阶段一：基础概念（Foundation Phase）

1. MiDaS - Towards Robust Monocular Depth Estimation

- ArXiv 论文: https://arxiv.org/abs/1907.01341
- GitHub 代码: https://github.com/isl-org/MiDaS
- PyTorch Hub: https://pytorch.org/hub/intelisl_midas_v2/

这篇论文是理解传统深度估计和尺度歧义问题的基石。它展示了如何通过混合数据集来提升零样本泛化能力，这为后续的深度基础模型铺平了道路。

2. Vision Transformer (ViT) - An Image is Worth 16x16 Words

- ArXiv 论文: https://arxiv.org/abs/2010.11929
- Google Research: https://research.google/pubs/an-image-is-worth-16x16-words-transformers-for-image-recognition-at-scale/
- GitHub 实现: https://github.com/jeonsworld/ViT-pytorch

这篇开创性论文将 Transformer 引入计算机视觉，为后续的 DPT 和深度基础模型奠定了理论基础。理解 ViT 的 patch embedding 和自注意力机制对理解后续架构至关重要。

3. DPT - Vision Transformers for Dense Prediction

- ArXiv 论文: https://arxiv.org/abs/2103.13413
- GitHub 代码: https://github.com/isl-org/DPT
- CVF 会议版本: https://openaccess.thecvf.com/content/ICCV2021/papers/Ranftl_Vision_Transformers_for_Dense_Prediction_ICCV_2021_paper.pdf
- Hugging Face 模型: https://huggingface.co/docs/transformers/en/model_doc/dpt

DPT 是 Prompt Depth Anything 直接基于的架构。这篇论文展示了如何将 ViT 适配到像深度估计这样的密集预测任务，理解其多尺度特征重组和解码过程是核心。

阶段二：深度基础模型（Foundation Models）

4. Depth Anything V1 - Unleashing the Power of Large-Scale Unlabeled Data

- ArXiv 论文: https://arxiv.org/abs/2401.10891
- 项目主页: https://depth-anything.github.io/
- GitHub 代码: https://github.com/LiheYoung/Depth-Anything

这是 Prompt Depth Anything 直接扩展的基础模型。深入理解其数据扩展策略、辅助监督机制，以及为什么它能产生优秀的相对深度但在度量深度上存在局限性。

阶段三：核心论文（Core Paper）

5. Prompt Depth Anything - Prompting Depth Anything for 4K Resolution Accurate Metric Depth Estimation

- 这是你想要理解的核心论文（已在上传的文档中）
- 项目主页: https://PromptDA.github.io/

在理解了前面的基础之后，你现在可以深入研究这篇论文的创新点：提示融合架构、边缘感知损失、以及如何通过 LiDAR 提示来解决度量深度问题。

阶段四：相关技术（Related Technologies）

6. ControlNet - Adding Conditional Control to Text-to-Image Diffusion Models 需要搜索这篇论文来理解视觉模型中的条件控制机制

7. Zip-NeRF 用于生成伪真值深度的重建方法

8. ARKitScenes 和 ScanNet++ 数据集论文 理解真实世界 iPhone LiDAR 数据的特性

Let me search for a few more essential papers to complete your reading list:现在让我为你整理完整的阅读计划，包括所有关键论文的链接：

完整的论文阅读路径（Complete Reading Path）

阶段一：基础概念建立（Foundation Phase）

1. MiDaS: 传统深度估计的里程碑

- ArXiv 论文：https://arxiv.org/abs/1907.01341
- GitHub 代码：https://github.com/isl-org/MiDaS
- PyTorch Hub：https://pytorch.org/hub/intelisl_midas_v2/

这篇论文是理解深度估计基础的关键起点。它向你展示了尺度歧义性问题的本质——为什么传统方法能产生相对深度但难以获得绝对度量深度。理解这个问题是把握Prompt Depth Anything创新价值的前提。

2. Vision Transformer：视觉革命的开端

- ArXiv 论文：https://arxiv.org/abs/2010.11929
- Google Research：https://research.google/pubs/an-image-is-worth-16x16-words-transformers-for-image-recognition-at-scale/
- GitHub 实现：https://github.com/jeonsworld/ViT-pytorch

这篇开创性论文将Transformer从自然语言处理引入计算机视觉。你需要深入理解patch embedding的概念——图像如何被切分成小块并转换为序列，以及自注意力机制如何在视觉任务中发挥作用。这些概念直接影响后续所有基于Transformer的深度模型。

3. DPT：密集预测的Transformer架构

- ArXiv 论文：https://arxiv.org/abs/2103.13413
- GitHub 代码：https://github.com/isl-org/DPT
- CVF 会议版本：https://openaccess.thecvf.com/content/ICCV2021/papers/Ranftl_Vision_Transformers_for_Dense_Prediction_ICCV_2021_paper.pdf
- Hugging Face 文档：https://huggingface.co/docs/transformers/en/model_doc/dpt

DPT是Prompt Depth Anything的直接架构基础。你必须理解它如何解决ViT在密集预测任务中的挑战——如何将不同阶段的token重新组装成多分辨率的图像表示，以及如何通过卷积解码器逐步融合这些表示来生成最终的深度图。

阶段二：深度基础模型理解（Foundation Models）

4. Depth Anything V1：大规模无标签数据的力量

- ArXiv 论文：https://arxiv.org/abs/2401.10891
- 项目主页：https://depth-anything.github.io/
- GitHub 代码：https://github.com/LiheYoung/Depth-Anything

这是Prompt Depth Anything的直接前身。你需要深入理解两个关键策略：如何通过数据增强创造更具挑战性的优化目标，以及如何通过辅助监督让模型继承预训练编码器的丰富语义先验。更重要的是，理解为什么这个模型在相对深度上表现优秀，但在度量深度估计上仍有局限性。

5. Depth Anything V2：更强大的基础模型

- GitHub 代码（主要）：https://github.com/DepthAnything/Depth-Anything-V2
- HTML 版本：https://arxiv.org/html/2406.09414v1

这个版本在前者基础上进一步改进了细节处理和鲁棒性。理解它如何在合成数据和真实数据之间建立桥梁，以及如何处理透明物体和反射表面这些传统方法的弱点。

阶段三：条件控制理解（Conditional Control）

6. ControlNet：条件控制的典范

- ArXiv 论文：https://arxiv.org/abs/2302.05543
- GitHub 代码：https://github.com/lllyasviel/ControlNet
- CVF 会议版本：https://openaccess.thecvf.com/content/ICCV2023/papers/Zhang_Adding_Conditional_Control_to_Text-to-Image_Diffusion_Models_ICCV_2023_paper.pdf
- Hugging Face 页面：https://huggingface.co/papers/2302.05543

虽然ControlNet用于扩散模型，但它的核心思想——使用额外输入信号来引导预训练基础模型——与Prompt Depth Anything的理念高度一致。特别要理解"零卷积"（zero convolution）的概念和为什么它能确保不损害原始模型的能力。

阶段四：核心论文深入（Core Paper Study）

7. Prompt Depth Anything

- 这是你的目标论文（已在上传文档中）
- 项目主页：https://PromptDA.github.io/

现在你已经具备了理解这篇论文的所有基础知识。重点关注：

- 提示融合架构如何在多尺度上集成LiDAR信息
- 边缘感知损失如何结合伪真值深度和FARO标注深度的优势
- 为什么这种简单的架构扩展就能解决度量深度的难题

学习策略建议（Learning Strategy）

在这个学习过程中，我建议你采用渐进式理解的方法。不要试图一次性掌握所有细节，而是要在每个阶段建立扎实的概念基础。

首先，当你阅读MiDaS时，专注于理解什么是尺度歧义性，以及为什么仅仅从单张图像很难获得绝对的度量信息。然后在学习ViT时，重点把握patch embedding和自注意力的核心思想，不必过分纠结于具体的数学公式。

在学习DPT时，最重要的是理解它如何将全局感受野的优势与多分辨率处理相结合。这个架构思想直接影响了后续所有工作。学习Depth Anything时，要深入思考大规模数据训练的重要性，以及为什么无标签数据能显著提升模型的泛化能力。

学习ControlNet时，重点不在于理解扩散模型的细节，而在于把握如何在不损害原始模型能力的前提下添加条件控制。零初始化的思想在Prompt Depth Anything中同样被采用。

最后，当你研读Prompt Depth Anything时，你会发现它的创新既简洁又深刻——通过在已有架构中巧妙地融入LiDAR提示，就解决了长期困扰该领域的度量深度问题。这种"四两拨千斤"的设计思路本身就值得深入体会。

这样的学习路径会让你不仅理解论文的技术细节，更能理解它在整个深度估计领域发展历程中的地位和意义。每一篇论文都是理解下一篇论文的必要基础，形成了一个完整的知识体系。