

PR #26044 完整报告

sgl-project/sglang

ci: guard diffusion gt publishing

合并时间: 2026-05-22 19:07

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/26044>

执行摘要

- 一句话: 为 diffusion CI 添加 GT 图像质量门, 防止发布低质量 / 噪声图像
- 推荐动作: 值得精读, 尤其是质量指标计算和远程 Blob API 交互的实现。展示了 CI 流程中引入数据质量门的实践模式, 可复用于其他需要资产质量验证的场景。

功能与动机

防止低质量或随机噪声的生成图像被发布为 CI 的 ground truth, 从而影响后续一致性测试的有效性。PR body 明确指出需要“add a quality gate before publishing changed diffusion GT images”和“reject no-reference low-detail/random-noise outputs before creating GitHub blobs”。

实现拆解

1. 质量指标定义: 在 `scripts/ci/utils/diffusion/publish_diffusion_gt.py` 中添加 `ImageQualityMetrics` 和 `OldNewMetrics` 两个 frozen dataclass, 封装亮度标准差、熵、梯度 P95、邻域相关系数等图像质量维度, 以及 SSIM 和平均绝对差异用于新旧对比。
2. 远程 blob 操作: 新增 `get_remote_image_entries` 函数 (从原来的 `get_remote_blob_shas` 重构而来) 获取远程目录下所有图像文件的元数据; 新增 `get_remote_blob_content` 函数通过 GitHub Blob API 下载原始内容用于对比。
3. 质量检查入口: `main` 函数增加 `--check-only` 模式, 遍历待发布图像, 对每张图计算质量指标: 低于低细节阈值 (低 STD、低熵、低梯度) 或随机噪声特征 (高邻域相关、低频率占比) 则拒绝; 如果有远程旧图, 则计算 SSIM 和均值差, 若漂移过大则拒绝。
4. workflow 集成: 在 `.github/workflows/diffusion-ci-gt-gen.yml` 中, 为每个发布 job (`FLUX.1-dev`、`FLUX.2-klein-base`、`LTX Video`、`Mochi 1 Preview`) 添加 `Validate generated GT images` 步骤, 在发布前调用 `publish_diffusion_gt.py --check-only`。
5. 特殊情形支持: 允许通过 `--allow-replace` 标记覆盖拒绝, 以便运维人员手动替换可疑的旧 GT。该逻辑在检查步骤后, 发布步骤前嵌入交互式决策 (实际 CI 中由环境变量控制)。

关键文件:

- `scripts/ci/utils/diffusion/publish_diffusion_gt.py` (模块 CI 工具; 类别 `infra`; 类型 `infrastructure`; 符号 `ImageQualityMetrics`, `OldNewMetrics`, `get_remote_image_entries`, `get_remote_blob_content`): 核心实现文件, 新增了图像质量检测、远程 blob 比较、旧图替换等逻辑, 是质量门的执行引擎。

- [.github/workflows/diffusion-ci-gt-gen.yml](#) (模块 CI 配置; 类别 infra; 类型 infrastructure) : CI workflow 文件, 在每个模型 (FLUX、LTX Video 等) 的 GT 发布步骤前插入验证环节。

关键符号: ImageQualityMetrics, OldNewMetrics, get_remote_image_entries, get_remote_blob_content, _load_quality_image, _image_to_rgb_array, _luminance, _neighbor_correlation

关键源码片段

[scripts/ci/utils/diffusion/publish_diffusion_gt.py](#)

核心实现文件, 新增了图像质量检测、远程blob比较、旧图替换等逻辑, 是质量门的执行引擎。

```
# dataclass 定义质量指标
@dataclass(frozen=True)
class ImageQualityMetrics:
    luminance_std: float # 亮度标准差
    entropy: float # 信息熵
    blur_residual: float # 模糊残差 (Laplacian 响应)
    gradient_p95: float # 梯度强度 P95
    neighbor_correlation: float # 邻域相关系数
    low_frequency_ratio: float # 低频能量占比

@dataclass(frozen=True)
class OldNewMetrics:
    ssim: float # 结构相似性
    mean_abs_diff: float # 平均绝对差异

# 核心质量检查: 对每张待发布图像
# 1. 低细节检测
if (metrics.luminance_std < LOW_DETAIL_STD_THRESHOLD or
    metrics.entropy < LOW_DETAIL_ENTROPY_THRESHOLD or
    metrics.blur_residual < LOW_DETAIL_BLUR_RESIDUAL_THRESHOLD or
    metrics.gradient_p95 < LOW_DETAIL_GRADIENT_P95_THRESHOLD):
    reason = "low_detail"
# 2. 随机噪声检测
elif (metrics.neighbor_correlation < RANDOM_NOISE_CORRELATION_THRESHOLD and
      metrics.low_frequency_ratio < RANDOM_NOISE_LOW_FREQUENCY_THRESHOLD and
      metrics.blur_residual < RANDOM_NOISE_BLUR_RESIDUAL_THRESHOLD):
    reason = "random_noise"
# 3. 新旧漂移检测 (当存在远程旧图时)
if old_metrics is not None:
    if old_metrics.ssim < OLD_NEW_MIN_SSIM or old_metrics.mean_abs_diff > OLD_NEW_MAX_MEAN_ABS_DIFF:
        drift_reason = "old_new_drift"
```

[.github/workflows/diffusion-ci-gt-gen.yml](#)

CI workflow 文件, 在每个模型 (FLUX、LTX Video 等) 的 GT 发布步骤前插入验证环节。

```
# 新增步骤 (每个发布 job 相同模式)
- name: Validate generated GT images
  env:
    GITHUB_TOKEN: ${ secrets.GH_PAT_FOR_NIGHTLY_CI_DATA }
  run: |
    python scripts/ci/utlils/diffusion/publish_diffusion_gt.py \
      --source-dir python/${ env.OUTPUT_NAME } \
      --target-dir "${ env.PUBLISH_TARGET_DIR }" \
      --check-only
# 原有发布步骤保持不动
```

评论区精华

此 PR 没有多人 review 讨论 (review_comments_count=0) , 主要交互来自作者本人的多次提交迭代, 包括添加允许替换、移除未使用的测试等。CI bot 的评论仅提示配额限制, 无技术讨论。

- PR 本身无 review 讨论 (other): 无技术争议, 纯基础设施变更。

风险与影响

- 风险:
 1. 阈值误调风险: 低细节 / 噪声检测阈值 (如 LOW_DETAIL_STD_THRESHOLD=0.075) 是硬编码的, 可能对某些合法图像过于严格或放过异常, 需要后续根据实际数据调优。
 2. 依赖风险: 新增 PIL 和 numpy 依赖, 但 CI 环境通常已安装; 若环境缺失会导致 check-only 步骤失败。
 3. 网络异常: get_remote_image_entries 和 get_remote_blob_content 依赖 GitHub API , 若网络波动或遭遇 rate limit 可能导致 false positive 拒绝。已有 HTTPError 处理, 但未涵盖所有异常。
 4. 性能开销: 每张图需要下载远程 blob 并计算指标, 可能增加 CI 发布步骤耗时 (约数秒至数十秒) , 但相对发布流程可接受。- 影响: 开发者: 需确保生成的 GT 图像通过质量门, 否则发布步骤会失败; 可以通过 --allow-replace 手动覆盖。系统: GT 仓库 sgl-project/ci-data 中的图像质量将得到保障, 减少因坏图导致的一致性测试误报。CI 流程: 增加一个无侵入的验证环节, 不影响正常发布流程。- 风险标记: 硬编码阈值可能误判, 新增 PIL/numpy 依赖, 网络 API 调用可能失败

关联脉络

- PR #25661 [diffusion] model: support FLUX.2-klein-base: 同为 diffusion 模块的变更, 该 PR 增加的新模型触发了 GT 发布需求, 本 PR 为 GT 发布添加质量门, 形成闭环。