DeepSeek Janus-Pro 深度解读

ZOMI

DeepSeek除夕狂飙大招：开源多模态掀翻全场！256张A100训两周碾压DALL-E 3

新智元 新智元 2025年01月28日 13:14 陕西

新智元报道

编辑：Aeneas 好困

DeepSeek深夜炸场！开源视觉模型，一统图像生成与视觉理解

原创 含萧 夕小瑶科技说 2025年01月28日 11:45 北京

视频目录大纲

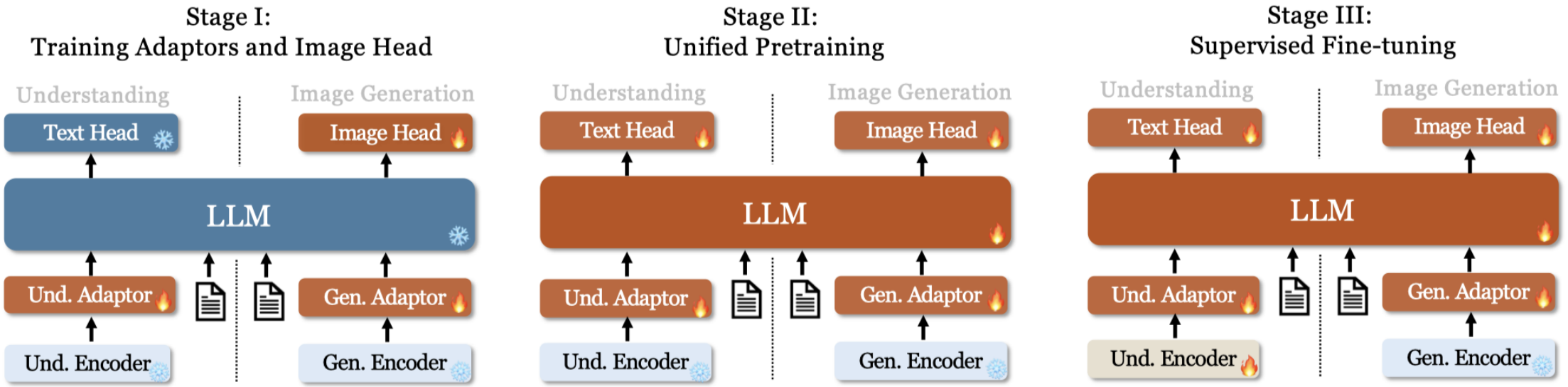
1. DeepSeek Janus-Pro 技术文章解读
2. 从 DeepSeek Janus 到 JanusFlow 终极 Janus-Pro
3. DeepSeek Janus-Pro 实测效果
4. 对产业的思考与小结

01

Janus-Pro
技术文章解读

Janus 的训练分为三个阶段

- **第一阶段：**使用 Image Caption 和 ImageNet 文生图数据，对 understanding adaptor, generation adaptor, image head 三个模块进行训练，起到 warm up 的效果。
- **第二阶段：**打开 LLM 和 text head，使用大量纯文本、图生文和文生图数据进行预训练。对文生图数据，让 ImageNet 出现在文生图数据之前，先学习像素依赖关系，然后学习场景生成 。
- **第三阶段：**打开 understanding encoder，使用指令跟随数据进行有监督的微调。



02

Janus、JanusFlow、Janus-Pro

DeepSeek Janus、JanusFlow、Janus-Pro

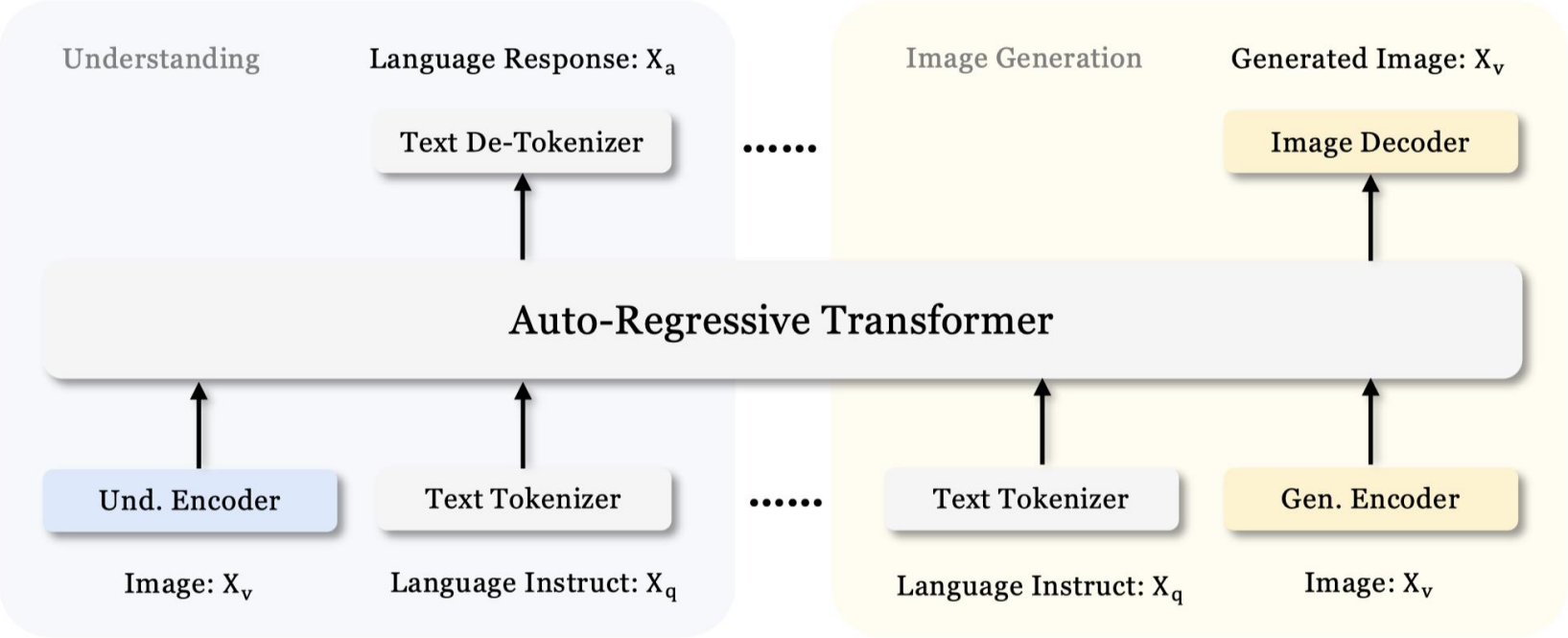
- DeepSeek VLM 系列发布路线：

Janus 1.3b → Janus Flow 1.3b → Janus-pro-7b

2024.10 **2024.11** **2025.01.28**

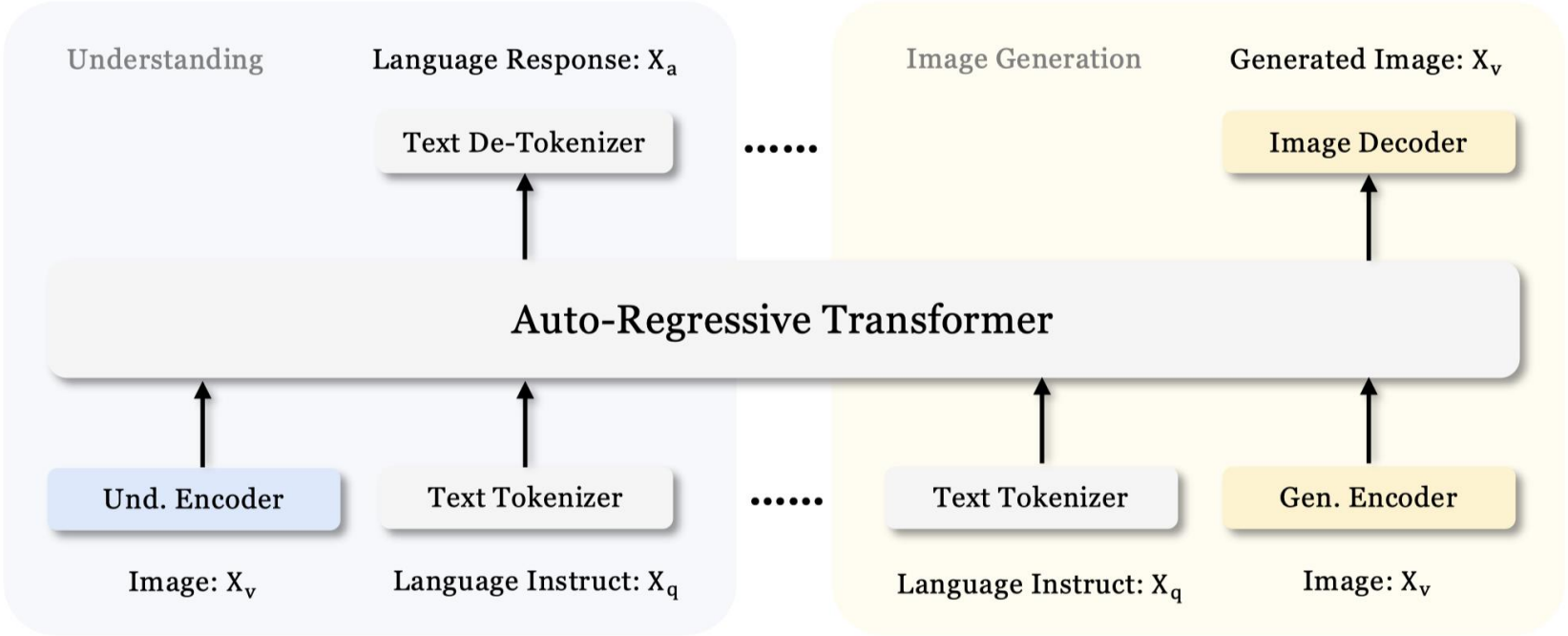
DeepSeek Janus

- 基于自回归的多模态理解与生成统一模型。Janus 的核心思想是对理解和生成任务的视觉编码进行解耦，在提升了模型的灵活性的同时，有效缓解了使用单一视觉编码导致的冲突和性能瓶颈。



DeepSeek Janus

- 为简化整个模型，对多模态理解任务来说，使用 SigLIP-Large-Patch16-384 编码图像特征。对视觉生成任务来说，使用 Llama Gen 中标准 VQ Tokenizer 编码图像特征。
- 编码后信息分别经过一个 adaptor，然后送入 LLM 。整个模型是使用 Next-Token-Prediction 方式进行训练，采用 causal attention mask，和 LLM 训练方式一致。



DeepSeek JanusFlow

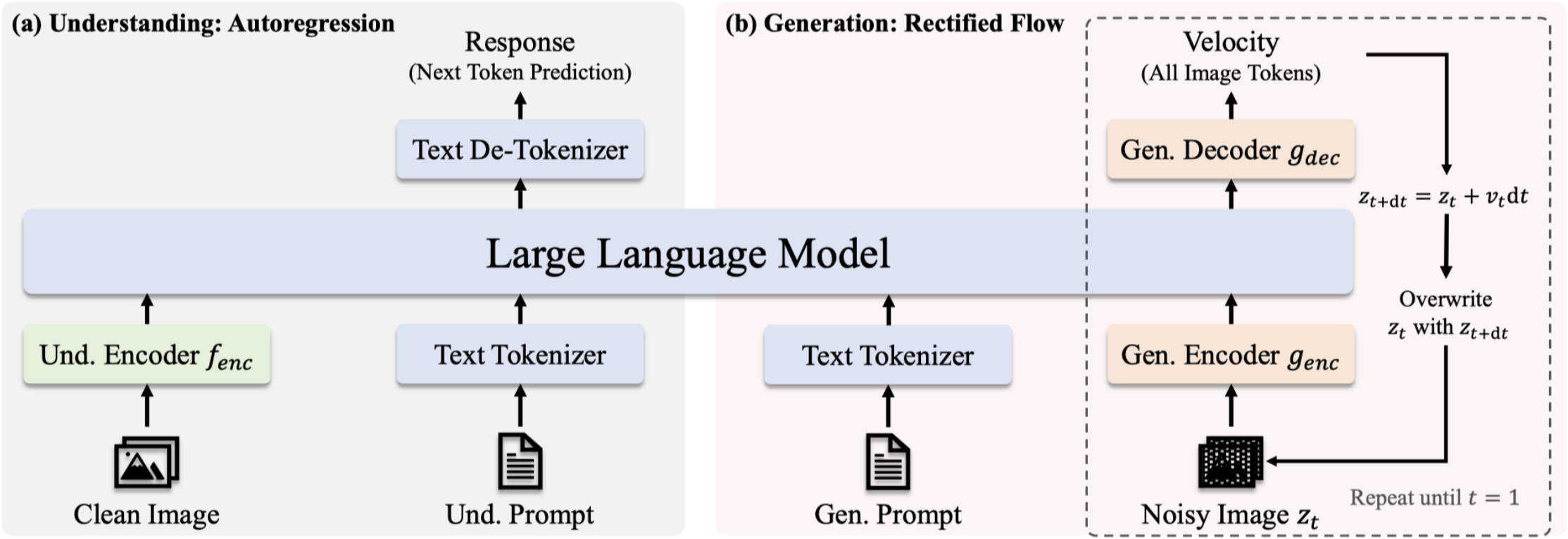
1. 在多模态AI领域，基于预训练视觉编码器与 MLLM （LLaVA 系列）在视觉理解任务上展现出卓越性能。
2. 基于 Rectified Flow 模型（如Stable Diffusion 3及其衍生版本）则在视觉生成方面取得重大突破。

- 能否将这两种简单的技术范式统一到单一模型中？DeepSeek JanusFlow：
- **在LLM框架内直接融合这两种结构，实现视觉理解与生成能力的有效统一**

DeepSeek JanusFlow

- **视觉理解编码器（Und. Encoder）：**使用 SigLIP 将输入图片转换成 Visual embeddings；专注于视觉理解任务的特征提取。
- **视觉生成编解码器（Gen. Encoder/Decoder）：**轻量级模块，总参数量约70M；基于 SDXL-VAE 的 latent space 进行生成；编码器：利用双层 ConvNeXt Block 将输入 latent z_t 转换为 visual embeddings；解码器：通过双层 ConvNeXt Block 将处理后的 embeddings 解码为 latent space 中速度 v 。
- **注意力机制：**初步实验中，发现生成任务中 causal attention 和 bidirectional attention 效果相当；基于效率和简洁性考虑，统一采用 causal attention 处理两类任务。

DeepSeek JanusFlow



Janus-Pro 主要改进

- **训练策略**
  - Stage 1：增加训练步数，在 ImageNet 上充分训练；
  - Stage 2：不再使用 ImageNet，直接使用常规文本到图像数据的训练数据；
  - Stage 3：修改微调过程中数据集配比，将多模态数据、纯文本数据和文本到图像比例从 7:3:10 改为 5:1:4
- **模型规模**
  - 将模型参数扩展到 70 亿参数规模；

Janus-Pro主要改进

- **数据规模**
  - Stage 2：增加 9000 万个样本，包括图像字幕数据 YFCC、表格图表文档理解数据 Doc-matrix；
  - Stage 3：加入 DeepSeek-VL2 额外数据集，如 MEME 理解等；
- **多模态理解**
  - 视觉生成：真实世界数据可能包含质量不高，导致文本到图像的生成不稳定，产生美学效果不佳的输出，Janus-Pro 使用 7200 万份合成美学数据样本，统一预训练阶段（Stage 2）真实数据与合成数据比例 1:1；

03

Janus-Pro
实测效果

Janus-Pro vs flux Pro

- https://huggingface.co/spaces/deepseek-ai/Janus-Pro-7B
- https://www.fluxpro.ai/dashboard

04

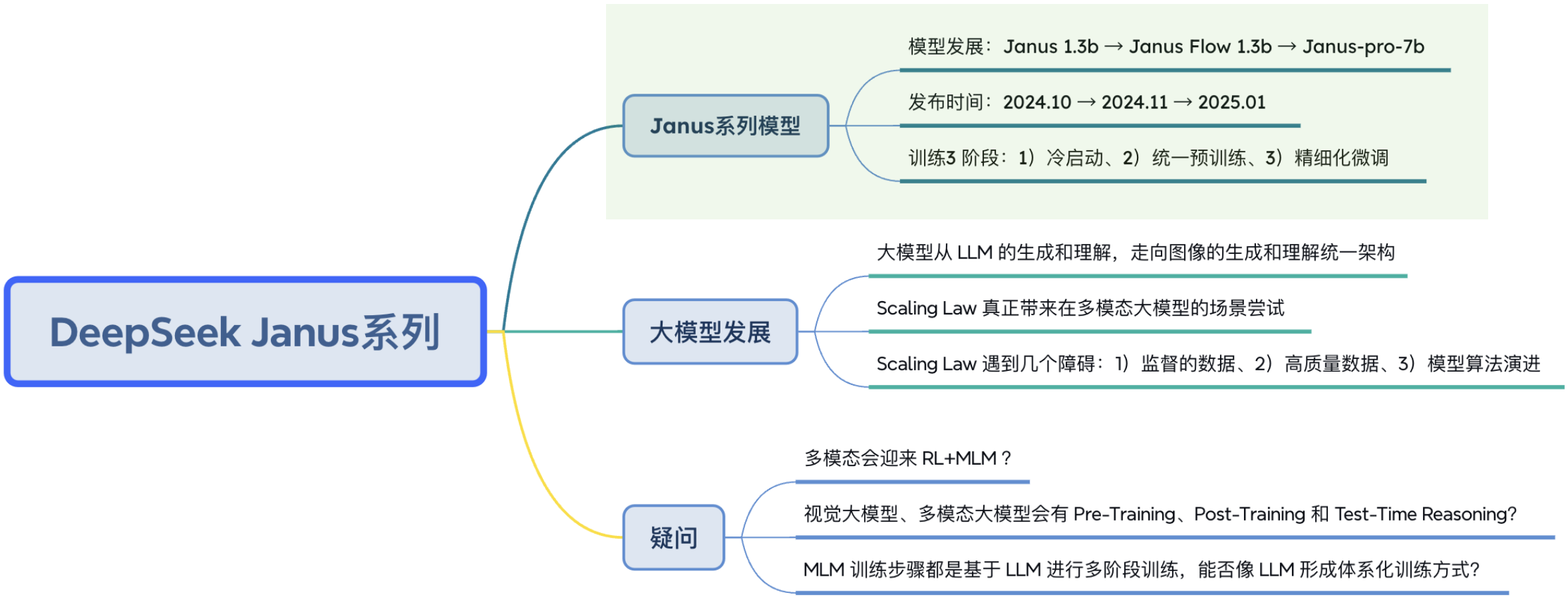
思考与小结

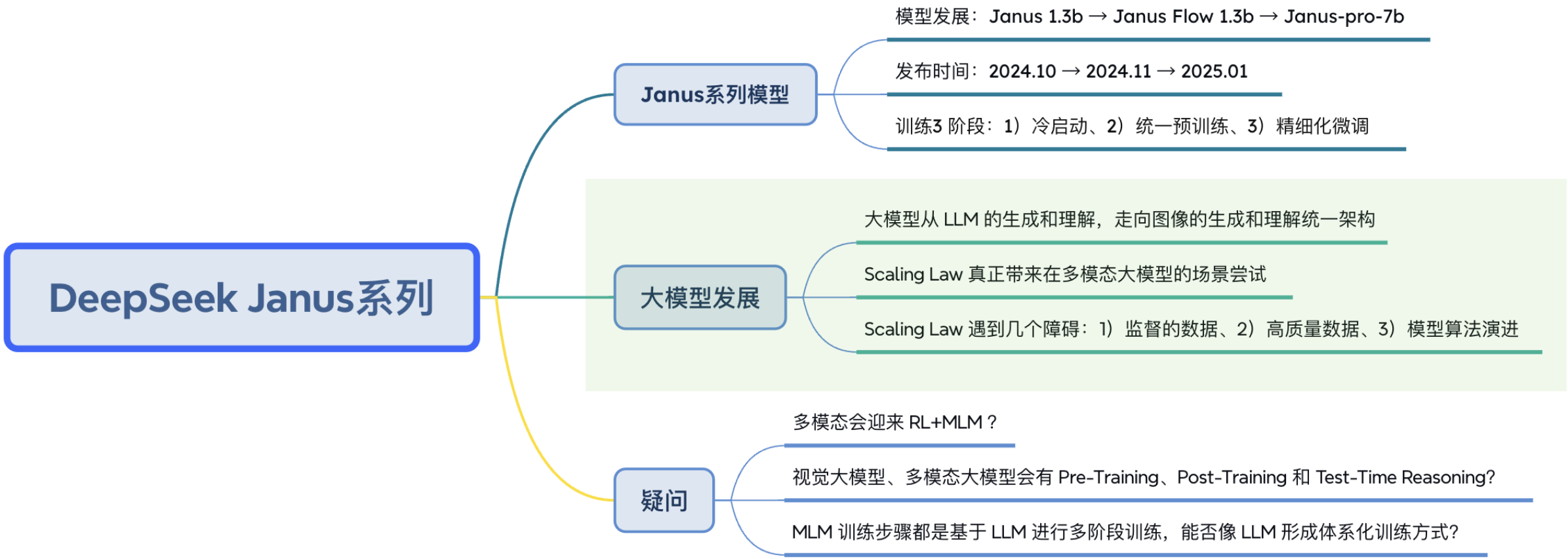
离散和连续的问题

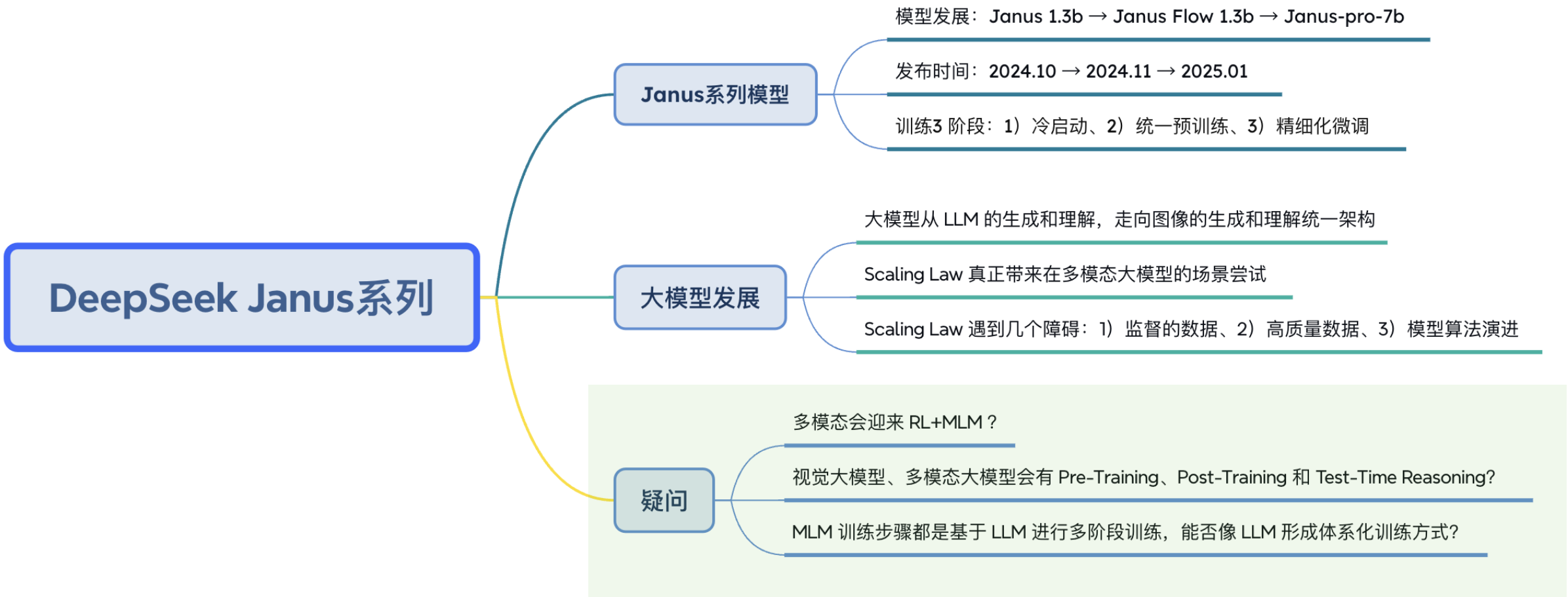
- JanusFlow 采用的是连续的 visual feature，Janus-Pro又回到了离散的 visual feature。
- DeepSeek 内部也不是很确定统一的视觉大模型应该是采用连续还是离散的特征，都在试。
- 连续 feature 上限更高直觉上更符合物理事实，离散 token + casual attention 对硬件更友好。

多模态 Scaling Law

- 现在简单罗列性能已经有点无聊了。当前图像 understanding or generation 都很难达到 SOTA 的能力。
- 更激动人心的应该是研究多模态 scaling law，看看图像理解 / 图像生成 / 文字生成 任务能不能相互促进。


DeepSeek Janus系列
Janus系列模型
模型发展：Janus 1.3b → Janus Flow 1.3b → Janus-pro-7b
发布时间：2024.10 → 2024.11 → 2025.01
训练3 阶段：1）冷启动、2）统一预训练、3）精细化微调
大模型发展
大模型从 LLM 的生成和理解，走向图像的生成和理解统一架构
Scaling Law 真正带来在多模态大模型的场景尝试
Scaling Law 遇到几个障碍：1）监督的数据、2）高质量数据、3）模型算法演进
疑问
多模态会迎来 RL+MLM？
视觉大模型、多模态大模型会有 Pre-Training、Post-Training 和 Test-Time Reasoning?
MLM 训练步骤都是基于 LLM 进行多阶段训练，能否像 LLM 形成体系化训练方式?


DeepSeek Janus系列
Janus系列模型
模型发展：Janus 1.3b → Janus Flow 1.3b → Janus-pro-7b
发布时间：2024.10 → 2024.11 → 2025.01
训练3 阶段：1）冷启动、2）统一预训练、3）精细化微调
大模型发展
大模型从 LLM 的生成和理解，走向图像的生成和理解统一架构
Scaling Law 真正带来在多模态大模型的场景尝试
Scaling Law 遇到几个障碍：1）监督的数据、2）高质量数据、3）模型算法演进
疑问
多模态会迎来 RL+MLM？
视觉大模型、多模态大模型会有 Pre-Training、Post-Training 和 Test-Time Reasoning?
MLM 训练步骤都是基于 LLM 进行多阶段训练，能否像 LLM 形成体系化训练方式?


DeepSeek Janus系列
Janus系列模型
模型发展：Janus 1.3b → Janus Flow 1.3b → Janus-pro-7b
发布时间：2024.10 → 2024.11 → 2025.01
训练3 阶段：1）冷启动、2）统一预训练、3）精细化微调
大模型发展
大模型从 LLM 的生成和理解，走向图像的生成和理解统一架构
Scaling Law 真正带来在多模态大模型的场景尝试
Scaling Law 遇到几个障碍：1）监督的数据、2）高质量数据、3）模型算法演进
疑问
多模态会迎来 RL+MLM？
视觉大模型、多模态大模型会有 Pre-Training、Post-Training 和 Test-Time Reasoning?
MLM 训练步骤都是基于 LLM 进行多阶段训练，能否像 LLM 形成体系化训练方式?

DeepSeek Janus

- 论文: https://arxiv.org/pdf/2410.13848
- 项目主页：https://github.com/deepseek-ai/Janus
- 模型下载：https://huggingface.co/deepseek-ai/Janus-1.3B
- 在线 Demo：https://huggingface.co/spaces/deepseek-ai/Janus-1.3B

DeepSeek JanusFlow

- 论文链接：https://arxiv.org/abs/2411.07975
- 项目主页： https://github.com/deepseek-ai/Janus
- 模型下载： https://huggingface.co/deepseek-ai/JanusFlow-1.3B
- 在线 Demo： https://huggingface.co/spaces/deepseek-ai/JanusFlow-1.3B

DeepSeek Janus-Pro

- 论文链接：https://github.com/deepseek-ai/Janus/janus_pro_tech_report.pdf
- 项目主页： https://github.com/deepseek-ai/JanusJanus
- 模型下载： https://huggingface.co/deepseek-ai/ Janus-Pro-7B
- 在线 Demo： https://huggingface.co/spaces/deepseek-ai/Janus-Pro-7B

Reference & 参考内容

1. https://mp.weixin.qq.com/s/Ao5V0ICGX3X2HWfIw23YAQ
2. https://mp.weixin.qq.com/s/NLk9JUJskvCIy7IiMCQAUw
3. https://mp.weixin.qq.com/s/XuJh7huN2Ndsy2b3Mr1pbQ
4. https://www.zhihu.com/question/10723450802
5. https://zhuanlan.zhihu.com/p/20764953677