



沈阳建筑大学

硕士学位论文开题报告

姓 名 : 王佳伟

学 号 : 2434290024

学科 类别 领域: 计算机科学与技术

指 导 教 师 : 袁帅

培 养 类 别 : 全日制 非全日制

学 位 类 型 : 学术学位 专业学位

开 题 日 期 : 2025 年 10 月 31 日

所在培养单位 : 计算机科学与工程学院

沈阳建筑大学研究生院制

填表说明

1. 本表由研究生在导师指导下如实填写；
2. 封面“学科类别领域”应严格按照招生录取时的学科类别领域名称填写；
3. 封面“培养类别”和“学位类型”在相应框内划“√”；
4. 表中所有签字部分，均须本人签字，不得代签；
5. 除表中明确注明可附页的部分外，不得更改本表格式；
6. 本表一式一份，全部内容填写及签字完成后，由研究生扫描成 pdf 文件，并将原件及扫描完成的电子版报所在培养单位存档。

论文题目	基于语义先验图与目标边缘增强的 条件扩散模型单目深度估计
课题来源	国家自然科学基金项目(62073227),辽宁省科技厅项(2023JH2/101300212)

一、选题背景与意义

日常生活中，人类通过双眼感知世界时能够自然地判断物体的远近关系，从而获得三维空间中的深度信息。然而，相机拍摄的图像本质上是三维场景在二维平面上的投影，丢失了至关重要的深度维度。为了使计算机视觉系统具备类似人类的空间感知能力，深度估计技术应运而生。深度估计旨在从单张或多个视角的图像中恢复场景中每个像素点到摄像机的距离信息，重建出具有几何意义的深度图。简而言之，深度估计就是从二维图像中“还原”三维结构的过程，其目标是尽可能准确地恢复原始场景的空间布局。作为计算机视觉领域的一项基础性任务，深度估计不仅是理解三维场景的关键环节，也为诸多下游应用提供了不可或缺的信息支持，具有广泛的应用价值，主要体现在以下几个方面：

(1) **自动驾驶。**智能车辆需要实时感知周围环境以实现路径规划、障碍物避让、车道保持等关键功能。仅依靠 RGB 图像难以判断前方车辆、行人或其他障碍物的实际距离，而通过深度估计技术可以从摄像头采集的图像中推断出场景的三维结构，为决策系统提供精确的距离信息，显著提升行车安全性。

(2) **机器人导航与交互。**服务机器人、仓储搬运机器人等需要在复杂动态环境中自主移动并完成抓取、避障等操作。深度估计能够帮助机器人构建环境的三维地图，识别可通行区域与障碍物位置，增强其空间认知能力，是实现精准定位与导航的核心技术之一。

(3) **增强现实 (AR) 与虚拟现实 (VR)。**在 AR/VR 应用中，虚拟对象需与真实场景无缝融合，这就要求系统精确了解真实环境的几何结构。深度估计提供了场景的表面形状与空间层次信息，使得虚拟物体能够正确遮挡或被真实物体遮挡，极大提升了沉浸感与交互真实性。

(4) **无人机飞行与航拍建模。**无人机在执行地形测绘、建筑建模、灾害评估等任务时，常需基于视觉信息进行三维重建。深度估计可辅助生成高精度的点云模型与数字高程图，克服传统 GPS 定位精度不足的问题，尤其适用于低空精细作业场景。

在当前技术快速发展的背景下，深度估计已成为连接二维视觉感知与三维空间理解的重要桥梁。传统方法依赖立体匹配或多视角几何约束，近年来随着深度学习的发展，基于卷积神经网络和 Transformer 架构的深度估计算法不断涌现，显著提升了估计精度与泛化能力。然而，深度估计的效果往往直接影响后续任务的性能表现，输入图像的质量与场景特性也给深度估计带来了诸多挑战，主要包括以下几类：

(1) **弱纹理或无纹理区域。**如白墙、天空、雪地等表面缺乏明显纹理特征，导致局部梯度变化微弱，传统立体匹配算法难以找到可靠的对应点，深度网络也容易在此类区域产生模糊或错误的预测。

(2) **光照变化剧烈与噪声干扰图像。**低光照条件下图像信噪比降低，细节缺失；强光反射或阴影则造成亮度突变，这些都会误导深度估计模型对边缘和轮廓的判断，影响整体深度图的连续性与准确性。

(3) **重复纹理与周期性结构。**例如瓷砖地面、玻璃幕墙、书架等场景中存在大量相似图案，使得匹配过程出现歧义，导致误匹配，进而引发深度跳跃或伪影现象。

(4) **透明或反光物体。**玻璃、水面、镜面等材质会折射或反射背景信息，破坏图像一致性假设，使得基于外观一致性的深度估计方法失效，成为极具挑战性的难题。

(5) **运动模糊与动态场景。**当场景中存在快速移动的物体或相机自身抖动时，图像会出现拖影或

失真，严重影响特征提取与视差计算，导致深度图出现断裂或漂移。

其中，动态场景下的深度估计由于前景物体与背景之间存在相对运动，打破了静态场景假设，导致多帧间对应关系复杂化，难以准确分离刚性背景与非刚性运动物体的深度信息，因此成为当前研究的重点与难点。

综上所述，在人工智能与计算机视觉迅猛发展的今天，深度估计作为实现机器“看懂”三维世界的基础技术，已在自动驾驶、机器人、智能交互等多个前沿领域展现出巨大潜力。尽管已有诸多进展，但在复杂真实场景下面临的挑战依然严峻。如何高效、鲁棒、精确地完成深度估计任务，不仅关乎模型本身的性能提升，更直接影响上层应用的安全性与可靠性。因此，深度估计是一项兼具理论价值与实践意义的研究课题，具有广阔的发展前景和重要的研究必要性。

二、国内外研究现状及发展趋势

1. 数据驱动的深度估计方法

单目深度估计 (Monocular Depth Estimation, MDE) 的核心挑战在于从二维图像中恢复三维几何结构，这一任务本质上是有潜在问题的 (ill-posed)，因为深度信息在成像过程中被严重压缩。早期研究主要依赖多视角几何或结构光等硬件辅助手段，而随着深度学习的发展，基于大规模标注数据的数据驱动方法逐渐成为主流。

这类方法通常采用端到端的卷积神经网络 (CNN) 或视觉 Transformer 架构，通过在大规模图像-深度对数据集上进行监督训练，学习从 RGB 图像到深度图的映射关系。Eigen 等人[7]首次将卷积神经网络引入深度估计任务，提出了分层预测框架，在 NYU Depth 和 KITTI 等基准数据集上取得了突破性进展。后续工作从多个方向进行了改进：(1) 网络结构优化，如引入残差连接[8]、多尺度特征融合[2]、注意力机制[18]以及 Vision Transformer (ViT) 结构[21]，以增强模型的全局感知能力；(2) 损失函数设计，包括分类-回归混合范式、尺度不变损失、边缘感知梯度损失等，以提升预测的几何一致性；(3) 多任务学习策略，联合估计表面法向量、语义分割等辅助任务，以提供额外的几何与语义约束。

为了提升模型在未见场景中的泛化能力，部分人开始构建更大规模、更丰富的数据集。MiDaS[10]提出预测相对深度而非绝对深度，通过在多个异构数据集上联合训练，实现了跨域零样本迁移。Omnidata[6]进一步扩展了训练数据至约 1450 万张图像，训练出具有强泛化能力的通用深度估计模型。Depth Anything[18]则利用 6200 万张无标签图像进行自监督预训练，显著提升了模型在复杂真实场景下的鲁棒性。

尽管数据驱动方法在特定数据集上表现优异，但其性能高度依赖于训练数据的质量与覆盖范围，且模型难以适应光照、场景布局剧烈变化的真实世界环境。此外，采集高质量深度标签成本高昂，限制了模型的持续优化。因此，近年来研究重点逐渐转向无需大量标注数据的模型驱动方法。

2. 模型驱动的深度估计方法：基于扩散模型

为突破数据依赖的瓶颈，模型驱动方法开始兴起，其核心思想是利用在超大规模图像-文本对上预训练的生成模型（如 Stable Diffusion）中蕴含的强大先验知识，将其迁移至深度估计等判别性任务中。这类方法无需在深度数据上进行大规模训练，即可实现“零样本” (zero-shot) 深度估计，展现出卓越的泛化能力。

Marigold[22]是模型驱动方法的开创性工作，首次将单目深度估计重构为扩散-去噪过程。具体而言，输入图像被编码至潜在空间，随后扩散模型通过多步迭代去噪生成深度图。该方法充分利用了扩散模型对细节的高保真重建能力，在细节保留和结构合理性方面显著优于传统判别模型。GeoWizard[9]在此基础上引入法向量估计作为几何先验，进一步提升了深度图的几何一致性。

然而，基于迭代去噪的扩散方法推理速度慢，难以满足实时应用需求。为此，DepthFM[12]引入流匹配（Flow Matching）技术，减少采样步数以加速推理。更进一步，GenPercept[29]提出单步确定性范式（single-step deterministic paradigm），即直接将图像潜在表示输入去噪 U-Net，一步输出深度潜在图，大幅提升了推理效率，同时保持了与多步方法相当的性能。Lotus[14]也采用单步框架，并设计细节保持分支以缓解因跳过迭代过程而导致的预测模糊问题。

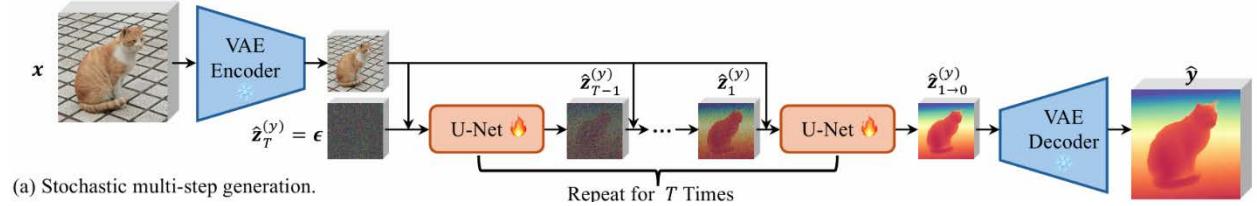


图 1.利用扩散模型进行深度估计的过程

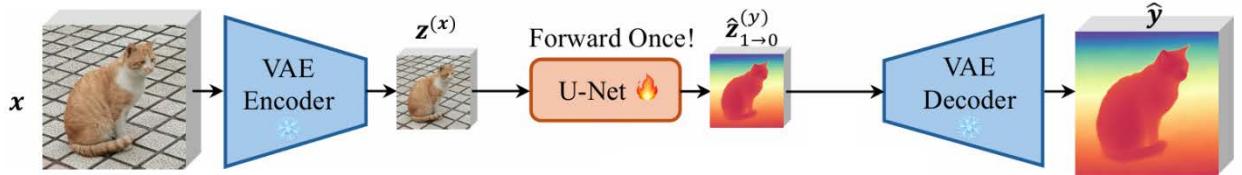


图 2.单步推理的扩散模型范式

尽管这些方法在效率与性能之间取得了良好平衡，但它们普遍忽视了生成式特征与判别式任务之间的本质差异。扩散模型的去噪网络在预训练阶段专注于图像重建，倾向于关注纹理细节而非高层语义结构，导致深度预测可能出现“伪纹理”现象，缺乏真实几何结构。此外，单步推理框架难以在一次前向传播中同时捕捉全局结构与高频细节，导致预测结果模糊。

针对上述问题，DepthMaster[28] 提出了一套系统性解决方案：首先，引入特征对齐模块（Feature Alignment），通过引入 DINOv2 等高质量外部语义编码器的特征，引导去噪网络关注语义结构，缓解对纹理细节的过拟合；其次，设计傅里叶增强模块（Fourier Enhancement），在频域中自适应融合低频结构与高频细节，模拟多步去噪过程的渐进式恢复机制；最后，采用两阶段训练策略，先在潜在空间优化全局结构，再在像素空间精修细节，实现了性能与质量的双重提升。

综上所述，基于扩散模型的深度估计方法正从“直接迁移生成模型”向“针对性改造生成架构以适配判别任务”演进。通过融合外部语义信息、引入频域建模、优化训练流程，新一代方法在保持零样本泛化能力的同时，显著提升了深度预测的准确性与视觉质量，为无需标注数据的通用深度感知提供了新的技术路径。

三、研究内容

我们针对现有基于扩散模型的单目深度估计（MDE）方法在推理速度和细节保留方面的局限性，提出了创新的解决方案。为大幅提升推理效率，我们提出了确定性的单步推理范式，将 640*480 图像的推理时间从 24 秒加速到 121 毫秒。针对单步推理中固有的局部细节丢失问题，根据我们的重参数化实验可知，扩散模型的知识先验都存储在 U-Net 网络中，因此我们选择提供给扩散模型更多的语义信息，该方法能够为 U-Net 提供更丰富的特征表示有效平衡推理效率和细节保留。

此外，为优化整体结构和局部细节，我们还提出了一种提升细节特征的指导方式；传统的扩散模型通过 MSE 优化全局信息，但 MSE 对全局像素都“一致对待”，导致训练过程中对边缘细节的处理不到

位。由此我们计划提出一种关注细节的指导方式，在 MSE 梳理全局细节的同时利用其它指导方法来增强图像细节。

四、理论依据与技术方案

1. 扩散模型 (Diffusion models) 作为一类生成模型，已成为数据合成和密集预测任务的强大框架。其核心原理包含一个双步骤过程：一个前向过程 (forward pass)，逐渐向数据中注入噪声；以及一个反向过程 (reverse pass)，学习如何去噪，从而有效地从噪声中重建数据。

扩散模型的发展经历了几个关键里程碑。其概念最初受到非平衡热力学的启发。一个重大突破是提出的去噪扩散概率模型 (Denoising Diffusion Probabilistic Models, DDPMs)[14]，它通过预测注入的噪声，建立了一个简单且稳定的训练目标。随后的工作通过基于分数的生成建模 (score-based generative modeling) 和随机微分方程 (stochastic differential equations) 推广了这一观点。最近引入的潜在扩散模型 (Latent Diffusion Models, LDMs) 通过在低维潜在空间中执行扩散过程，进一步提高了计算效率，实现了高分辨率图像合成，并促进了其在包括单目深度估计在内的各种条件生成任务中的应用。

扩散模型通过定义一个逐渐用高斯噪声破坏的数据 z_0 的前向过程，然后学习一个反向过程来恢复数据，这个反向的过程拟合的分布我们可以用 $p(z_0)$ 来定义。

前向过程是一个固定的马尔可夫链 (Markov chain)，在 T 个步骤中添加噪声，其加噪的过程满足加噪公式

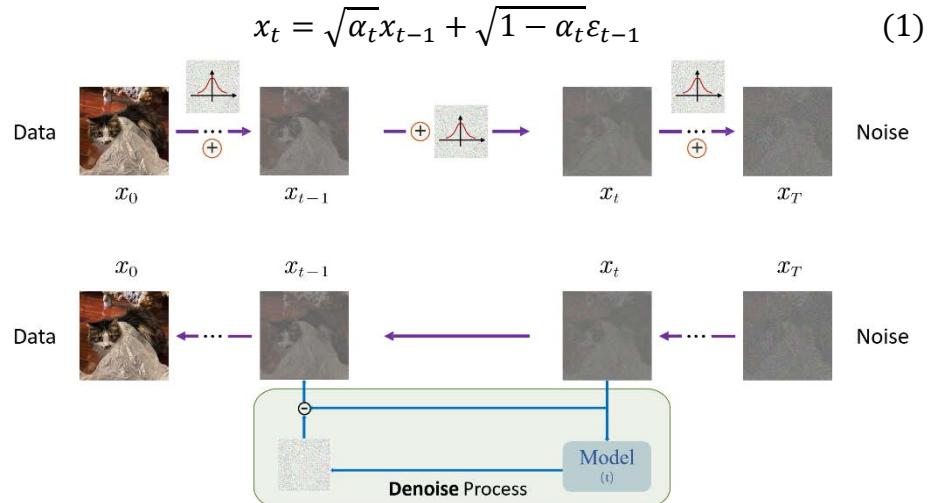


图 3. 扩散模型的简单示意图

其中的 ε_t 是一个服从 01 高斯分布的纯噪声

反向过程是一个参数化的马尔科夫链，从 $p(z_T) = \mathcal{N}(z_{T-1}; 0, I)$ 开始并学习去噪过程

$$p_\theta(z_{t-1}|z_t) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t), \Sigma_\theta(z_t, t)) \quad (2)$$

目标是匹配真实的反向分布 $q(z_{T-1}|z_T, z_0)$

2. 条件扩散模型 (Conditional Diffusion Models) 是对标准扩散模型的一种扩展，使其能够根据特定的输入信息 (即“条件”) 来生成目标数据，而不是仅仅随机生成。在像单目深度估计这样的预测任务中，模型需要根据输入的 RGB 图像来生成对应的深度图。

具体来说，这是通过在去噪过程的每一步中，都将这个条件信息 (例如，RGB 图像的特征) 馈送给噪声预测神经网络来实现的。这样，模型在预测噪声时不仅会考虑当前的含噪数据 z_t 和时间步 t ，还

会考虑输入的条件 x (即 $\epsilon(z_t, t, x)$)。

这种机制引导生成过程，确保最终的输出结果（例如深度图）与输入的条件信号（例如 RGB 图像）保持一致。这使得模型能够学习到从输入到输出的条件分布 $p(z_0|x)$ ，这对于完成各种具体的预测任务至关重要。

五、创新点和预期结论

1. 创新点：

(1) 我们提出了一种确定性的单步推理 (deterministic single-step inference) 范式，用以替代传统扩散模型在深度估计任务中常见的多步随机去噪过程。该方法通过一次前向传播即可直接从潜空间映射到干净的深度预测结果，无需迭代采样，从而极大地降低了推理时间——在 640×480 分辨率下，推理速度由约 24 秒提升至 121 毫秒。同时，该方法仍保持了多步扩散过程所具有的全局结构一致性与空间连贯性。这一创新显著提升了模型在自动驾驶与机器人导航等实时应用场景中的可用性。

(2) 针对单步推理中容易出现的局部细节缺失问题，我们引入了一个预训练的、数据驱动的单目深度估计模型作为辅助特征提取模块。通过将其编码特征与扩散模型的 U-Net 表示进行融合，模型能够同时利用来自 Stable Diffusion 的大规模文本-图像先验知识，以及数据驱动模型从真实深度标注数据中学习到的非线性纹理-深度与语义相关性。这种双先验融合机制显著丰富了扩散模型潜空间的特征表达能力，从而在保持整体一致性的同时增强了细节恢复与语义结构感知能力。

(3) 我们提出了一种两阶段的训练策略，在一阶段使用 MSE 损失函数，指导模型学习全局结构，在二阶段使用更加注重边缘细节的损失函数，例如利用傅里叶快速变换在频域内学习高频(边缘)信息。通过这样的训练策略，可以让模型在一次的推理过程中有效注重全局和边缘细节。

2. 预期结论：

本文提出的深度估计方法通过单步推理，有效解决了扩散模型方法在深度估计领域应用时推理时间过长的问题，引入的预训练模型辅助以及梯度损失函数，有效地解决了单步推理带来的细节缺失问题。使得模型在零样本性能和推理效率上实现了显著提升。

六、论文工作计划

第一阶段 (2025.09-2025.11)：准备阶段。查阅国内外相关文献，分析研究运用扩散模型的深度估计方法的具体步骤和深度学习框架。

第二阶段 (2025.12-2026.03)：设计阶段。明确研究思路，设计程序调试程序。

第三阶段 (2026.05-2026.06)：实现阶段。实现在第二阶段所设计的算法，并完成所设计的实验，记录实验数据，撰写小论文。

第四阶段 (2026.07-2026.09)：设计阶段。在第二阶段所设计的算法的基础上，完成实验，记录实验数据。

第五阶段 (2026.10-2027.01)：修正完善阶段。对存在的问题进行分析与修正，对所设计的算法及实验进行进一步的完善和改进，撰写毕业论文初稿。

第六阶段 (2027.02-2027.06)：总结阶段。系统的总结之前的工作，完善毕业论文，准备毕业答辩。

参考文献

- [1] Ashutosh Agarwal and Chetan Arora. Attention attention everywhere: Monocular depth prediction with skip attention. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 5861–5870, 2023.
- [2] Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. Adabins: Depth estimation using adaptive bins. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 4009–4018, 2021.
- [3] Yohann Cabon, Naila Murray, and Martin Humenberger. Virtual kitti 2. arXiv:2001.10773, 2020. arXiv preprint
- [4] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5828–5839, 2017.
- [5] Yiquan Duan, Xianda Guo, and Zheng Zhu. Diffusiondepth: Diffusion denoising approach for monocular depth estimation. In European Conference on Computer Vision, pages 432–449. Springer, 2024.
- [6] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 10786–10796, 2021.
- [7] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. Advances in neural information processing systems, 27, 2014.
- [8] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, and Dacheng Tao. Deep ordinal regression network for monocular depth estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2002–2011, 2018.
- [9] Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. Geowizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In European Conference on Computer Vision, pages 241–258. Springer, 2025.
- [10] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE transactions on pattern analysis and machine intelligence, 44:1623–1637, 2020. Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. The international journal of robotics research, 32(11):1231–1237, 2013
- [11] Ming Gui, Johannes Schusterbauer, Ulrich Prestel, Pingchuan Ma, Dmytro Kotovenko, Olga Grebenkova, Stefan Andreas Baumann, Vincent Tao Hu, and Björn Ommer. Depthfm: Fast monocular depth estimation with flow matching, 2024.
- [12] Jing He, Haodong Li, Wei Yin, Yixun Liang, Leheng Li, Kaiqiang Zhou, Hongbo Zhang, Bingbing Liu, and Ying-Cong Chen. Lotus: Diffusion-based visual foundation model for high-quality dense prediction. arXiv preprint arXiv:2409.18124, 2024.
- [13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. Advances in neural information processing systems, 33:6840–6851, 2020.
- [14] Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua

- Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [15] Yuanfeng Ji, Zhe Chen, Enze Xie, Lanqing Hong, Xihui Liu, Zhaoqiang Liu, Tong Lu, Zhenguo Li, and Ping Luo. Ddp: Diffusion model for dense visual prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 21741–21752, 2023.
- [16] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9492–9502, 2024.
- [17] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. From big to small: Multi-scale local planar guidance for monocular depth estimation. arXiv preprint arXiv:1907.10326, 2019.
- [18] Xuecheng Li, Renze Deng, Yulin Fan, Peng Chen, Siyuan Chen, Zhangjun Liu, Li Han, Shuai Zhu, Hao Sun, Yiyi Lu, and Qizhou Li. Depth anything v2. arXiv preprint arXiv:2404.14442, 2024.
- [19] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. arXiv preprint arXiv:2210.02747, 2022.
- [20] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In Proceedings of the IEEE/CVF international conference on computer vision, pages 12179–12188, 2021.
- [21] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2:
- [22] Simon Shi, Luc Van Gool, and Jonathon Luiten. Marigold: Repurposing diffusion models for monocular depth estimation. arXiv preprint arXiv:2404.09015, 2024.
- [23] Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [24] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9492–9502, 2024.
- [25] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. From big to small: Multi-scale local planar guidance for monocular depth estimation. arXiv preprint arXiv:1907.10326, 2019.
- [26] Ziyang Song, Zerong Wang, Bo Li, Hao Zhang, Ruijie Zhu, Li Liu, Peng-Tao Jiang, and Tianzhu Zhang. Depthmaster: Taming diffusion models for monocular depth estimation. arXiv preprint arXiv:2501.02576, 2025.
- [27] Guangkai Xu, Yongtao Ge, Mingyu Liu, Chengxiang Fan, Kangyang Xie, Zhiyue Zhao, Hao Chen, and Chunhua Shen. What matters when repurposing diffusion models for general dense perception tasks? arXiv preprint arXiv:2403.06090, 2024.
- [28] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, “Depth anything: Unleashing the power of large-scale unlabeled data,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 10371–10381.
- [29] A. Vahdat, K. Kreis, and J. Kautz, “Score-based generative modeling in latent space,” *Advances in neural information processing systems*, vol. 34, pp. 11287–11302, 2021.

- [30] Wenliang Zhao, Yongming Rao, Zuyan Liu, Benlin Liu, Jie Zhou, and Jiwen Lu. Unleashing text-to-image diffusion models for visual perception. In ICCV, 2023. 3

本人签字：

年 月 日

导师审核意见及建议：

导师签字：

年 月 日

开题报告专家组提出的主要问题及建议

记录人签字：

年 月 日

注：本表可另附页

开题报告专家组评价意见						
报告日期	2025 年 10 月 31 日		报告地点	丙 1-310		
答辩起始时间	时 分 — 时 分		开题成绩			
专家组评价意见：						
综合结论：						
是否通过开题报告（划√）：			<input type="checkbox"/> 通过		<input type="checkbox"/> 不通过	
组长签字：						
年 月 日						
专家组成员	姓名	职称	签字	姓名	职称	签字
	王永会	教授		李鹏	教授	
	曹科研	教授		韩子扬	副教授	
	袁帅	教授				
	马龙	教授				