🚀 VLA 领域学习路线图：从入门到跟上 2026 前沿

> 🗓️ 生成日期：2026年1月
> 🎯 目标读者：大二计算机本科生，已完成 SayCan / Inner Monologue / VIMA / VLA综述 的阅读
> 📌 课题方向：面向工业机械臂的模糊指令语义消歧

📚 Part 1: 需要补齐的核心知识点（少而精）

1.1 🔥 Diffusion Models（扩散模型）—— 优先级：⭐⭐⭐⭐⭐

为什么重要？

- 当前最先进的 VLA 模型（如 π0、OpenVLA）都使用 Diffusion 或 Flow Matching 来生成机器人动作
- 你的课题涉及"模糊指令"，Diffusion 天然擅长建模**多模态分布**（一句话可能对应多种合理动作）

需要掌握的核心概念：

概念	中文翻译	一句话解释
Denoising	去噪	从纯噪声一步步"擦干净"变成清晰图像/动作
Score Function	得分函数	告诉模型"往哪个方向去噪"的指南针
Conditional Diffusion	条件扩散	生成时加上条件（如语言指令），让输出符合要求
Classifier-Free Guidance	无分类器引导	一种让条件生成更"听话"的技巧

推荐学习资源：

1. 🎬 **视频入门**：Lil'Log 的 Diffusion 博客（有中文翻译）
2. 📖 **代码实践**：跑一遍 `diffusers` 库的 DDPM 教程（用 PyTorch，你熟悉）
3. ⏱️ 预计学习时间：3-5 天

1.2 🎯 Flow Matching（流匹配）—— 优先级：⭐⭐⭐⭐

为什么重要？

- π0（Physical Intelligence 的王牌模型）使用 Flow Matching 而非传统 Diffusion
- 比 Diffusion 训练更稳定、推理更快（50Hz 实时控制！）

核心思想（一句话）：

> Diffusion 是"随机游走去噪"，Flow Matching 是"走直线最短路径"

需要掌握的概念：

概念	中文翻译	对比 Diffusion
Optimal Transport	最优传输	Diffusion 是随机的，OT 是最优路径
Continuous Normalizing Flow	连续归一化流	用 ODE 而非 SDE 建模
Rectified Flow	矫正流	Flow Matching 的一种具体实现

推荐学习资源：

1. 📄 Flow Matching 原始论文（Lipman et al., 2022）- 只需看 Section 1-3
2. 🧪 π0 的 HuggingFace 博客有很好的图解
3. ⏱️ 预计学习时间：2-3 天（在理解 Diffusion 之后）

1.3 🤖 强化学习基础（聚焦 VLA 相关）—— 优先级：⭐⭐⭐⭐

为什么重要？

- 2025-2026 的趋势：**RL for VLA**，用强化学习微调 VLA 模型
- 你的课题可能需要：当机械臂犯错时，如何通过反馈自我纠正？

你只需掌握这几个算法（不需要全面学 RL）：

算法	中文名	用途	学习优先级
PPO	近端策略优化	VLA 在线微调的标准算法	⭐⭐⭐⭐⭐
DPO	直接偏好优化	不需要 Reward Model，更简单	⭐⭐⭐⭐
RLHF	人类反馈强化学习	概念性理解即可	⭐⭐⭐
GRPO	群体相对策略优化	DeepSeek 提出，2025 年新趋势	⭐⭐⭐

核心公式只需理解这一个（PPO 的 Clipped Objective）：

$$L^{CLIP}(\theta) = \mathbb{E}_t \left[\min \left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t \right) \right]$$

中文解读：

- $r_t(\theta)$：新策略比旧策略"好多少"的比值

- $\hat{A}_t$：优势函数（这个动作比平均水平好多少）
- clip：限制更新幅度，防止一步走太远

推荐学习资源：

1. 📺 李宏毅老师的 PPO 讲解（B站有）
2. 📄 OpenAI Spinning Up 的 PPO 文档
3. ⏱️ 预计学习时间：3-4 天

1.4 🧩 Action Tokenization（动作分词）—— 优先级：⭐⭐⭐

为什么重要？

- VLA 的核心问题：如何把连续的机器人动作（关节角度、末端位置）转成 LLM 能处理的 token？
- 2025 年热门话题：FAST Tokenizer（π0 使用）

两种主流方案对比：

方案	代表模型	优点	缺点
离散化 Binning	RT-2, OpenVLA	简单，兼容 LLM	精度损失，高频控制困难
Flow/Diffusion Head	π0	连续输出，高精度	需要额外的 Action Head

FAST Tokenizer 核心思想：

> 用 DCT（离散余弦变换）把时间序列动作压缩到频域，减少冗余

推荐学习资源：

1. 📄 π0-FAST 论文的 Section 3
2. 🔗 HuggingFace LeRobot 的 π0 教程
3. ⏱️ 预计学习时间：1-2 天

1.5 📦 Imitation Learning / Behavior Cloning（模仿学习）—— 优先级：⭐⭐⭐

为什么重要？

- 这是所有 VLA 的训练基础（SayCan、VIMA 你已经看过了）
- 理解 BC 的局限性，才能理解为什么需要 RL

核心概念：

概念	中文	解释
Behavior Cloning	行为克隆	直接模仿专家动作，监督学习
Distribution Shift	分布偏移	模型犯错后进入"没见过"的状态
Compounding Error	累积误差	小错误滚雪球变成大错误
DAgger	数据聚合	让专家来纠正模型的错误

你应该已经懂了： SayCan 的 affordance scoring 其实就是在缓解 distribution shift！

📄 Part 2: 必读论文清单（少而精，共 5 篇）

📌 论文阅读优先级排序

#	论文	年份	为什么必读	预计时间
1	**Diffusion Policy**	2023	Diffusion 用于机器人的开山之作	2天
2	**π0**	2024	当前最强 VLA，你的对标模型	2天
3	**OpenVLA**	2024	开源可跑，学习代码架构	2天
4	**RT-2**	2023	VLA 概念的起源，理解历史脉络	1天
5	**VLA-RL**	2025	RL 微调 VLA 的最新范式	1天

📖 论文 1：Diffusion Policy (Chi et al., RSS 2023)

论文全名： *Diffusion Policy: Visuomotor Policy Learning via Action Diffusion*

一句话概括： 把扩散模型用于机器人，生成一段连续动作序列而非单步动作

必看部分（按优先级）：

- ✅ Figure 1, 2：理解整体流程
- ✅ Section 3.1：Diffusion 如何条件化在视觉观测上
- ✅ Section 4：实验结果（尤其是和 BC 的对比）
- ⏭️ 可跳过：数学推导细节（Appendix）

和你课题的联系：

> 模糊指令 → 多种合理动作 → Diffusion 的多模态建模能力正好适用！

资源链接：

- 📄 论文：https://arxiv.org/abs/2303.04137
- 🌐 项目页：https://diffusion-policy.cs.columbia.edu/
- 💻 代码：GitHub 有完整实现

📖 论文 2：π0 (Physical Intelligence, 2024)

论文全名： *π₀: A Vision-Language-Action Flow Model for General Robot Control*

一句话概括： 用 Flow Matching + 预训练 VLM，训练出能控制多种机器人的通用策略

必看部分（按优先级）：

- ✅ Figure 1, 3：模型架构图（必须看懂！）
- ✅ Section 3：Flow Matching Action Head 的设计
- ✅ Section 4.1：训练数据来源（Open X-Embodiment）
- ⏭️ 可跳过：具体的超参数调优细节

关键架构理解：

```
输入: 图像 + 语言指令 + 机器人状态
  ↓
预训练 VLM (PaliGemma) → 语义理解
  ↓
Flow Matching Head → 连续动作序列 (50Hz)
  ↓
输出: 关节角度 / 末端位置
```

和你课题的联系：

> π0 展示了如何把"语言理解"和"动作生成"端到端结合——你的"模糊指令消歧"可以插入到这个流程中

资源链接：

- 📄 论文：https://arxiv.org/abs/2410.24164
- 💻 开源代码：https://github.com/Physical-Intelligence/openpi
- 🤗 HuggingFace：有预训练权重可下载

📖 论文 3：OpenVLA (Stanford, 2024)

论文全名： *OpenVLA: An Open-Source Vision-Language-Action Model*

一句话概括： 7B 参数的开源 VLA，可以在消费级 GPU 上微调

必看部分（按优先级）：

- ✅ Figure 2：模型架构（DINOv2 + SigLIP + Llama）
- ✅ Section 3：动作离散化的具体做法
- ✅ Section 5：微调实验（你可以复现！）
- ⏭️ 可跳过：所有 baseline 的详细对比

为什么推荐这篇：

1. **开源可跑**：你可以在自己电脑上跑 demo
2. **架构清晰**：理解 VLA 的标准设计模式
3. **社区活跃**：GitHub issues 里有很多调试经验

资源链接：

- 📄 论文：https://arxiv.org/abs/2406.09246
- 💻 代码：https://github.com/openvla/openvla
- 🤗 模型：HuggingFace 上有 7B 权重

📖 **论文 4：RT-2 (Google DeepMind, 2023)**

论文全名： *RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control*

一句话概括： 第一个证明"互联网知识可以迁移到机器人"的工作，VLA 这个词的起源

必看部分（按优先级）：

- ✅ Figure 1：核心 idea 一图流
- ✅ Section 2：如何把动作表示成文本 token
- ✅ Section 4.3：Emergent Capabilities（涌现能力）
- ⏭️ 可跳过：PaLI-X 的具体训练细节

历史意义：

> RT-2 证明了：预训练 VLM 的"常识"（比如"苹果是红色的"）可以帮助机器人理解指令！

和你课题的联系：

> "把那个红色的东西拿过来"——RT-2 展示了 VLM 如何利用颜色常识消歧

📖 论文 5：VLA-RL (Tsinghua & NTU, 2025)

论文全名： *VLA-RL: Towards Masterful and General Robotic Manipulation with Scalable Reinforcement Learning*

一句话概括： 用在线 RL 微调预训练 VLA，让模型能从错误中学习

必看部分（按优先级）：

- ✅ Figure 2：RL 训练框架图
- ✅ Section 3.2：Trajectory-level RL 的形式化
- ✅ Section 3.3：Process Reward Model（过程奖励模型）
- ⏭️ 可跳过：GPU 调度优化的工程细节

为什么推荐这篇（2025 最新！）：

- 代表了 VLA 领域的最新趋势：**从模仿学习到强化学习**
- 展示了如何解决"分布偏移"问题
- 在 LIBERO 上超过了 π0-FAST！

资源链接：

- 📄 论文：https://arxiv.org/abs/2505.18719
- 💻 代码：https://github.com/GuanxingLu/vlarl

🛠️ Part 3: 实践建议

3.1 推荐的 Demo 跑通顺序

```
Week 1: 跑通 Diffusion Policy 的仿真 demo
    ↓
Week 2: 跑通 OpenVLA 的推理（用 HuggingFace）
    ↓
Week 3: 尝试在 LIBERO 上微调 OpenVLA
    ↓
Week 4: 阅读 π0 的开源代码，理解架构
```

3.2 推荐的仿真环境

环境	特点	适合练习
LIBERO	40个任务，VLA研究标配	VLA 微调和评估
SimplerEnv	真实感强，Google 出品	Sim-to-Real 实验

环境	特点	适合练习
RoboCasa	家庭场景，长时任务	长程任务规划
MetaWorld	经典，任务多	快速验证想法

3.3 硬件要求参考

任务	最低配置	推荐配置
跑 OpenVLA 推理	RTX 3090 (24GB)	A100 (40GB)
微调 OpenVLA	A100 (40GB)	4× A100
跑 π0 推理	A100 (40GB)	A100 (80GB)

🗓️ Part 4: 建议学习时间表（4周计划）

周次	知识点	论文	实践
Week 1	Diffusion Models 基础	Diffusion Policy	跑通 DP 仿真 demo
Week 2	Flow Matching + Action Tokenization	π0	阅读 π0 代码
Week 3	PPO + DPO 基础	OpenVLA, RT-2	跑通 OpenVLA 推理
Week 4	RL for VLA	VLA-RL	尝试微调实验

🔗 Part 5: 实用资源汇总

代码仓库

- 📦 **Awesome VLA 列表**：https://github.com/jonyzhang2023/awesome-embodied-vla-va-vln
- 📦 **Diffusion for Robotics 文献**：https://github.com/mbreuss/diffusion-literature-for-robotics
- 📦 **LeRobot (HuggingFace)**：https://github.com/huggingface/lerobot

博客 & 教程

- 📝 **State of VLA at ICLR 2026**：最新趋势总结
- 📝 **HuggingFace π0 博客**：架构图解非常清晰
- 📝 **Lil'Log Diffusion 教程**：入门必读

学术追踪

- 🔔 **arXiv cs.RO**：每天刷一刷
- 🔔 **Twitter/X @EmbodiedAIRead**：精选论文推送
- 🔔 **CoRL / RSS / ICRA**：顶会关注

💡 Part 6: 和你课题的结合点

你的课题是**面向工业机械臂的模糊指令语义消歧**，以下是知识点的直接应用：

知识点	你的课题应用
Diffusion Policy	模糊指令对应多种动作分布，用 Diffusion 建模
Flow Matching	工业场景需要实时控制（≥50Hz），Flow 比 Diffusion 快
PPO / RL	机械臂犯错后，如何通过反馈自我纠正？
Action Tokenization	工业机械臂通常 6-7 DoF，如何高效编码？
Affordance (SayCan)	你已经学过！判断"这个动作能不能做"

一个可能的研究方向：

> 结合 SayCan 的 affordance scoring + Diffusion Policy 的多模态动作生成
> → 当指令模糊时，先用 affordance 过滤不可行动作，再用 diffusion 从剩余动作中采样

✅ 检查清单

- [ ] 能用自己的话解释 Diffusion 的去噪过程
- [ ] 能说出 Flow Matching 和 Diffusion 的区别
- [ ] 能写出 PPO 的 clipped objective 并解释每一项
- [ ] 能跑通 Diffusion Policy 的 demo
- [ ] 能跑通 OpenVLA 的推理
- [ ] 能画出 $\pi 0$ 的架构图

> 📌 **记住**：知识点和论文都是为你的课题服务的。学的时候时刻问自己："这个东西怎么用到模糊指令消歧上？"

祝学习顺利！有任何问题随时问我 🎓