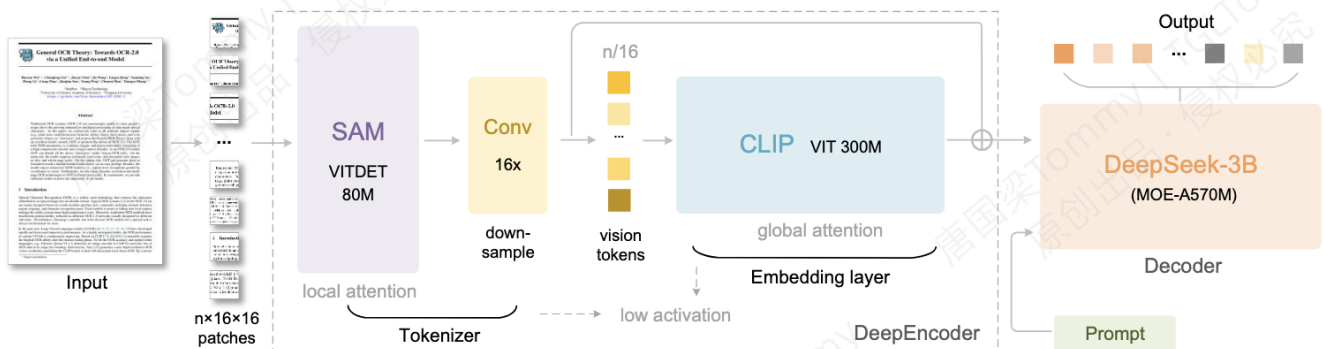


DeepSeek-OCR 2.0 快速解读版

1. 核心痛点

传统的 Vision Transformer (ViT) 处理图像时采用**光栅扫描 (Raster-Scan)** 顺序（即从左到右、从上到下）。这是一种**物理坐标**的顺序，而非**语义逻辑**的顺序。对于复杂文档（如报纸排版、数学公式、表格），物理顺序往往是乱序的，强行 Flatten 导致模型必须花费大量参数去“纠正”空间位置与逻辑语义的错位。

DeepSeek OCR V1.0

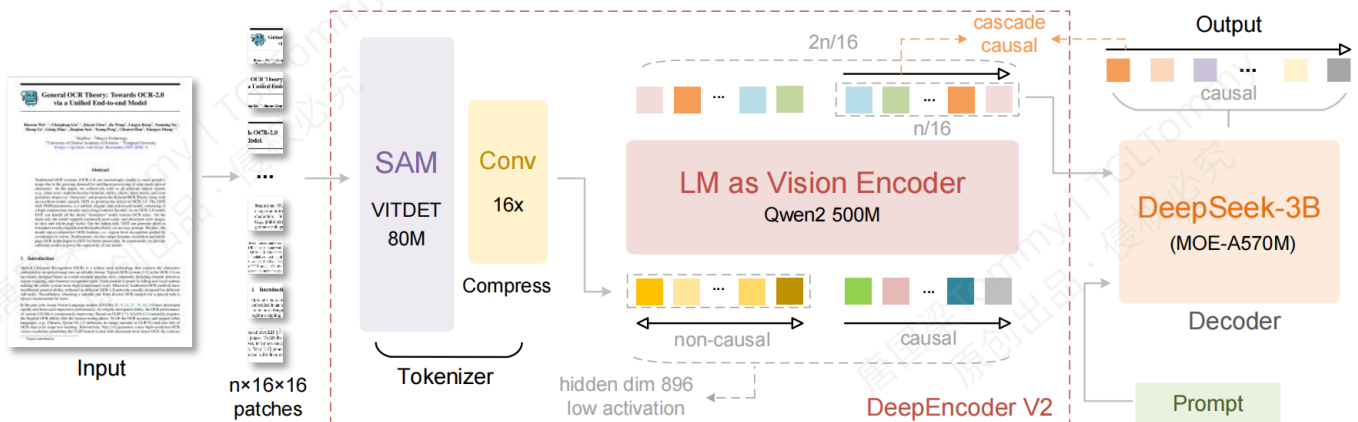


2. 核心方案

DeepEncoder V2 是一个基于 LLM (Qwen2-0.5B) 魔改的视觉编码器。

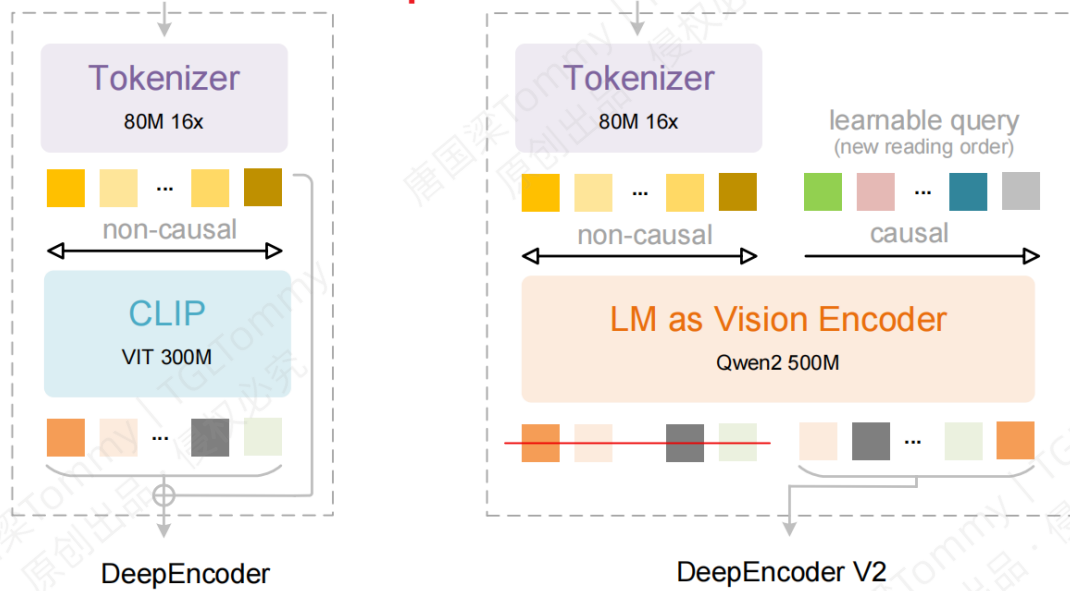
它引入了可学习的 **Causal Queries**，在编码阶段就完成了视觉 Token 的“重排序”和“压缩”，让编码器具备了类似人类视线移动的**逻辑推理能力**，而非死板的像素扫描。

DeepSeek OCR V2.0



- **DeepEncoder V2**: 负责**视觉逻辑推理**。它将无序的 2D 图像块，按照语义逻辑（如阅读顺序）重组为有序的 1D 序列。
- **DeepSeek-MoE Decoder**: 负责**任务逻辑推理**。它基于已经理顺的视觉信息，进行文本生成（Markdown 输出）。

DeepSeek OCR V2.0



- **前缀拼接**: 将 V 作为前缀, 后面拼接一组可学习的 **Causal Flow Queries (Q)**。

$$X_{input} = [\underbrace{V_1, V_2, \dots, V_N}_{\text{Visual Tokens (Prefix)}}, \underbrace{Q_1, Q_2, \dots, Q_M}_{\text{Causal Queries (Suffix)}}] \quad (1)$$

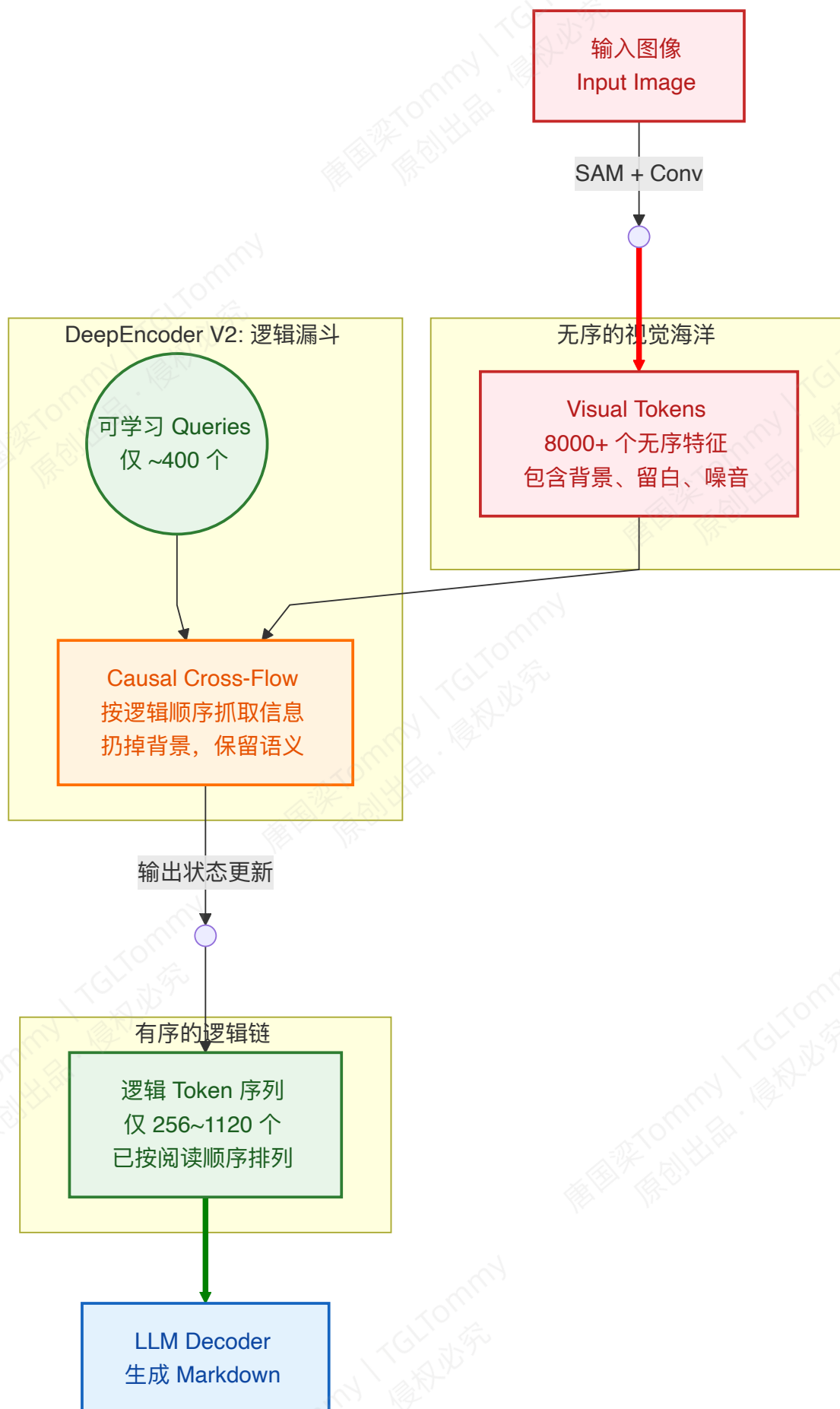
- V (**Visual Tokens**): 由前端的 SAM-base + 卷积层生成, 保留了 2D 空间信息。
- Q (**Causal Flow Queries**): 随机初始化的可学习向量 (Learnable Embeddings), 承载逻辑顺序。

- **混合注意力 (关键)**:

- V 内部: **双向可见 (Bidirectional)**, 像 ViT 一样, 每个像素都能看到全局。
- Q 对 V : **全可见**, Query 可以“看”到整个图像来提取信息。
- Q 对 Q : **因果可见 (Causal/Autoregressive)**, 即 Query i 只能看到 Query $0 \dots i-1$ 。

- **输出**: 仅输出 Q 经过编码后的状态。此时, 这些 Token 已经按**逻辑顺序**排列, 且不再包含冗余的背景信息。

3. 数据流



AI进阶学习

 B站课堂: <https://space.bilibili.com/3546748815411393/pugv>

 官方网站: <https://www.tgltommy.com/> (国内访问需科学上网)

如果你正在关注大模型 Agent、强化学习后训练、RLHF、DPO、GRPO、RLVR 等前沿方向, 欢迎关注我最新上线的精品课程:
[《大模型 Agent 强化学习实战: 从后训练、奖励设计到工业级对齐系统》](#)

这门课程围绕当前大模型 Agent 强化学习的核心技术路线展开, 不只是介绍零散算法和论文概念, 而是希望帮助学习者建立一套完整的技术地图: 从基础原理、论文脉络、关键算法, 到开源项目源码解析与工业级系统实践, 逐步理解 Agent RL 为什么重要、如何演进, 以及在真实应用中如何落地。

课程配套专属飞书知识库《Agent RL 学习宝典》, 包含学习路线、论文资料、项目说明、源码阅读辅助资料和后续持续更新内容, 方便长期查阅和跟进前沿技术。

Agent RL 变化很快, 但越是变化快, 越需要一套系统的学习框架。如果你希望系统进入 Agent RL 方向, 而不是停留在碎片化了解阶段, 这门课会是一个很好的起点。

课程官网: <https://www.tgltommy.com/p/agent-rl> (国内访问需科学上网)

B站课堂: <https://www.bilibili.com/cheese/play/ss842375604>

更多课程详情, 扫描以下二维码咨询课程助理:

新课首发

大模型 **Agent**
强化学习实战:
| 从后训练、奖励设计到工业级对齐系统

唐国梁Tommy



官方微信课程助理



B站/公众号/YouTube: 唐国梁Tommy

**关注微信公众号
获取更多学习资源**



官方网站: TGLTommy.com(需科学上网)