

图像模态：场景结构化与人脸情绪融合说明

多模态信息结构化提取与情绪变量构造系统 · 按当前代码整理

先回答一个问题：一张图片里有哪些可研究的对象、人物和事件？如果有脸，场景情绪与人脸情绪是否相同？

一、理解版：从问题到结果



一张图片不能直接拿去做回归分析，就像一张街景照片不能直接告诉我们“有几个人、发生了什么、气氛如何”。系统先请一位观察员描述整张图片，再在有人脸时请另一位观察员专门检查面部表情，最后把两份观察记录合在一起。

第一位观察员会回答一组固定问题：画面里有什么？有哪些人？发生了什么动作？哪些细节让人觉得开心、紧张或严肃？这样，原本只能“看”的图片，就变成了可以逐行比较的文字记录。

如果画面里有人脸，系统再请一位只看脸部表情的专门观察员。它会记录看到了几张脸、每张脸更像哪种情绪，以及这个判断有多确定。没有人脸、没有配置这项服务，或者服务暂时不可用时，就保留第一位观察员的场景判断。

最后采用一个很容易解释的规则：只要人脸观察员给出了有效答案，就优先使用它；否则使用场景观察员的答案。这里不是把两个答案按比例相加，而是把两种来源都留下，让研究者知道最终标签究竟来自哪里。

因此，图片分析的结果不仅有“最终情绪”，还有场景摘要、人物、物体、动作、情绪线索、人脸数量和判断理由。研究者可以先看最终变量，再回到这些原始线索检查它是否合理。

边界提醒：图像主流程没有本地视觉情绪模型，也没有自动裁剪人脸；百度接口是可选的外部服务。若没有百度凭证，fusion/variables.csv 中的 API 字段为空是正常回退，不代表 Qwen 没有识别情绪。

二、学术版：可直接放入方法部分的表述

图像首先依据扩展名和 MIME 类型读取并编码为 base64 data URL，输入 Qwen qwen3.6-plus 提取场景级对象、人物、事件、情绪线索和语义情绪；在配置百度人脸 V3 凭证时，进一步请求 emotion、expression、face_shape 和 quality 字段，获得逐脸情绪类别及概率；最终按照“有效人脸 API 标签优先、否则采用 Qwen 标签”的规则构造融合情绪变量，并保留人脸数量与概率信息。

三、实际使用的包、模型与职责

组件/模型	作用	本项目实际怎么用
Python pathlib / mimetypes / base64	输入读取与 编码	检查扩展名，猜测 MIME 类型，读取原始字节并生成 data:image/...;base64,...。
OpenAI-compatible client	调用 Qwen	以 qwen3.6-plus、DashScope compatible base URL 和 JSON schema 完成场景结构化与情绪判断。
百度人脸 V3 detect（可 选）	人脸情绪核 对	换取 access_token 后请求 emotion、expression、face_shape、quality，最多 10 张脸。
Python csv / json / tqdm	批处理和结 果保存	支持断点、重试、失败图片归档，并输出 llm/ 和 fusion/ 两组结果。

四、按代码复现：每一步输入什么、做什么、得到什么

步骤	人话说明与复现动作	得到什么
1. 准备输入	上传 .jpg、.jpeg、.png、.webp 或 .bmp。程序默认读取 data/images，也可用 --input 指定单图。	原始图片
2. 编码图片	用 mimetypes 判断 MIME，用 base64.b64encode 读取文件字节；当前没有额外裁剪、缩放或 Pillow 预处理。	图片 data URL
3. Qwen 场景提取	向 qwen3.6-plus 发送 image_url + JSON schema；成功响应应提取第一个 JSON 对象，列表字段以 展平。	outputs/llm/per_image/ *.json 与 batch_results.csv
4. 可选人脸接口	若配置百度 API Key/Secret Key，先换 access_token，再请求人脸情绪；若无 token，API 结果会明确记录 unavailable。	num_faces、face_emotions、 face_probabilities
5. 规则融合	emotion_label_from_api 非空时作为 final_emotion_label；否则使用 emotion_label_from_llm，并写出 fusion_reason。	outputs/fusion/ variables.csv
6. 检查失败与重试	网络超时、连接错误、429/5xx 等可重试；失败图片复制到 images_failed，progress.json 记录阶段状态。	run_summary.json / progress.json

五、公式、权重与融合规则

$$FinalLabel_i = \begin{cases} FaceLabel_i, & \text{有有效的人脸判断} \\ LLMLabel_i, & \text{否则} \end{cases}$$

LaTeX 写法（可复制到论文编辑器）

```
\operatorname{FinalLabel}_i = \begin{cases} \text{FaceLabel}_i, & \text{numFaces}_i > 0 \\ \text{LLMLabel}_i, & \text{否则} \end{cases}
```



```
p_{i,r,k} = \Pr(\text{第 } r \text{ 张脸属于情绪 } k)
```

优先级融合： $\text{FinalLabel}_i = \text{FaceLabel}_i$ ，当 $\text{num_faces} > 0$ 且 API 返回有效标签；否则 $\text{FinalLabel}_i = \text{LLMLabel}_i$ 。这是可解释的条件规则，不是数值加权。

人脸概率： 对第 r 张脸，接口返回 $p_{i,r,k}$ （情绪类别 k 的概率）；当前 CSV 原样保存这些概率，代码没有再对多张脸求平均。

公式中的英文变量名与实际输出列保持一致；若某一步是规则优先级而不是连续加权，文中已明确标出。

六、关键输出变量：英文名与中文解释

英文变量名	中文解释
image_id	图片编号（文件名去掉扩展名）
summary	场景的一句话摘要
objects	图片中识别出的物体或视觉元素
people	人物或人物角色
events	可见动作或事件
emotion_cues	支持情绪判断的视觉线索
emotion_label / emotion_reason	Qwen 的场景情绪标签及理由
num_faces	百度接口检测到的人脸数
emotion_label_from_api	百度人脸情绪标签
face_probabilities_raw	每张脸的情绪类别概率原始串
final_emotion_label	人脸标签可用时优先，否则使用 Qwen 标签
fusion_reason	说明最终标签来自 API 还是 LLM
time / location / risk_tags	Qwen 可选结构化字段：时间、地点、风险标签

七、结果文件与研究用途

结果文件	用途
outputs/llm/batch_results.csv	场景结构化与 Qwen 情绪

结果文件	用途
outputs/llm/per_image/*.json	每张图的完整 JSON
outputs/fusion/variables.csv	人脸情绪、Qwen 情绪和最终融合标签
outputs/fusion/per_image/*.json	每张图的融合详情和概率
outputs/llm/run_summary.json / outputs/fusion/ run_summary.json	两阶段成功/失败/重试摘要

objects、people、events 可以转成计数、类别或主题变量；emotion_cues 和 emotion_reason 用于人工复核；final_emotion_label 可用于图片情绪分组。做论文时应区分“没有人脸”和“接口失败”，不能把两者都解释为中性。

本篇说明作为项目复现参考，更多字段、原始响应和运行日志请结合下载结果中的 JSON、CSV、progress.json 与 config_snapshot.json 一起核对。