PR #21390 完整报告

sgl-project/sglang

[diffusion] Fix Wan2.2-I2V-A14B video max size and calculate generated video size from the given width and height

合并时间：2026-03-31 21:49

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/21390

执行摘要

- 一句话：修复 Wan2.2-I2V-A14B 视频分辨率过小问题，支持用户指定宽度和高度以计算输出尺寸。
- 推荐动作：建议技术管理者和工程师精读此 PR，关注 input_validation.py 中处理用户尺寸的逻辑设计（如长宽比保持和面积限制），以及如何通过配置继承来管理不同模型的分辨率限制。此外，review 中的讨论展示了 API 设计中的权衡（如 width/height 与 size 的覆盖关系）和向后兼容性考虑，值得学习。

功能与动机

根据 PR body 中的描述：'While using the https://huggingface.co/Wan-AI/Wan2.2-I2V-A14B model, I observed that the generated video resolution is smaller than expected.' 以及作者在 review 中解释，添加 width 和 height 是为了与 ImageGenerationsRequest 行为对齐，方便客户使用相同 JSON dict 进行 T2I（文本到图像）和 I2V（图像到视频）任务。

实现拆解

实现方案分为四个层次：1) 配置层：修改 python/sglang/multimodal_gen/configs/pipeline_configs/wan.py，将 Wan2_2_I2V_A14B_Config 继承自 WanI2V720PConfig（max_area = 1280×720），替换原 480P 配置；2) API 层：在 python/sglang/multimodal_gen/runtime/entrypoints/openai/protocol.py 的 VideoGenerationsRequest 类中添加 width 和 height 可选字段，并在 video_api.py 中传递这些参数；3) 核心逻辑层：修改 python/sglang/multimodal_gen/runtime/pipelines_core/stages/input_validation.py 的 preprocess_condition_image 函数，新增处理用户指定 width/height 的逻辑：基于条件图像的长宽比计算目标面积（受 max_area 限制），并确保输出尺寸对齐到 mod 值；4) 测试层：新增单元测试文件 python/sglang/multimodal_gen/test/unit/test_input_validation.py 验证分辨率计算逻辑，并更新性能基准文件 perf_baselines.json 以反映分辨率变化后的性能数据。

关键文件：

- python/sglang/multimodal_gen/configs/pipeline_configs/wan.py（模块 multimodal_gen/configs）：修改 Wan2_2_I2V_A14B_Config 的继承关系，从 480P 改为 720P，直接影响模型的最大支持分辨率，是修复分辨率过小问题的关键配置变更。

- python/sglang/multimodal_gen/runtime/pipelines_core/stages/input_validation.py（模块 multimodal_gen/runtime/stages）：核心逻辑变更，在 preprocess_condition_image 函数中新增处理用户指定 width/height 的逻辑，计算输出视频尺寸，是分辨率控制的核心实现。
- python/sglang/multimodal_gen/test/unit/test_input_validation.py（模块 multimodal_gen/test）：新增单元测试，验证分辨率计算逻辑的正确性，包括长宽比处理、面积限制和 mod 对齐，确保变更的可靠性和回归防护。
- python/sglang/multimodal_gen/runtime/entrypoints/openai/protocol.py（模块 multimodal_gen/runtime/entrypoints）：在 VideoGenerationsRequest 中添加 width 和 height 字段，扩展 API 以支持用户输入，是实现功能的基础接口变更。

关键符号：preprocess_condition_image, _calculate_dimensions_from_area

评论区精华

Review 中的核心讨论包括：1) gemini-code-assist[bot] 指出 input_validation.py 中初始逻辑只处理同时提供 width 和 height 的情况，忽略单个维度输入，建议基于图像长宽比计算缺失维度，此问题在后续提交中修正；2) mickqian 询问是否需要在 VideoGenerationsRequest 中添加 width 和 height 字段，因为已有 size 字段，yeahdongcn 解释是为了与图像生成 API 保持一致，最终决定保留字段并确保 width/height 覆盖 size；3) 在 Issue 评论中，mickqian 要求添加测试用例，作者响应并新增了单元测试，同时更新了性能基准阈值。

- 处理用户指定 width/height 的逻辑修正 (correctness): 在后续提交中修正了逻辑，添加了处理单个维度的代码，确保用户输入被尊重。
- 是否在 VideoGenerationsRequest 中添加 width 和 height 字段 (design): 保留 width 和 height 字段，并确保它们覆盖 size 字段，以保持 API 一致性。

风险与影响

- 风险：技术风险包括：1) 核心路径变更风险 – input_validation.py 中的 preprocess_condition_image 函数修改了视频分辨率计算逻辑，影响所有使用 Wan2.2-I2V-A14B 模型的视频生成任务，若逻辑错误可能导致输出尺寸异常；2) 兼容性风险 – 新增 width 和 height 字段可能与现有 size 字段产生冲突，但 review 中已明确 width/height 应覆盖 size，需确保 API 文档更新；3) 性能风险 – 由于最大分辨率从 480P 提升到 720P，视频生成时间可能增加，perf_baselines.json 的更新显示了 InputValidationStage 和 DenoisingStage 耗时增长，需监控实际部署中的资源消耗；4) 测试覆盖风险 – 新增单元测试覆盖了基本场景，但可能未覆盖所有边界条件，如无效输入或极端长宽比。
- 影响：对用户的影响：用户现在可以通过指定 width 和 height 参数更精确地控制视频输出分辨率，提升使用体验和灵活性，尤其对于需要与图像生成保持 API 一致的客户；对系统的影响：视频生成管道现在支持额外的输入参数，可能轻微增加请求处理开销，但增强了功能一致性；对团队的影响：代码库在 multimodal_gen 模块中更统一，为后续扩展视频生成功能提供了基础，同时通过添加测试提升了代码质量。
- 风险标记：核心路径变更，API 兼容性，性能影响

关联脉络

- PR #21746 [diffusion] Fix typo: 同属 diffusion 模块，涉及代码修复，展示团队对 diffusion 模块的持续维护和代码质量关注。
- PR #21664 [diffusion] Fix Flux.2: 同属 diffusion 模块的 bugfix，关联类似的分辨率或权重加载问题，反映团队在扩散模型方面的集中改进。