

AI 绘画评估脚本使用教程与原理说明

第一部分：使用教程 (How to use)

步骤 1：准备环境 (安装依赖)

你需要安装一些 Python。打开你的端（Terminal 或 CMD），运行以下命令：

1. 安装基础数据处理和图像处理库

```
pip install numpy pandas matplotlib seaborn tqdm opencv-python pillow
```

2. 安装 AI 相关的库

MediaPipe 用于姿态识别

```
pip install mediapipe
```

```
pip install torch torchvision
```

LPIPS 用于计算图片感知差异

```
pip install lpips
```

OpenAI 的 CLIP 用于计算文本匹配度（直接从 GitHub 安装）

```
pip install git+https://github.com/openai/CLIP.git
```

步骤 2：准备数据文件夹 (关键！)

脚本的核心依是 `metadata.csv` 文件。你需要按照以下构整理你的文件：

```
MyProject/
|
├── scripts/
|   └── evaluate.py          <-- 把上面的代码保存为这个文件|
├── outputs/
|   └── experiments/
|       └── exp_v1_test/     <-- 这是一个实验文件夹
|           ├── metadata.csv <--- 【必须有这个文件】
|           └── orig_001.jpg <--- 原图
```

└─ gen_001.jpg <--- 生成图
└─ ...

metadata.csv 必包含表，每一行对应一对“原图-生成图”。你可以用 Excel 然后另存 CSV。

original_image_path	generated_image_path	prompt	seed
orig_001.jpg	gen_001.jpg	girl running	123
orig_002.jpg	gen_002.jpg	boy sitting	456

步骤 3：运行脚本

在终端中，输入 MyProject 根目录，运行以下命令：

```
python scripts/evaluate.py
```

```
--exp_dir outputs/experiments/exp_v1_test
```

步骤 4：查看结果

运行结束后，去 **eval_results** 文件夹里看：

1. summary_exp_v1_test.md: 包含通过率，平均分，结论。
2. results_exp_v1_test.csv: 每一张图片的具体打分。
3. charts/: 里面有自动画好的统计图表。

第二部分：原理白话文解释

1. 姿态评委 (Pose Consistency)

负责检测生成的任务姿态有没有变

原理：使用 MediaPipe 技术提取人体 33 个关节点（骨架），把两张图骨架堆叠起来，计算对应的关节距离，进行打分，MSE 越接近 0，动作越还原

2. 风格评委 (Style/LPIPS)

负责检查图像的结构和感觉像不像原图

原理：使用深度神经网络（vgg/alexnet）来提取特征，LPIPS 越接近 0，越准确。

3. 语义评委 (CLIP Score)

负责检查生成的图是不是你想要的“意思”（与 Prompt 的匹配度）。

原理：使用 CLIP 模型将图片和文字映射到同一个数学向量空间里的坐标，CLIP Score 计算生成图和提示词的余弦相似度。分数越高，生成图越符合文字描述。

总结速查表

指标名称	中文名	怎么看好坏？	实际含义
Pose MSE	姿态误差	越小越好 (接近 0)	动作准不准。>0.05 说明动作严重变形。
LPIPS	感知差异	越小越好	结构像不像原图。 0.3~0.6 是正常风格化范围。
CLIP Score	文本匹配度	越大越好	是否符合 Prompt。>0.25 算合格。
Style Fidelity	风格符合度	越大越好	含不含特定风格元素。

常见报错与解决

- 1. FileNotFoundError: Metadata not found -> 路径填错了。
- 2. CUDA out of memory -> 显存不足
- 3. OSError: Can't load LPIPS/CLIP ->网络问题。