AI大模型应用：NLP与大模型

2025年

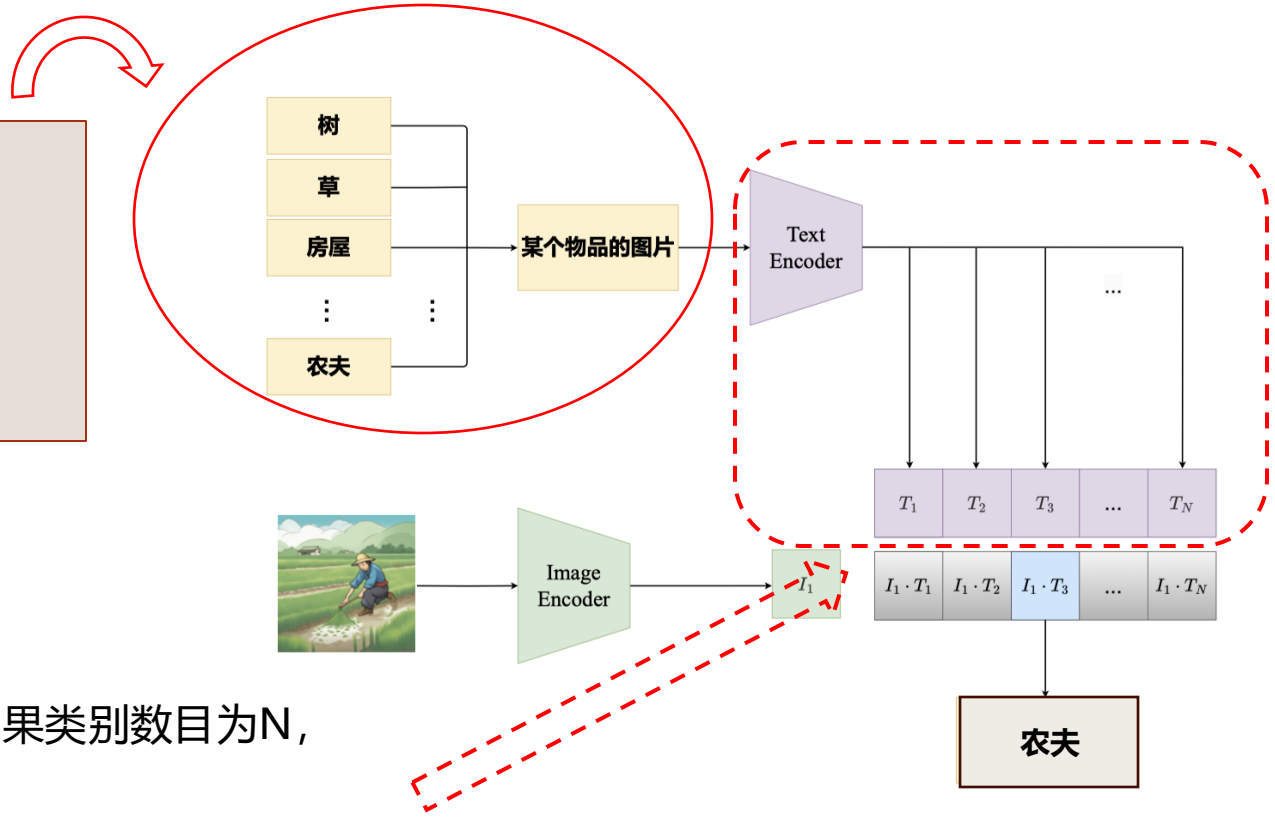
内部资料，请勿外传

CLIP模型

- **利用CLIP模型做零样本图像分类步骤：**

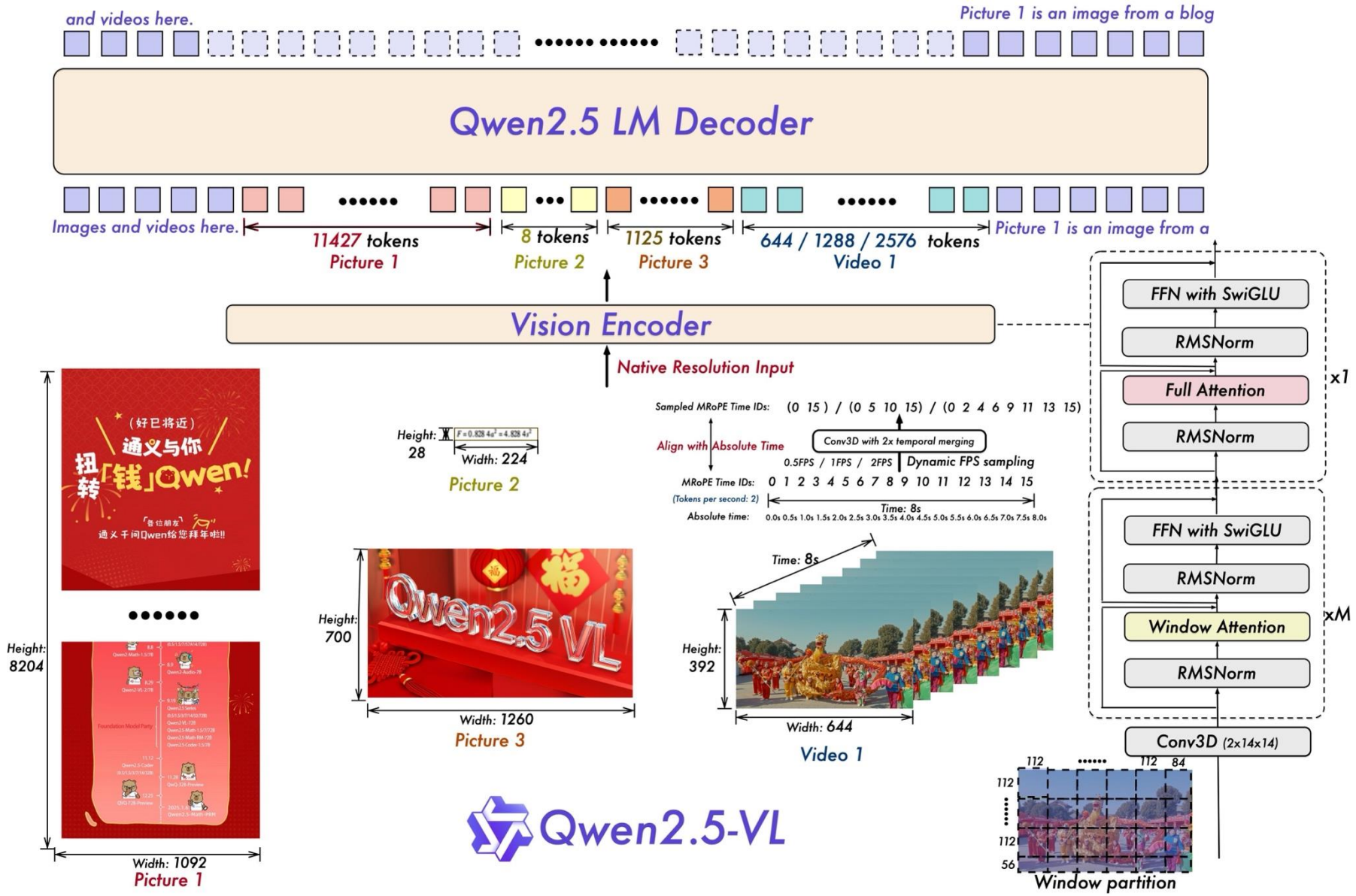
  - **分类标签转换**：根据任务的分类标签构建每个类别的描述文本：A photo of {label}/某个物品的图片

例如： 对于事物分类任务，标签集可以是[“树”，“草”，···，“农夫”]：**为每个标签生成文本描述**，如“树的图片”和“草的图片” 等。



  - **特征抽取**：将这些文本送入文本编码器，如果类别数目为N，将得到N个类别特征

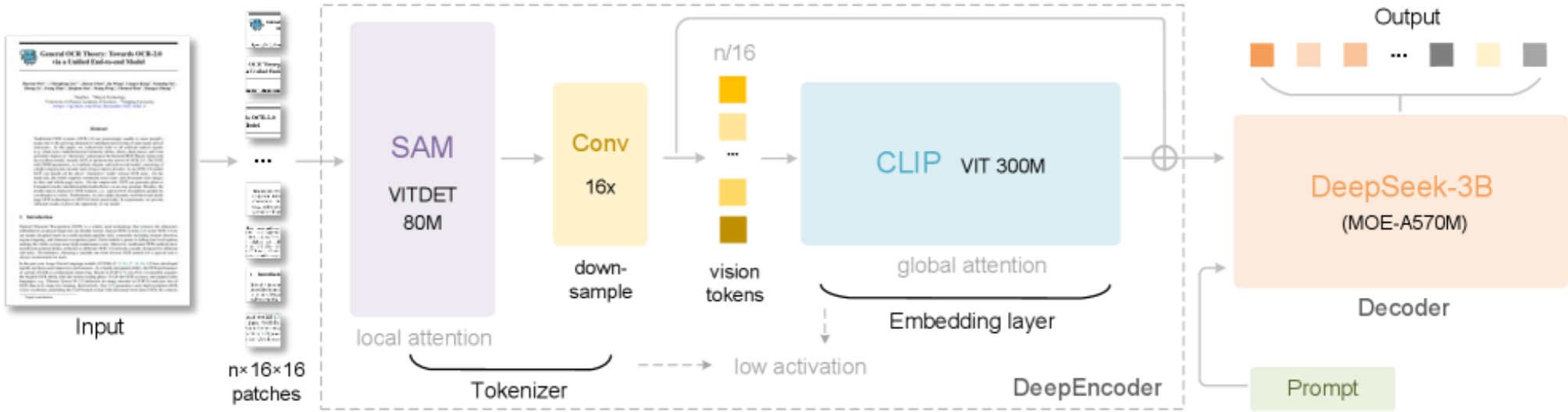
Qwen-VL (202504)



DeepSeek-OCR

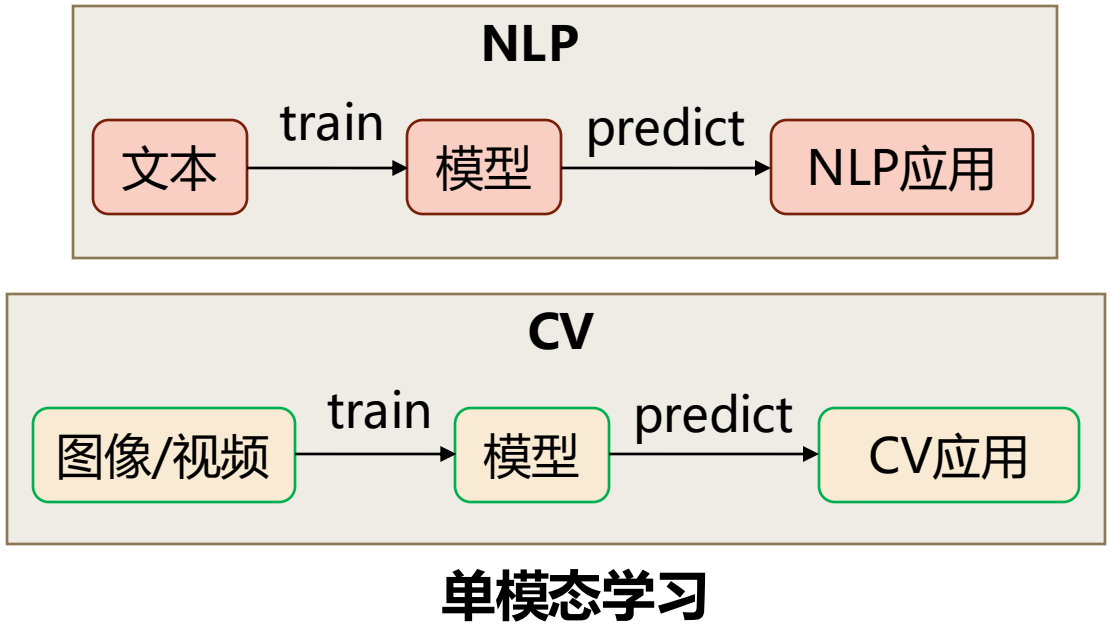
DeepEncoder（视觉编码器）： 负将二维文本图像映射成数量大大减少的视觉 tokens。
窗口注意力 和 **全局注意力编码器** 的串行连接，以高效处理高分辨率输入。
卷积压缩器 在进入全局注意力之前大幅减少视觉 tokens的数量。

DeepSeek3B-MoE-A570M（解码器）： 紧凑的 **专家混合（MoE）** 架构模型，它接收压缩后的视觉 tokens，并将其解码回原始的文本信息。MoE 设计使其能够高效推理并保持高准确率。

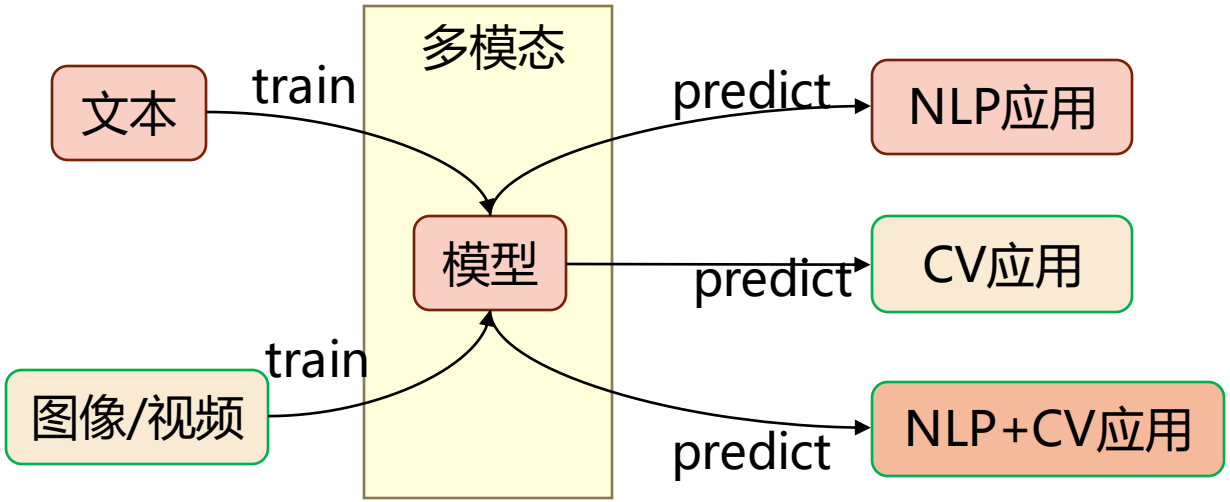


多模态大模型

多模态大模型是一种能够处理和理解多种类型数据（如文本、图像、音频和视频）的人工智能模型。



单模态学习



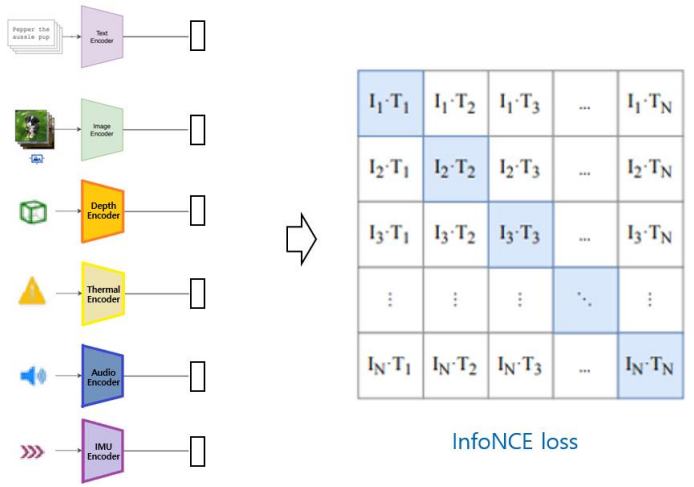
多模态学习

多模态任务解决范式

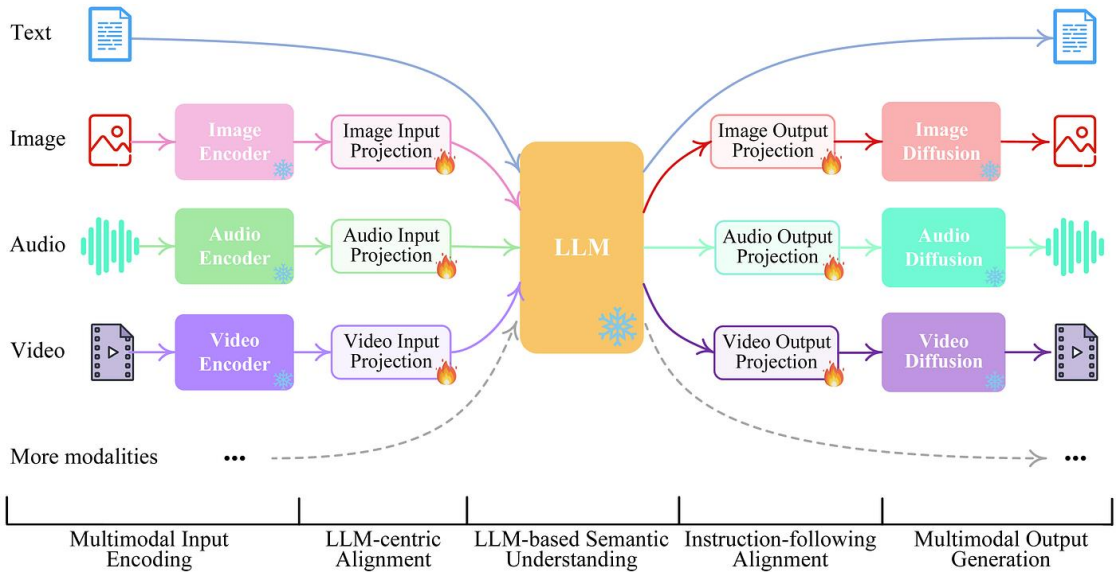
案例1：医疗健康行业下影像诊断与报告生成

案例2：电子商务/零售行业下商品内容理解与搜索

- ✓ 传统的解决范式（基于模块化或特征融合）
  - ✓ 独立的特征提取与决策融合
  - ✓ 跨模态特征融合



- ✓ 基于大型预训练模型（LMPs/LMMs）的范式
  - ✓ 共享嵌入空间（CLIP）
  - ✓ 统一的生成架构下的提示学习/微调范式（Qwen-VL）



案例1：医疗健康行业下影像诊断与报告生成

行业应用	传统范式案例	新范式（大模型）案例
疾病诊断辅助	**模块化诊断系统**： 1. **图像模块：** 训练一个 **CNN** 模型（如ResNet）专门识别影像中的病灶区域。 2. **文本模块：** 训练一个 **LSTM/RNN** 模型处理病历中的既往史、症状描述。 3. **融合：** 将两个模块的结果简单加权或拼接，判断最终的疾病分类（如“肺部感染”）。	**多模态医疗LLM (e.g., Med-PaLM/GPT-4V)**： 1. 统一模型摄入影像和病历。 2. **推理能力**：模型能综合判断：“尽管X光片病灶不明显，但结合患者高烧、咳血的描述（文本），应优先考虑结核病，建议进行进一步检查。” 3. **生成**：自动生成结构化诊断报告和下一步建议。
优势/差异	**优势：** 针对特定疾病和影像类型准确度高，模型结构简单，易于监管。 **局限：** 无法理解复杂的临床情境，**缺乏推理和整合能力。**	**优势：** 强大的**跨模态推理**和**临床知识整合**能力，能处理多张不同类型的影像和长篇病历。 **差异：** 从分类器升级为**智能临床助手。**

案例2:电子商务/零售行业下商品内容理解与搜索

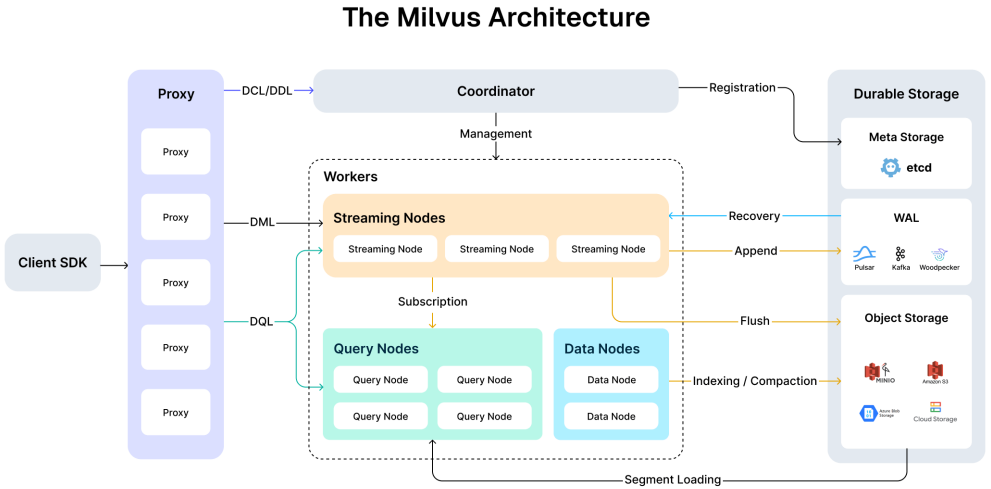
行业应用	传统范式案例	新范式（大模型）案例
跨模态搜索/推荐	**独立检索系统**： 1. **图像检索：** 基于 VGG/ResNet 提取商品图片特征，实现**以图搜图。** 2. **文本检索：** 使用 **TF-IDF/BERT** 等模型处理用户查询和商品标题，进行文本匹配。 3. **结果排序：** 将图像和文本的相似度分数简单相加，进行最终排名。	**统一嵌入空间模型 (e.g., CLIP in E-commerce)**： 1. 将用户查询（如“适合春天的淡蓝色裙子”）和所有商品的图片/标题**映射到统一语义空间。** 2. **语义匹配**：直接计算查询向量和商品向量的相似度，实现更精准的“以文搜图”或“以图搜文”。
优势/差异	**优势：** 结构清晰，便于优化各自的检索精度。 **局限：**无法理解“淡蓝色”在图片中的具体体现，**模态间的语义鸿沟大。**	**优势：** 极强的**泛化性和语义理解**能力，能捕捉细微的视觉和文本概念差异，实现**概念级**的搜索和推荐。 **差异：** 从特征匹配升级为**语义理解。**

Milvus介绍与使用

Milvus 是一个**开源的向量数据库**，专门用于存储、索引和管理海量**向量数据**，并对其执行快速、高效的**相似性搜索**。它是构建推荐系统、图像/视频/音频搜索、自然语言处理（NLP）、检索增强生成（RAG）等各种 **GenAI（生成式 AI）应用**的核心组件。

在 Milvus 中，**Collection** 相当于传统数据库中的**表**，用于存储具有相同 Schema（模式）的向量和元数据。创建 Collection 时，你需要定义字段（包括向量字段的维度）。

- ✓ **定义 Schema：** 确定需要存储的字段及其数据类型。
- ✓ **创建 Collection：** 指定 Schema、集合名称和索引参数。



https://docs.zilliz.com.cn/docs/quick-start

Kafka介绍与使用

Apache Kafka 是一个开源的**分布式事件流平台（Distributed Event Streaming Platform）**，由 LinkedIn 开发并贡献给 Apache 基金会。

- ✓ **分布式和可扩展性**：Kafka 集群可以横向扩展到多台服务器（称为 **Broker**），以处理 PB 级别的数据量。
- ✓ **高吞吐量**：能够以极低的延迟处理数百万条/秒的消息。
- ✓ **持久性**：数据会被持久化（写入磁盘），并且可以设置保留时间，保证数据不会丢失。
- ✓ **容错性**：数据在集群中会被**复制 (Replicated)**，以防止单点故障。
- ✓ **实时性**：支持实时发布、订阅和处理数据流。

概念	解释
Topic (主题)	一种特定类型或类别的数据流。生产者将消息发布到主题，消费者从主题订阅消息。
Producer (生产者)	负责创建和发送消息（记录）到 Kafka Topic 的应用程序。
Consumer (消费者)	负责订阅 Topic 并处理消息（记录）的应用程序。
Broker (代理)	Kafka 集群中的一台服务器实例。Broker 负责存储 Topic 的数据。
Partition (分区)	Topic 被切分成一个或多个分区，分区是 Kafka 存储和并行处理的基本单位。
Record (记录/消息)	Kafka 中传输的最小数据单元，通常包含一个 **Key**、一个 **Value** 和一个**时间戳**。

https://kafka1x.apachecn.org/documentation.html

实操项目1: 商品检索与图文匹配

随着电子商务和内容平台的快速发展，用户对商品和信息的检索需求不再局限于简单的关键词匹配。传统的文本搜索（如基于商品标题或描述）面临“**语义鸿沟**”问题，即用户搜索的词语可能与商品描述的词语不完全一致，导致召回率低。

- ✓ **构建统一的语义空间：** 利用深度学习模型（如 CLIP、BERT 等）将商品图片和文本标题编码成统一的向量嵌入(Embedding)，消除模态差异。
- ✓ **实现全方位的检索能力：** 在此语义空间的基础上，实现图-图、文-文、图-文、文-图 四种模式的**跨模态检索。**

向量编码： PyTorch, Hugging Face (CLIP等模型)

向量数据库： Milvus / Elasticsearch k-NN

服务框架： Python (FastAPI/Flask)

数据集： 实际电商商品图片和标题数据集 或 公司内部数据集

实操项目1: 商品检索与图文匹配

编号	功能模块	描述	关键产出
A1	图搜索图	输入一张商品图片，检索并返回库中 最相似 的 N张商品图片（应用于相似款查找、重复商品检测）。	相似图片列表
A2	文本搜索文本	输入一段文本标题，检索并返回库中 语义最相关 的 N个商品标题（应用于语义搜索、近义标题匹配）。	语义相似标题列表
A3	图搜索文本	输入一张商品图片，检索并返回库中 语义最匹配 的 N个商品标题（应用于图片描述、商品自动打标题）。	基于图的合适标题 列表
A4	文本搜索图	输入一段文本标题，检索并返回库中 最符合描述 的 N 张商品图片（应用于关键词搜索，但基于语义理解）。	基于标题的合适图片 列表
A5	文本生成图 (文生图)	输入一段商品标题或描述文本，生成一张与其语义内容相符的商品图片（应用于商品设计辅助、内容创意）。	基于标题生成的合适图片

实操项目2: 财报多模态问答

企业财报是高度结构化和非结构化数据混合的文档，包含大量的文字叙述、复杂的表格（如资产负债表、利润表）以及各类图表（如趋势图、饼图）。本项目旨在构建一个能够深度理解财报 PDF 中**文本、表格和图像**三种模态信息的智能问答系统，以支持金融分析师和投资者的高效决策。传统的 RAG（检索增强生成）系统在处理 PDF 文档时，往往仅关注纯文本，导致以下问题：

- ✓ **信息丢失：** 表格和图表中的关键数据和趋势无法被有效提取和利用。
- ✓ **上下文不完整：** 提问关于“去年销售增长情况”时，如果回答仅依赖文本描述而忽略了关键的销售趋势图，答案的准确性和深度将受到限制。
- ✓ **格式错乱：** PDF 解析器难以准确识别复杂的表格结构，导致数据混淆。

PDF 解析/结构化： MinerU、PaddleOCR、DeepSeek-OCR

向量化： Text Embeddings

向量数据库： Milvus / Elasticsearch k-NN

LLM： GPT-4, Claude, 或 Llama/Mistral 等开源模型

实操项目2: 财报多模态问答

编号	功能模块	描述	关键示例问题
A1	**复杂文本问答**	基于财报的文字叙述部分进行 RAG 问答。	"公司本年度的战略目标和主要挑战是什么？"
A2	**表格数据提取与问答**	能够识别并解析 PDF 中的表格结构，并基于表格内容进行精确问答。	"**第三季度**的**净利润**相比**第二季度**的**增长率**是多少？"
A3	**图表内容提取与问答**	能够识别财报中的各类图表（柱状图、折线图等），提取其核心数据、趋势，并进行问答。	"从**近五年营收趋势图**来看，最高增速发生在哪一年？"
A4	**多模态整合问答**	**结合文本、表格、图像三种信息，进行综合性推理并回答问题。**	**"根据附注中的表格和营收增长图，请分析本年度增长的主要驱动力并总结其影响。"**