制难以彻底解决极限退化下的感知失效,迫使研究范式转向底层成像机理的深入挖掘。

2.3 基于物理先验与多模态融合的显式感知

面对极限成像情况,引入显式物理先验与多模态机制成为突破感知瓶颈的关键。

在纯视觉层面,构建特定光学场景的物理先验模型能有效弥补感知短板。例如,Shen等[19]提出的DNA-Depth,利用傅里叶变换将图像解耦,通过在频域置换振幅谱实现昼夜风格平滑;Zheng等[20]提出的STEPS引入光照感知掩码机制,依据物理光照分布自适应抑制极暗、过曝区域的损失权重,强制网络聚焦于光照适中的有效特征区域,有效抑制了车灯眩光导致的深度伪影。

此外,物理先验的应用已从宏观驾驶领域延伸至微观感知任务,如D^{2}FNet提出了退化感知的扩散特征引导架构[21],通过扩散特征生成模块合成退化不变特征,并结合无风格内容重建块成功抑制在内窥镜等极具挑战的微观场景中环境噪声引发的风格畸变。

然而,面对浓雾或全黑环境下的物理感知极限,多模态异构融合成为实现全天候感知的核心方向。PromptDA方法构建了一种度量提示架构,摒弃了传统的分支特征融合范式,将稀疏雷达或LiDAR物理测距点作为几何提示,引导大模型推理[22],有效缓解了单目系统在环境退化下的尺度坍塌问题。为强化传感器间的几何关联,Sun等[23]提出的GET-UP采用注意力增强的图神经网络(graph neural networks, GNN)显式建模雷达点的3D几何结构,并利用异构特征聚合在视觉信息缺失区域重建局部几何一致性。

尽管如此,在复杂气象与全景畸变的叠加状态下,跨模态融合仍面临严峻挑战:1)严苛的物理对齐壁垒。目前的融合研究大多局限于标准透视场,在全景非欧空间中,ERP引发的非线性畸变与稀疏测距点间的空间映射极易错位,若雷达点云与全景图像间的投影逻辑未经过精确校准,物理退化导致的高频细节丢失将引发特征点对齐失效。2)锚点漂移的全局误差。多模态提示范式过度依赖锚点的绝对精度,一旦物理测距信号受气象噪声干扰发生漂移,将引发不可逆的全局推断误差。各类方法的核心策略、优势与局限性如表1所示。

3 全景视场的单目深度估计

全景单目深度估计面临ERP引发的球面非欧几里得畸变,严重制约传统二维卷积网络的局部感受野与全局结构理解。为应对该挑战,相关研究经历了3个关键的技术转折点:第一阶段(图像级)通过局部透视重投影或多面体切片规避畸变,实现局部高精度特征提取与深度预测;第二阶段(特征级)通过隐空间多投影融合、频域解耦等策略,强化全景深度的全局一致性和连续性;第三阶段(先验级)依托视觉基础大模型知识蒸馏或原生全景大模型,通过统一光线几何建模和跨投影特征增强,实现零样本泛化与绝对尺度预测,并探索在环境退化叠加全景畸变下的鲁棒性。

3.1 基于监督学习的单目全景深度估计

在具备真实深度标签的条件下,监督学习方法构成了全景单目深度估计的研究基础,其技术演进经历了从局部重投影、隐空间多投影融合到数据表示重构与多任务正则化的过程。

3.1.1 局部透视重投影与高分辨率拼接

早期研究通过将球面图像拆解为多个局部透视切面来规避畸变。OmniFusion利用球极平面投影提取非均匀分布的切向图像块,结合几何嵌入

表1　复杂天气下各类单目深度估计方法对比

Tab. 1　Comparison of monocular depth estimation methods under adverse weather conditions

技术路径	代表性工作	核心策略	优势	局限性
生成式增强与大模型先验	文献[10]	采用单步扩散模型生成雨雾图像	能够严格保持几何结构,推理速度提升显著	在极端物理遮挡(如大雨滴)情况下生成图像易失真
	文献[2]	在大规模数据集(6200 万幅无标签图像)上进行训练	具有卓越的全局结构恢复能力与极强的零样本泛化能力	无法显式处理极端物理遮挡与纹理缺失问题;处理高反光或透明物体时易产生伪影
	文献[12]	冻结扩散模型以提取生成式先验	实现零样本跨域迁移,边缘锐度显著优于 DPT 与 ZoeDepth 方法	马尔可夫链迭代导致推理延迟极高;输出为相对深度,缺乏绝对尺度感知能力
	文献[13]	提取 ViT 的全局图像特征并嵌入骨干模型	突破局部特征局限,擅长恢复连贯物理结构	扩散采样开销大;高度依赖的透视全局先验在全景 ERP 极区极易发生扭曲
	文献[14]	引入光谱门控机制	具有显著的抗域偏移能力,在极端天气基准上展现出极强的零样本鲁棒性	频域转换与多级扩散推理带来较高显存开销;缺乏对叠加条件的应对机制
自监督跨域适应	文献[15]	在隐空间的代价体内传递合成域与真实域特征	克服了单纯风格迁移的几何失真,精度优于 md4all 和 MonoDepth2 等传统蒸馏基线	仅为隐式统计适应,对光学退化的物理机理建模不足;在动态场景中性能显著下降
	文献[16]	白昼-环境退化跨域一致性蒸馏	实现稳定的跨域特征对齐,规避了夜间光度损失不可靠的难题	精度上限受限于教师网络;易继承教师网络在反光或动态物体上的预测错误
	文献[18]	采用反射感知三元组挖掘与自适应掩码	免疫非朗伯体反射干扰,能够忽略水面反光等失效信号,突破光度梯度爆炸瓶颈	掩码区域退化为依赖全局先验,难以保证高频细节;缺乏对全景拉伸反光畸变的物理建模
	文献[17]	引入多教师模型协同蒸馏大模型先验	显著提升模型的零样本泛化能力,在保留边缘锐度的同时实现绝对尺度的恢复	在全景大视场与环境退化叠加场景中,其高度依赖的透视几何先验会发生非线性扭曲
物理约束与多模态融合	文献[19]	利用傅里叶变换将图像解耦,进行振幅置换	实现无监督昼夜自适应,有效降低光照干扰	依赖全局频域解耦,对局部极端眩光处理能力有限;频域转换等操作带来额外的计算与显存开销
	文献[20]	引入光照感知掩码机制,构建联合增强学习	针对夜间环境进行专项提升优化,在夜间测试数据集上的指标显著提升	场景适配性有限,白昼场景泛化极差;图像增强模块无法恢复极度黑暗区域
	文献[23]	雷达点云上采样与 2D-3D 图神经网络特征联合聚合	克服雷达极度稀疏瓶颈,支持不固定点数输入,远距离边界预测更清晰	训练过程依赖高成本 LiDAR 真值;未解决全景非欧几何空间中点云拓扑拉伸所引发的特征映射偏移
	文献[22]	将稀疏 LiDAR 点云转化为几何提示,以引导大模型	突破分辨率极限的度量感知,实现了 4K 分辨率下的绝对尺度解码	高度依赖雷达锚点精度;在雨、雾和全景大视场下易发生非线性漂移,误差将全局传播
	文献[21]	扩散特征引导与无风格内容重建	能够抵御内窥镜手术等极端场景中的物理光照干扰,展现出极强的鲁棒性	架构针对特定硬件高度定制,向宏观全景感知场景的迁移能力尚待验证

网络注入 3D 球面位置编码[24],并通过视觉 Transformer 算法[25]实现块级全局聚合,最终获得高锐度的边界深度信息。PanoFormer 进一步优化了切向块间的特征交互,提出了配备可学习 Token 流的全景 Transformer 方法[26]。针对高分辨率全景场景,360MonoDepth 方法[27]采用正二十面体提取切面图像,调用透视域基础模型 MiDaS[28-29]预测局部视差,随后通过多尺度可变形对齐场与泊松梯度融合生成无缝的高分辨率全景深度图,突破了传统 CNN 的显存与分辨率瓶颈。

3.1.2　跨投影域特征融合与隐空间对齐

为避免烦琐的重投影及其拼接伪影,研究者逐步转向隐空间内多投影表征协同。BiFuse 方法[30]首次提出 ERP 与立方体投影之间的双向互补融合理念。UniFuse 方法[31]则证实从 CMP(cubemap projection format)到 ERP 的单向融合更为高效。为进一步提升融合质量,GLPanoDepth 方法[32]利用立方体视觉 Transformer 提取的长程特征与 CNN 局部结构特征,并首创门控融合模块在多尺度上自适应聚合。Cross360 方法[33]则构建跨投影特征对齐模块,利用交叉注意力模块融合局部切线投影特征与具有 360°视场的 ERP 特征,从算子层面缓解了球面投影边缘的不连续性。此外,针对局部投影特征统一至等距柱状格式时引入的二次畸变问题,PanoGabor 方法[34]引入畸变感知 Gabor 滤波器,利用其方向与尺度敏感的局部空间频率分析能力,并通过通道与空间的单向融合有效克服了 ERP 的非欧几何形变误差。最后,Elite360D 方法[35]采用几何分布绝对均匀的二十面体投影离散点集作为全局特征载体,通过双投影与双重注意力融合模块,以极低的参数增量实现跨域全局特征对齐,具备良好的边缘设备部署潜力。

3.1.3　室内先验与序列化空间建模

除了直接应对畸变问题,部分研究从网络架构适配性方面寻求突破。SliceNet 方法[36]借助室内重力对齐先验,将全景图切分为垂直切片序列并利用 LSTM 网络[37]建模空间连续性,实现无显

式球面投影的深度解码。SphereDepth方法[38]则直接在三维球面网格上利用可变形卷积提取特征并预测深度，消除二维投影所引入的几何失真。

3.1.4 多任务结构正则化与几何先验引导

引入额外的结构约束，开展多任务联合学习，是提升监督模型泛化能力的重要途径。早期方法多依赖特定的室内强几何假设以纠正局部畸变，如文献[39]引入平面感知机制，以主曲率作为边界指示器，强制同一平面内的深度保持法线一致性。LayoutDepth方法[40]将三维房间布局与深度估计任务深度绑定，利用布局回归生成的深度图作为强空间注意力先验，有效纠正了因纹理遮挡引发的深度畸变。HRDFuse方法[41]将传统深度回归任务转化为深度分布直方图分类，大幅抑制深度离群值的产生。BGDNet方法[42]则借助视觉大模型离线提取具有明确物理意义且精确的确定性背景深度，在推理阶段对前景深度预测结果施加强约束，为真实场景泛化提供了新方向。

3.2 基于自监督学习的单目全景深度估计

受限于全景场景高密度深度真值的高昂采集成本，基于光度一致性与自监督重构的方法逐步成为该领域的研究热点。此类方法无需依赖深度标注，主要通过视图合成或隐式场重建实现网络的无监督优化。

3.2.1 全景域的可微重投影与位姿优化

Pano-SfMLearner[43]标志着端到端全景自监督深度估计框架的成熟，它构建了基于ERP的可微视图合成模块，并联合自监督超像素语义分割，利用弱语义结构引导弱纹理区域的深度收敛。针对全景视频序列中真实光照剧变所产生的噪声梯度，BiFuse++方法[44]在双投影融合架构中引入对比度感知光度损失，大幅提升了模型在非理想光照条件下自监督学习的收敛稳定性。此外，针对自监督运动恢复结构框架在平移估计中存在固有的尺度模糊缺陷，PanoPose方法[45]构建了全景相对位姿估计网络，通过将深度网络提取的稠密尺度信息注入位姿网络，以输出绝对尺度平移量，并利用全景视场的连续性进行预训练，显著降低了位姿漂移。

3.2.2 神经渲染与隐式场重建

受神经渲染的启发，CUBE360方法[46]从单张全景图学习包含密度与颜色信息的连续立方场，将ERP图投影至6个CMP面，并利用神经网络预测每个面的多平面图像[47]，结合射线立方体采样和神经渲染，生成支持物理交互的高质量深度表征并实现动态新视角合成，展现出全景深度感知向真实三维物理重建过渡的发展趋势。

3.2.3 掩码预训练与跨模态重建

在全景深度补全任务中，M3PT方法[48]引入掩码自编码器[49]，在自监督预训练阶段，对全景RGB图像与稀疏深度输入同步施加随机掩码，迫使网络利用上下文信息推断并重构被掩盖的稀疏深度信号。这种跨模态预测的自监督模式有效引导网络学习到全景空间结构耦合关系，为后续监督微调提供鲁棒的初始化参数。

3.3 半监督与视觉大模型驱动的全景单目深度估计

将在海量透视图像上预训练的视觉基础大模型（visual foundation model，VFM）所蕴含的通用几何知识高效、无损地迁移至全景域，是突破全景深度估计性能上限的核心前沿方向。针对真实全景深度真值稀缺问题，近期研究实现了从“利用透视大模型进行视角蒸馏”到“构建原生全景视觉大模型与统一相机感知”的跨越式发展。

3.3.1 透视大模型与数据增强的知识蒸馏

利用强泛化的透视域大模型（Depth Anything V2[50]、Depth anything 3[51]、DINOv3[52]）生成伪标签已成为应对全景真值数据极度缺乏的主流方法之一。Depth anywhere[53]构建了半监督知识蒸馏的全新范式，通过透视大模型为CMP生成高质量深度伪标签：离线阶段利用SAM[54]精准剔除不可靠伪标签区域；在线阶段引入随机球面旋转数据增强策略，迫使全景学生网络自主弥合各投影面之间的尺度差异。PanDA[55]框架则创新性地引入莫比乌斯空间数据增强，显著提升了模型对球面极区几何畸变及尺度变换的抗干扰能力。这种依托透视大模型先验来提升全景模型性能上限的技术范式，标志着全景深度估计正式迈入大模型知识迁移的新阶段。

3.3.2 全景通用大模型与统一相机感知

全景深度估计的零样本泛化与尺度模糊难题正被学术界正视。Depth Any Panoramas（DAP）模型[56]通过对海量全景图像特征的深度挖掘与对齐消除投影畸变，展现出卓越的零样本（Zero-shot）预测能力。Depth anything in 360°方法[57]针对无约束全景环境中的尺度模糊问题，深入探讨了自然

场景下的尺度不变性。此外,Depth Any Camera提出了一种零样本度量深度估计框架[58],通过统一的光线几何建模,彻底打破了透视与全景投影的架构壁垒,无缝接收全景相机的成像并直接端到端输出高精度的绝对全景深度图,标志着全景单目深度估计正迈向通用架构的统一。

3.3.3　叠加态环境下的鲁棒性机制探讨

当全景空间畸变叠加雨雪、大雾等极端环境退化时(即“叠加态”),传统方法的性能会急剧下降。在此背景下,视觉大模型的价值进一步凸显:一方面,透视教师模型在海量多样化数据中已学习到物理退化的强先验抗性;另一方面,在全景蒸馏框架(如Depth Anywhere方法[53])中引入离线无效区域掩码,可显式过滤因极区拉伸或恶劣天气遮挡所导致的错误标签。通过在局部无畸变投影面上施加可靠的伪标签监督,并结合全景格式的全局特征约束,当前的知识蒸馏架构正逐步具备抵抗“环境退化叠加空间畸变”的鲁棒性,推动单目全景感知系统走向全天候真实部署。

总体而言,从图像级畸变规避、特征级隐空间融合到先验级大模型蒸馏,全景深度估计的技术演进本质上是一场持续对抗球面非欧几何畸变的过程。各阶段方法在精度、泛化能力和计算效率上各有优势,但在应对环境退化与全景畸变叠加、轻量化部署以及跨场景零样本泛化等方面仍面临挑战。代表性方法的核心策略、优势与局限性如表2所示。

表2　应对全景视场畸变的单目深度估计代表性方法对比

Tab. 2　Comparison of representative monocular depth estimation methods for panoramic spatial distortion

技术路径	代表性工作	核心策略	优势	局限性
监督学习(局部重投影)	文献[24]	提取切片图像特征,引入Transformer进行全局特征聚合	有效缓解了畸变问题,边缘预测精度得到提升	切线补丁的合并计算烦琐;在雨雾等场景中,局部特征缺乏物理全局约束
监督学习(跨投影融合)	文献[30]	提取ERP与CMP双分支特征,进行双向互补融合	兼顾了全局视野与局部无畸变特征提取优势,显著降低了极区误差	双分支网络架构导致参数量翻倍;未能原生适配球面非欧几何特性
	文献[33]	构建跨投影机制,融合多尺度注意力进行深度解码	从算子层面缓解了球面投影的几何不连续性,提升了全局结构的推断精度	模型的计算复杂度限制了分辨率的提升;在恶劣天气遮挡下,易产生重构伪影
监督学习(频域解耦)	文献[34]	引入Gabor滤波器提取频域纹理,进行通道与空间的单向融合	扩展感受野并增强深度线索,解决了投影特征统一至ERP时引入的二次畸变误差	多组滤波器并行导致显存消耗巨大;极端光照等因素会导致频域滤波器失效
自监督学习(视图合成)	文献[43]	基于ERP的可微视图合成模块,联合语义分割进行多任务学习	无需采集高成本的密集深度值,利用语义一致性引导,提升弱纹理区域的深度表现	光度重构损失严格依赖静态环境假设,在全景大视场下易受动态物体干扰
自监督学习(神经渲染)	文献[46]	从单张全景图进行连续立方体场学习与多平面图像预测	支持连续六自由度视角的射线投射采样与沉浸式物理交互渲染	训练阶段的光线投射计算耗时较长;仅适用于静态室内场景
半监督与大模型(知识蒸馏)	文献[53]	结合在线-离线掩码与随机球面旋转数据增强进行联合训练	突破了全景数据稀缺的瓶颈,显著提升了全景深度的零样本泛化性能与精度上限	性能受限于教师模型的精度;在恶劣天气下,生成的错误伪标签易全局放大
半监督与大模型(全景大模型)	文献[58]	基于统一的光线几何建模与海量全景特征对齐	能够适配全景成像模型,消除投影畸变,实现零样本度量深度输出	模型参数量与计算复杂度极高,难以在边缘设备或自动驾驶场景中应用

4　总结与展望

系统梳理了非标准场景下单目深度估计的技术发展脉络,深入剖析了复杂天气引发的环境退化与全景视场所带来的空间畸变两大核心挑战。现有研究进展表明,无论是通过域适应与物理先验嵌入提升模型对环境退化的鲁棒性,还是通过球面重投影与隐式场建模适配全景几何特性,单目深度估计技术正加速突破传统理想成像假设的桎梏。面向未来全天候、全方位的通用深度估计系统,该领域将呈现出4大融合演进趋势。

(1)基础大模型驱动的统一先验迁移

针对非标准场景数据稀缺、分布偏移等问题,依托亿级数据预训练的透视大模型(如Depth

Anything[2])正成为应对上述两类挑战的共性技术方案。在环境退化场景下，大模型可通过轻量级微调[59]提升对局部噪声干扰的鲁棒性；在全景视场下，大模型可通过半监督知识蒸馏为ERP畸变图像提供高质量深度伪标签。

未来，设计更加通用的适配器模块，将大模型的强语义先验同步迁移至环境退化与全景视场的耦合场景，是实现通用鲁棒单目深度感知的核心路径。然而，当前该路径面临着严峻的计算瓶颈。视觉大模型庞大的参数量导致其在车载或边缘计算平台上的微调训练与实时推理成本极高；此外，大模型先验在应对未见过的极端畸变域时，仍存在生成高置信度错误伪标签的风险，如何设计轻量级且抗干扰的适配器，仍是亟待攻克的难题。

(2)环境与空间的叠加态感知挑战

当前研究大多将环境退化鲁棒性与全景感知能力解耦处理。但在自动驾驶等真实场景中，系统往往面临雨夜环境下的全景感知这类叠加态挑战。ERP几何畸变叠加水滴遮挡、动态眩光等环境干扰，将导致自监督学习依赖的光度一致性在球面空间彻底失效。未来研究需探索融合球面几何一致性与物理光学属性的联合重建机制，例如在神经辐射场[60]或3D高斯泼溅[61]的全景渲染框架中，显式建模天气退化分量，以实现叠加态非标准场景下的端到端深度估计。当前的主要瓶颈在于“叠加态”真实配对数据的极度匮乏。早期数据集如ADUULM-360[62]已初步揭示主流算法在全景与恶劣天气叠加下性能呈断崖式下降。在叠加态场景中，透视教师模型极易产生高置信度的“深度幻觉”，且此类误差会随全景感受野的扩张而全局化。

如何在单目视觉下实现物理退化与几何畸变的底层解耦，是未来基础模型跨域迁移必须跨越的挑战。现有的NeRF或3DGS模型在处理静态场景时表现优异，但面对雨雪等动态气象遮挡以及全景相机的高速运动时，其空间一致性极易崩溃，如何在隐式场中实现对气象干扰的物理级解耦仍处于起步阶段。

(3)多模态异构融合与绝对尺度恢复

单目可见光全景相机在浓雾、全黑等极端环境中存在天然的物理感知极限，且单目视觉方法缺乏绝对尺度的恢复能力。为构建真正全天候鲁棒闭环，多模态异构融合是未来必然的技术方案。利用对天气不敏感且自带测距能力的毫米波雷达方案(如R4Dyn[63])、LiDAR等主动传感器，结合全景相机的广角语义信息，引导视觉模型在光学失效区域和畸变极区完成深度补全与校正，将是补齐单目深度估计短板的必由之路。

(4)真实物理数据集的评测基准标准化

阻碍非标准场景深度估计发展的一个隐性瓶颈在于：当前真实物理数据集的评测基准缺乏严格的标准化。在处理诸如Oxford RobotCar等包含真实恶劣天气的多模态数据集时，多模态传感器的时空同步极易引入系统误差。例如，不同算法在生成深度真值时，其底层投影逻辑(如是否对齐雷达与相机间的Z轴空间朝向)往往存在差异；同时，针对恶劣天气下极度稀疏的雷达点云，现有文献在评测阶段通常采用不同的截断策略(如将最大评估深度动态限制在60 m或80 m，视差上限设定各异)。这种底层投影机理与评估阈值的不对齐，严重削弱了现有前沿模型在叠加态下跨模型定量对比的公平性与学术严谨性。因此，构建一套包含严格物理对齐规范与统一评估度量的非标准场景基准，是推动该领域向纵深发展的先决条件。

参考文献：

[1] Narasimhan S G, Nayar S K. Vision and the atmosphere[J]. International Journal of Computer Vision, 2002, 48(3): 233-254.

[2] Yang Lihe, Kang Bingyi, Huang Zilong, et al. Depth anything: unleashing the power of large-scale unlabeled data [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22, 2024, Seattle, WA, USA. Piscataway, NJ: IEEE, 2024: 10371-10381.

[3] Zhou Tinghui, Brown M, Snavely N, et al. Unsupervised learning of depth and ego-motion from video[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. Piscataway, NJ: IEEE, 2017: 6612-6619.

[4] Godard C, Mac Aodha O, Firman M, et al. Digging into self-supervised monocular depth estimation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). October 27-November 2, 2019, Seoul, Korea. Piscataway, NJ: IEEE, 2019: 3827-3837.

[5] Lou Ange, Noble J. WS-SfMLearner: self-supervised monocular depth and ego-motion estimation on surgical videos with unknown camera parameters[J]. Journal of Medical Imaging, 2025, 12(2): 025003.

[6] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//2012

IEEE Conference on Computer Vision and Pattern Recognition. June 16-21,2012,Providence,RI,USA. Piscataway,NJ:IEEE,2012:3354-3361.
[7] Maddern W, Pascoe G, Linegar C, et al. 1 year, 1000 km: the Oxford RobotCar dataset[J]. The International Journal of Robotics Research, 2017, 36(1):3-15.
[8] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[PP/OL]. arXiv(2014-06-10)[2026-03-09]. https://arxiv.org/abs/1406.2661.
[9] Zhu Junyan, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017, Venice, Italy. Piscataway,NJ:IEEE,2017:2242-2251.
[10] Jiang Kui, Cao Jing, Yu Zhaocheng, et al. Always clear depth: robust monocular depth estimation under adverse weather[C]//Proceedings of the 34th International Joint Conference on Artificial Intelligence. August 16-22, 2025. Montreal,Canada. International Joint Conferences on Artificial Intelligence Organization,2025:1251-1259.
[11] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[PP/OL]. v2. arXiv(2020-12-16)[2026-03-09]. https://arxiv.org/abs/2006.11239.
[12] Ke Bingxin, Obukhov A, Huang Shengyu, et al. Repurposing diffusion-based image generators for monocular depth estimation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22, 2024,Seattle,WA,USA. Piscataway,NJ:IEEE,2024:9492-9502.
[13] Patni S, Agarwal A, Arora C. ECoDepth: Effective conditioning of diffusion models for monocular depth estimation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22,2024,Seattle,WA,USA. Piscataway,NJ:IEEE,2024:28285-28295.
[14] Cai Xinhao, Pei Gensheng, Sun Zeren, et al. Iris: bringing real-world priors into diffusion model for monocular depth estimation[PP/OL]. v2. arXiv(2026-03-19)[2026-05-14]. https://arXiv.org/abs/2603.16340.
[15] Yan Weilong, Li Ming, Li Haipeng, et al. Synthetic-to-real self-supervised robust depth estimation via learning with motion and structure priors[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 10—17,2025,Nashville,TN,USA. Piscataway,NJ:IEEE,2025:21880-21890.
[16] Gasperini S, Morbitzer N, Jung H J, et al. Robust monocular depth estimation under challenging conditions[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). October 1-6,2023,Paris,France. Piscataway,NJ:IEEE,2023:8143-8152.
[17] He Xiankang, Guo Dongyan, Li Hongji, et al. Distill any depth: distillation creates a stronger monocular depth estimator[PP/OL]. v2. arXiv (2025-04-21)[2026-05-08]. https://arXiv.org/abs/2502.19204.
[18] Choi W, Hwang K, Wei Peng, et al. Self-supervised monocular depth estimation robust to reflective surface leveraged by triplet mining[PP/OL]. arXiv (2025-02-20)[2026-04-21]. https://arXiv.org/abs/2502.14573.
[19] Shen Mengjiao, Wang Zhongyi, Su Shuai, et al. DNA-depth: a frequency-based day-night adaptation for monocular depth estimation[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 2530112.
[20] Zheng Yupeng, Zhong Chengliang, Li Pengfei, et al. STEPS: joint self-supervised nighttime image enhancement and depth estimation[C]//2023 IEEE International Conference on Robotics and Automation (ICRA). May 29-June 2, 2023,London,United Kingdom. Piscataway,NJ:IEEE,2023:4916-4923.
[21] Wang Zhongyang, Xu Shengjin, Xu Yechang, et al. D^2FNet: degradation-aware diffusion feature guided network against physical-world attack for monocular depth estimation[J/OL]. IEEE Transactions on Consumer Electronics, 2026-03-13. https://ieeexplore.ieee.org/document/11434522.
[22] Lin Haotong, Peng Sida, Chen Jingxiao, et al. Prompting depth anything for 4K resolution accurate metric depth estimation[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 10-17, 2025,Nashville,TN,USA. Piscataway,NJ:IEEE,2025:17070-17080.
[23] Sun Huawei, Wang Zixu, Feng Hao, et al. GET-UP: gEomeTric-aware depth estimation with radar points UPsampling[PP/OL]. v2. arXiv(2024-09-08)[2026-04-21]. https://arXiv.org/abs/2409.02720.
[24] Li Yuyan, Guo Yuliang, Yan Zhixin, et al. OmniFusion: 360 monocular depth estimation via geometry-aware fusion[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans,LA,USA. Piscataway,NJ:IEEE,2022:2791-2800.
[25] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[PP/OL]. v2. arXiv(2021-06-03)[2026-05-16]. https://arXiv.org/abs/2010.11929.
[26] Shen Zhijie, Lin Chunyu, Liao Kang, et al. PanoFormer: Panorama transformer for indoor 360 depth estimation[C]// Avidan S, Brostow G, Cissé M, et al. Computer Vision-ECCV 2022. Cham:Springer,2022:195-211.
[27] Rey-Area M, Yuan Mingze, Richardt C. 360MonoDepth: high-resolution 360° monocular depth estimation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24,2022,New Orleans,LA, USA. Piscataway,NJ:IEEE,2022:3752-3762.
[28] Ranftl R, Lasinger K, Hafner D, et al. Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(3):1623-1637.
[29] Ranftl R, Bochkovskiy A, Koltun V. Vision transformers for dense prediction[PP/OL]. (2021-03-24)[2026-03-07]. https://arXiv.org/abs/2103.13413.
[30] Wang Fuen, Yeh Y H, Sun Min, et al. BiFuse: monocular 360 depth estimation via Bi-projection fusion[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19,2020,Seattle,WA,USA. Piscataway,NJ:IEEE,2020:459-468.
[31] Jiang Hualie, Sheng Zhe, Zhu Siyu, et al. UniFuse: unidirectional fusion for 360° panorama depth estimation[J]. IEEE Robotics and Automation Letters, 2021, 6(2):1519-1526.
[32] Bai Jiayang, Qin Haoyu, Lai Shuichang, et al. GLPanoDepth: global-to-local panoramic depth estimation[J]. IEEE Transactions on Image Processing, 2024, 33:2936-2949.
[33] Huang Kun, Zhang Fanglue, Dodgson N. Cross360: 360° monocular depth estimation via cross projections across scales[PP/OL]. arXiv(2026-01-24)[2026-05-08]. https://arXiv.org/abs/2601.17271.
[34] Shen Zhijie, Lin Chunyu, Nie Lang, et al. Revisiting 360 depth estimation with PanoGabor: a new fusion perspective [PP/OL]. v5. arXiv(2026-02-04)[2026-05-08]. https://arXiv.org/abs/2408.16227.
[35] Ai Hao, Wang Lin. Elite360D: towards efficient 360 depth estimation via semantic- and distance-aware Bi-projection

fusion[PP/OL]. v2. arXiv(2024-05-25)[2026-03-07]. https://arXiv.org/abs/2403.16376.
[36] Pintore G, Agus M, Almansa E, et al. SliceNet: deep dense depth estimation from a single indoor panorama using a slice-based representation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20-25,2021,Nashville,TN,USA. Piscataway,NJ:IEEE,2021:11531-11540.
[37] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
[38] Yan Qingsong, Wang Qiang, Zhao Kaiyong, et al. SphereDepth: panorama depth estimation from spherical domain [C]//2022 International Conference on 3D Vision (3DV). September 12-16, 2022, Prague, Czech Republic. Piscataway,NJ:IEEE,2022:1-10.
[39] Eder M, Moulon P, Guan Li. Pano popups: indoor 3D reconstruction with a plane-aware network[C]//2019 International Conference on 3D Vision (3DV). September 16-19, 2019, Quebec City, QC, Canada. Piscataway, NJ: IEEE,2019:76-84.
[40] Zeng Wei, Karaoglu S, Gevers T. Joint 3D layout and depth prediction from a single indoor panorama image[C]// Vedaldi A, Bischof H, Brox T, et al. Computer Vision-ECCV 2020. Cham: Springer, 2020:666-682.
[41] Ai Hao, Cao Zidong, Cao Yanpei, et al. HRDFuse: monocular 360° depth estimation by collaboratively learning holistic-with-regional depth distributions[PP/OL]. v3. arXiv(2023-05-22)[2026-03-07]. https://arXiv.org/abs/2303.11616.
[42] Chen Jiajing, Wan Zhiqiang, Narayana M, et al. BGDNet: background-guided indoor panorama depth estimation [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). June 17-18, 2024,Seattle,WA,USA. Piscataway,NJ:IEEE,2024:1272-1281.
[43] Liu Mengyi, Wang Shuhui, Guo Yulan, et al. Pano-SfMLearner: self-supervised multi-task learning of depth and semantics in panoramic videos[J]. IEEE Signal Processing Letters, 2021, 28:832-836.
[44] Wang Fuen, Yeh Y H, Tsai Y H, et al. BiFuse++: self-supervised and efficient Bi-projection fusion for 360° depth estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(5):5448-5460.
[45] Tu Diantao, Cui Hainan, Zheng Xianwei, et al. PanoPose: self-supervised relative pose estimation for panoramic images[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22,2024, Seattle,WA,USA. Piscataway,NJ:IEEE,2024:20009-20018.
[46] Chang Wenjie, Ai Hao, Zhang Tianzhu, et al. CUBE360: learning cubic field representation for monocular 360 depth estimation for virtual reality[PP/OL]. arXiv(2024-10-08)[2026-03-07]. https://arXiv.org/abs/2410.05735.
[47] Zhou Tinghui, Tucker R, Flynn J, et al. Stereo magnification: learning view synthesis using multiplane images[J]. ACM Transactions on Graphics, 2018, 37(4):1-12.
[48] Yan Zhiqiang, Li Xiang, Wang Kun, et al. Multi-modal masked pre-training for monocular panoramic depth completion [C]//Avidan S, Brostow G, Cissé M, et al. Computer Vision- ECCV 2022. Cham: Springer, 2022:378-395.
[49] He Kaiming, Chen Xinlei, Xie Saining, et al. Masked autoencoders are scalable vision learners[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans, LA, USA. Piscataway,NJ:IEEE,2022:15979-15988.
[50] Yang Lihe, Kang Bingyi, Huang Zilong, et al. Depth anything V2[PP/OL]. v2. arXiv(2024-10-20)[2026-05-16]. https://arXiv.org/abs/2406.09414.
[51] Lin Haotong, Chen Sili, Liew J, et al. Depth anything 3: recovering the visual space from any views[PP/OL]. arXiv (2025-11-13)[2026-05-16]. https://arXiv.org/abs/2511.10647.
[52] Siméoni O, Vo H V, Seitzer M, et al. DINOv3[PP/OL]. arXiv(2025-08-13)[2026-05-16]. https://arXiv.org/abs/2508.10104.
[53] Wang Ninghsu, Liu Yulun. Depth anywhere: enhancing 360 monocular depth estimation via perspective distillation and unlabeled data augmentation[PP/OL]. v2. arXiv(2024-10-30)[2026-03-07]. https://arXiv.org/abs/2406.12849.
[54] Kirillov A, Mintun E, Ravi N, et al. Segment anything[PP/OL]. arXiv(2023-04-05)[2026-03-07]. https://arXiv.org/abs/2304.02643.
[55] Cao Zidong, Zhu Jinjing, Zhang Weiming, et al. PanDA: towards panoramic depth anything with unlabeled panoramas and möbius spatial augmentation[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 10-17,2025,Nashville,TN,USA. Piscataway,NJ:IEEE,2025:982-992.
[56] Lin Xin, Song Meixi, Zhang Dizhe, et al. Depth any panoramas: a foundation model for panoramic depth estimation [PP/OL]. arXiv(2025-12-18)[2026-05-14]. https://arXiv.org/abs/2512.16913.
[57] Jiang Hualie, Song Ziyang, Lou Zhiqiang, et al. Depth anything in 360°: towards scale invariance in the wild[PP/OL]. arXiv(2025-12-28)[2026-05-14]. https://arXiv.org/abs/2512.22819.
[58] Guo Yuliang, Garg S, Miangoleh S M H, et al. Depth any camera: zero-shot metric depth estimation from any camera [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 10-17, 2025, Nashville,TN,USA. Piscataway,NJ:IEEE,2025:26996-27006.
[59] Hu E J, Shen Yelong, Wallis P, et al. LoRA: low-rank adaptation of large language models[PP/OL]. v2. arXiv (2021-10-16)[2026-05-16]. https://arXiv.org/abs/2106.09685.
[60] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: representing scenes as neural radiance fields for view synthesis [J]. Communications of the ACM,2021,65(1):99-106.
[61] Kerbl B, Kopanas G, Leimkuehler T, et al. 3D gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4):1-14.
[62] Schön M, Ruof J, Wodtko T, et al. The ADUULM-360 Dataset: a multi-modal dataset for depth estimation in adverse weather[C]//2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC). September 24-27,2024,Edmonton,AB,Canada. Piscataway,NJ:IEEE,2024:1403-1409.
[63] Gasperini S, Koch P, Dallabetta V, et al. R4Dyn: exploring radar for Self-Supervised monocular depth estimation of dynamic scenes[C]//2021 International Conference on 3D Vision (3DV). December 1-3, 2021. London, United Kingdom. Piscataway,NJ:IEEE,2021:751-760.

责任编辑:申晓伟　郭红建