RESEARCH READING GUIDE

焕羽研究阅读指南与实验设计说明

壹典 IDENIFE · 方法说明 · 版本 1.0 · 2026-09-23

本文围绕跨模态表示、领域语义、条件生成与评价，提供原始论文阅读路线和可执行的实验设计建议。这是壹典的原创方法说明，不是已发表的学术论文，也不报告焕羽已完成的实验、性能指标或已实现的模型架构。所引论文来自外部研究者；引用表示方法参考，不表示其作者与壹典存在合作或背书关系。

1. 阅读路线：从表示到可检验的结果

研究问题	建议阅读	阅读时应区分的层次
图像与文字如何建立对应关系？	CLIP [1]	整体相似度与局部关系正确性
如何组织生成模型中的视觉序列？	DiT [2]	网络结构、潜空间压缩与计算规模
如何学习从噪声到样本的变化？	Flow Matching [3]	训练目标、概率路径与推理求解器
图像与文字如何在生成中交换信息？	MMDiT [4]	模态各自的参数与联合信息交换
如何把空间条件送入生成过程？	ControlNet [5]	条件符合度与视觉自然度
如何把图文一致性拆成具体问题？	TIFA [6]	问题质量、视觉判断与人工复核
模型置信度能否代表正确概率？	Calibration [7]	置信度、正确率与分布变化
如何在质量与成本之间选择模型？	RouteLLM [8]	路由策略、任务分布与质量约束

文献年份采用 arXiv 首发年份。DiT、Flow Matching 的首发年份是 2022；会议发表年份可能不同。配套的 BibTeX 文献库 (huanyu-references.bib) 对应同一组 arXiv 记录，并保留完整作者列表。

METHOD NOTE / 02

2. 跨模态表示：先检索，再核对关系

CLIP 通过图文配对学习可比较的视觉与文本表示，为检索和自然语言描述的视觉分类提供了基础。原始论文 1

面向工业设计，可以提出一个独立的实验问题：领域描述是否能帮助模型区分外观相似、功能关系不同的方案？以下是实验建议，不是 CLIP 论文中已经验证的工业设计结论。

01 以项目、产品系列和设计衍生关系为单位划分训练、开发与测试数据，避免同一方案的改色、裁切或连续版本跨集合出现。

02 为每个查询准备正确样本及难负例。难负例只改变一个关键关系，例如把顶部接口移到侧面，而保持产品类别和整体风格相近。

03 比较原始简报、标准化领域描述和去掉关系词的描述；固定候选库、编码器版本和检索设置。

04 报告 Recall@K、关系难例的错误类型和每个产品类别的样本数。多幅图均满足要求时，应保留多个相关项，避免错误地把合法答案当作负例。

高相似度只能作为候选筛选依据。接口数量、部件上下关系或材料未知状态，应由独立检查确认。

METHOD NOTE / 03

3. 领域语义：保留事实、要求与未知

领域语义层是本文提出的数据组织方法，并非上述论文直接给出的工业设计标准。建议每条样本保留：来源与授权范围、项目与版本标识、产品类别、可见部件、关系、简报要求、材料及工艺描述、未知字段，以及标注者和复核记录。

以下是虚构的说明性样本，不来自公司项目：

字段	示例值	证据类型
产品类别	便携式检测终端	简报给定
操作条件	戴手套使用	简报要求，不能仅凭渲染图验证
接口位置	顶部	可见图像区域与简报分别记录
外壳材料	未指定	未知，不推断为金属
内部装配	未知	无结构图或实物依据

进行标注试验时，先让至少两位标注者独立处理同一小批样本，再核对分歧原因并修订规范。把“看得见”“要求如此”“需要专业推断”分开，避免将视觉印象升级为制造事实。保留冲突与未知的数量，不应只统计字段填充率。

METHOD NOTE / 04

4. 条件生成：一次改变一个因素

DiT 研究在潜图像块序列上使用 Transformer；Flow Matching 研究通过向量场回归训练生成过程；MMDiT 使用模态各自的权重并让图文信息相互交流；ControlNet 则为预训练扩散模型引入空间条件控制。这些工作处理的问题互有关联，但网络结构、训练目标和条件输入不是同一个概念。论文 2、论文 3、论文 4、论文 5

可执行的比较应先限定一个支持所需条件的模型和任务：

实验组	输入变化	主要检查
A	文字条件	对象、数量与语义关系
B	同一文字 + 轮廓条件	布局符合度与整体形态
C	同一文字 + 深度条件	前后关系与遮挡一致性
D	局部编辑 + 明确保留区域	修改是否发生，以及未编辑区域漂移

条件模态不可用时，应记录“该模型不支持”，而非用不同模型替代后仍把差异归因于条件。跨模型比较须把模型本身作为独立变量。

在实验前固定数据划分、提示词、输出分辨率、推理预算、条件预处理和评价规则；支持时使用预先记录的随机种子列表，并保存实际输出。种子相同不保证不同系统产生可直接配对的随机过程。除条件符合度外，还要检查细小部件、文字、边界和未编辑区域；外观图像本身不能证明尺寸、装配或制造可行性。

METHOD NOTE / 05

5. 评价：每个判断都有问题和依据

TIFA 把文字描述转化为问答项，再用视觉问答判断图像是否满足描述，提供更细粒度的图文忠实度检查思路。原始论文 6

建议在生成前固定问题清单，避免看过输出后选择容易通过的问题。例如，“顶部是否有接口？”“可见接口有几个？”“显示区域与握持区域是否分离？”可用于图像检查；“长期戴手套是否舒适？”需要其他证据与实物验证。

每个问题记录预期答案、证据区域、模型答案、专家复核和失败类型。删除含糊或无法从图像回答的问题时，应保留删除原因。分别报告对象、属性、数量、关系、条件保持和证据不足等切片，不用一个平均分掩盖明确的关键条件失败。人工复核应打乱模型和左右位置，保留评审分歧。

评价器也可能误读图片。抽样核对自动判断与人工判断的一致性，并单独分析“生成正确但评价器判错”与“生成错误但评价器通过”。比较版本时，按项目或简报进行成组重采样来估计区间，避免把同一简报的多个输出当作完全独立的任务。

METHOD NOTE / 06

6. 置信度与路由：先定义正确，再设阈值

Calibration 论文研究分类器预测概率与实际正确率的关系，并讨论温度缩放等后处理方法。温度缩放需要相应的分类分数与独立校准数据；不能直接把自由文本里的“我有把握”当成已校准概率。原始论文 7

建议先定义可复核的正确事件，例如“顶部接口检查通过”，再划分开发、校准与最终测试数据。记录可靠性图、每个置信区间的样本数及校准误差的分箱规则，同时观察错误率与保留任务比例的关系。进入新产品类别或新输入质量时，应重新验证阈值的适用性。

RouteLLM 利用偏好数据学习在不同语言模型之间做质量与成本的取舍。将这一思路用于多模态任务，需要另行定义适用的输入特征、能力边界和偏好标注，不能直接继承其结论。原始论文 8

可比较“始终使用低成本模型”“始终使用高能力模型”和“按任务路由”三个基线。预先约定关键条件的合格要求，再报告通过率、总调用成本、延迟分布、复核比例和拒答或澄清比例。阈值在开发集选择，最终测试集只用于一次独立检验。

METHOD NOTE / 07

7. 最小实验记录

每次实验建议保存以下记录，便于团队复核或在条件一致时重复：

实验标识与日期：

问题与待检验假设：

数据版本、来源范围与分组划分：

模型、权重或服务版本：

基线与唯一改变的因素：

提示词、条件文件及预处理：

随机种子、采样设置与预算：

预先固定的评价问题、阈值和关键条件：

人工复核规则、分歧与未知状态：

输出文件、失败类型与成本记录：

结果及不确定性：待实验完成后填写

适用范围与后续问题：

REFERENCES

参考文献

以下均为外部原始论文，年份为 arXiv 首发年份；完整作者与引用字段见配套 BibTeX。

1. Radford et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. 论文条目 · 原文 PDF。引用键：radford2021clip。

2. Peebles & Xie (2022). Scalable Diffusion Models with Transformers. 论文条目 · 原文 PDF。引用键：peebles2022dit。

3. Lipman et al. (2022). Flow Matching for Generative Modeling. 论文条目 · 原文 PDF。引用键：lipman2022flowmatching。

4. Esser et al. (2024). Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. 论文条目 · 原文 PDF。引用键：esser2024mmdit。

5. Zhang, Rao & Agrawala (2023). Adding Conditional Control to Text-to-Image Diffusion Models. 论文条目 · 原文 PDF。引用键：zhang2023controlnet。

6. Hu et al. (2023). TIFA: Accurate and Interpretable Text-to-Image Faithfulness Evaluation with Question Answering. 论文条目 · 原文 PDF。引用键：hu2023tifa。

7. Guo et al. (2017). On Calibration of Modern Neural Networks. 论文条目 · 原文 PDF。引用键：guo2017calibration。

8. Ong et al. (2024). RouteLLM: Learning to Route LLMs with Preference Data. 论文条目 · 原文 PDF。引用键：ong2024routellm。

书目信息核对日期：2026-09-23。未标版本号的 arXiv 链接指向最新修订；复现实验时应另外记录实际使用的论文版本、代码版本和模型版本。