

说明书

发明名称: 一种大模型交互的智能图片或视频的生成方法和系统
技术领域

[0001] 本发明涉及大模型交互领域，特别涉及一种大模型交互的智能图片或视频的生成方法和系统。

背景技术

[0002] 目前人工智能的多模态大模型的发展迅速，例如stableVideo 、 openAI sora等模型已经实现了文本到图片或视频的生成功能。

[0003] 现有人工智能的多模态大模型的文本到图片或视频的生成功能流程是：用户提出prompt请求，进行图像或画面的语言描述提供给大模型，大模型即可返回生成的图像或视频。

[0004] 由于prompt请求的功能以文本/字符串模板为主，其中包含图片或视频的画面主题、主体描述、画面背景与环境描述、画风定义和模型参数等。

[0005] 如图1所示，为一种大模型根据prompt请求实现从文本到图片的生成功能时，用户提供如图1左侧的prompt请求，则大模型生成了如图1右侧所示的图片。

[0006] 如图2所示，为一种大模型根据prompt请求实现从文本到视频的生成功能，用户提供的prompt请求，则大模型生成了如图3所示的视频。

[0007] 但是，目前的人工智能的多模态大模型在实现文本到图片或视频的生成功能时，存在下述几个缺点：

[0008] 1、生成页面的交互方式单一，页面交互只适合单条、逐条输入的文生图、文生视频交互；画面的各维度信息包含在prompt内；新出的stableVideo将相机的状态配置放在二级生成初稿页面进行选择，导致各个画面的prompt请求缺少系统化管理；

[0009] 2、当prompt请求涉及到大段文字描述时，用户可能缺少描述维度，影响大模型对图片或视频的生成质量；

[0010] 3、大模型返回生成的内容时，播放交互方式单一，通常仅提供的图片和视频的浏览器，仅支持播放、缩放、倍速播放、跳跃播放等播放交互方式；

[0011] 4、不支持批量化生成，效率低下。

发明概述

技术问题

[0012] 本发明要解决的技术问题，在于提供一种大模型交互的智能图片或视频的生成方法和系统，为用户构建提供系统化分镜模板，从而实现系统化、结构化编辑与管理，同时可支持批量化生成，交互方式丰富多样，提升用户交互体验。

技术方案

[0013] 第一方面，本发明提供了一种大模型交互的智能图片或视频的生成方法，包括图片或视频的生成过程，所述图片或视频的生成过程包括下述步骤：

[0014] S11、接收用户的原始prompt请求，抽取所述原始prompt请求中的描述信息并填充到系统内部的分镜模板中；或

[0015] 直接导入原始prompt请求和系统外部的分镜模板，将系统外部的分镜模板映射到系统内部的分镜模板中；

[0016] S12、序列化、文字化所述分镜模板中的内容，得到至少一个prompt请求分段，每个prompt请求分段对应一图片或分镜视频；

[0017] S13、将prompt请求分段逐条或批量提交给大模型，由大模型生成图片或视频内容后返回。

[0018] 第二方面，本发明提供了一种大模型交互的智能图片或视频的生成系统，包括模板管理模块，所述模板管理模块用于执行下述步骤：

[0019] S11、接收用户的原始prompt请求，抽取所述原始prompt请求中的描述信息并填充到系统内部的分镜模板中；或

[0020] 直接导入原始prompt请求和系统外部的分镜模板，将系统外部的分镜模板映射到系统内部的分镜模板中；

[0021] S12、序列化、文字化所述分镜模板中的内容，得到至少一个prompt请求分段，每个prompt请求分段对应一图片或分镜视频；

[0022] S13、将prompt请求分段逐条或批量提交给大模型，由大模型生成图片或视频内容。

有益效果

- [0023] 本发明提供的一个或多个技术方案，至少具有如下技术效果或优点：为用户构建系统化的分镜模板，可自动抽取prompt请求的实体配置，导入、编辑配置生成分镜模板，或直接导入系统外部的分镜模板，从而实现prompt请求的系统化、结构化编辑与管理，提升生成页面的交互体验；同时可支持prompt请求分段的单个生成或批量化生成，大大提高了图片和视频的生成效率；基于结构化分镜模板的信息动态编辑功能，用户可随时补充描述维度，保证大模型对图片或视频的生成质量；且在生成结果播放过程中，可针对用户的播放交互输入，实时修改或新增所述分镜模板上的对应参数，并提交给大模型，实现播放动态交互，大大提升了播放交互体验。
- [0024] 上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

附图说明

- [0025] 下面参照附图结合实施例对本发明作进一步的说明。
- [0026] 图1为一种现有大模型根据prompt请求实现从文本到图片的生成功能的交互页面示意图。
- [0027] 图2为一种现有大模型根据prompt请求实现从文本到视频的生成功能的交互页面示意图。
- [0028] 图3为一种现有大模型根据prompt请求生成的视频的其中一帧的图示。
- [0029] 图4为本发明实施例一中方法中的流程图。
- [0030] 图5为本发明一实施例中系统内部的分镜模板的结构示意图。
- [0031] 图6为本发明一实施例中系统外部的分镜模板的结构示意图。
- [0032] 图7为本发明实施例二中方法中的流程图。
- [0033] 图8为本发明实施例三中的系统框图。
- [0034] 图9为本发明一实施例中播放动态交互过程的流程图。

本发明的实施方式

- [0035] 本申请实施例通过提供一种大模型交互的智能图片或视频的生成方法和系统，为用户构建提供系统化分镜模板，从而实现系统化、结构化编辑与管理，同时可支持批量化生成，交互方式丰富多样，提升用户交互体验。
- [0036] 本申请实施例中的技术方案，总体思路如下：为用户构建系统化的分镜模板，自动抽取prompt请求的实体配置，导入、编辑配置生成分镜模板，或直接导入系统外部的分镜模板，从而实现prompt请求的系统化、结构化编辑与管理，提升生成页面的交互体验；同时可支持prompt请求分段的单个生成或批量化生成，大大提高了图片和视频的生成效率；基于结构化分镜模板的信息动态编辑功能，用户可随时补充描述维度，保证大模型对图片或视频的生成质量；且在生成结果播放过程中，可针对用户的播放交互输入，实时修改或新增所述分镜模板上的对应参数，并提交给大模型，实现播放动态交互，大大提升了播放交互体验。

实施例一

- [0037] 如图4所示，本实施例提供一种大模型交互的智能图片或视频的生成方法，包括图片或视频的生成过程，所述图片或视频的生成过程包括下述步骤：
- [0038] S11、接收用户的原始prompt请求，抽取所述原始prompt请求中的描述信息并填充到系统内部的分镜模板中；或
- [0039] 直接导入原始prompt请求和系统外部的分镜模板，将系统外部的分镜模板映射到系统内部的分镜模板中；所述系统外部的分镜模板为文本文件，表格文件，或json、yaml的markdown配置文件。
- [0040] 如图5和图6所示，所述分镜模板以表格的形式展示于生成交互页面中，且处于可二次编辑状态；
- [0041] 所述分镜模板的内容包括场景编号、镜头编号、画面参考、对象、景别、机位编号、拍摄角度、运镜方式、设备、镜头长度、声音、画面描述、剪辑与过渡方式、灯光、生成次数编号、拍摄时间；从而涵盖了创意策划、画面设计、拍摄制作、后期剪辑对于画面理解沟通的工具，能够更准确传递视觉策划、制片或导演需求，加快沟通、制作效率。
- [0042] S12、序列化、文字化所述分镜模板中的内容，得到至少一个prompt请求分段，每个prompt请求分段对应一图片或分镜视频；

- [0043] 其中，所述prompt请求分段可被对应地填充在所述分镜模板中。
- [0044] S13、将prompt请求分段逐条或批量提交给大模型，即对应每个prompt请求分段设置一个单条执行的触发按键以及在整个分镜模板处设置一个批量执行的触发按键，在触发执行后，即可由大模型生成图片或视频内容后返回。
- [0045] 如图4所示，通常视频播放包括正常播放、缩放播放、倍速播放、跳跃播放等常规播放方式，而本实施例的生成方法，还可包括播放动态交互过程：
- [0046] S21、在生成视频结果播放时，实时接收用户的播放交互输入；
- [0047] S22、基于预设的播放交互规则或播放交互用例，根据所述用户的播放交互输入，基于时间戳，查找对应的视频参照帧和描述文本，实时修改或新增所述分镜模板上的对应参数，并提交给大模型；
- [0048] S23、大模型实时重新生成图片或视频内容后返回。
- [0049] 如图9所示，播放动态交互过程可以让用户和视频中的序列帧进行交互，通过播放交互输入，基于时间戳、索引对应的视频参照帧和描述文本，实时修改或新增所述分镜模板上的对应参数，实现画面内容交互，如基于鼠标交互模拟的视频画面镜头动作，或者是键盘快捷键对应的画面特效补充生成等，从而视频播放交互从静态序列帧播放交互拓展到生成内容内的沉浸式交互。
- [0050] 其中，用户的播放交互输入包括但不限于键盘输入、鼠标输入、音频输入、动作输入、视频输入等；所述播放交互规则或所述播放交互用例由系统提供或由用户自行定义。
- [0051] 所述播放交互规则以鼠标操作为例：
- [0052] 视频播放时，左键点击向左/右拖拽，运镜参数修改为推左/推右；
- [0053] 视频播放时，右键点击向左/右拖拽，运镜参数修改为摇左/摇右；右键点击向前/后拖拽，摇上/摇下；
- [0054] 视频播放时，滚轮滚动操作，为变焦拉近；滚轮点击向前/向后，前进/后退。
- [0055] 所述播放交互用例如可以是屏幕下雨或飘雪等。

实施例二

- [0056] 如图7所示，本实施例二提供一种大模型交互的智能图片或视频的生成方法，其是在实施例一的基础上，还提供生成建议prompt的功能，即在步骤S12中序列

化、文字化所述分镜模板中的内容，得到至少一个prompt请求分段后，每个所述prompt请求分段还能根据用户的触发生成建议prompt，并将所述建议prompt代替所述prompt请求分段提交给大模型。从而可弥补用户的prompt请求分段的不足，进一步提升大模型对图片或视频的生成质量。

- [0057] 其中，用户的触发可设在生成交互页面中的分镜模板对应每个所述prompt请求分段的位置。
- [0058] 本实施例二还提供将prompt请求分段逐条或批量提交给大模型的选择机会，即对应每个prompt请求分段设置一个单条执行的触发按键以及在整个分镜模板处设置一个批量执行的触发按键，在用户触发执行后，即可由大模型生成图片或视频内容后返回。

实施例三

- [0059] 如图8所示，在本实施例中提供了一种大模型交互的智能图片或视频的生成系统，是与实施例一中的方法对应的装置，包括模板管理模块，还可包括播放动态交互模块。
- [0060] 如图4所示，所述模板管理模块用于执行下述步骤：
- [0061] S11、接收用户的原始prompt请求，抽取所述原始prompt请求中的描述信息并填充到系统内部的分镜模板中；或
- [0062] 直接导入原始prompt请求和系统外部的分镜模板，将系统外部的分镜模板映射到系统内部的分镜模板中；所述系统外部的分镜模板为文本文件，表格文件，或json、yaml的markdown配置文件。
- [0063] S12、序列化、文字化所述分镜模板中的内容，得到至少一个prompt请求分段，每个prompt请求分段对应一图片或分镜视频；
- [0064] S13、将prompt请求分段逐条或批量提交给大模型，由大模型生成图片或视频内容。
- [0065] 本实施例二还提供将prompt请求分段逐条或批量提交给大模型的选择机会，即对应每个prompt请求分段设置一个单条执行的触发按键以及在整个分镜模板处设置一个批量执行的触发按键，在用户触发执行后，即可由大模型生成图片或视频内容后返回。

- [0066] 如图5和图6所示,所述分镜模板以表格的形式展示于生成交互页面中,且处于可二次编辑状态;
- [0067] 所述分镜模板的内容包括场景编号、镜头编号、画面参考、对象、景别、机位编号、拍摄角度、运镜方式、设备、镜头长度、声音、画面描述、剪辑与过渡方式、灯光、生成次数编号、拍摄时间;
- [0068] 所述模板管理模块还用于:
- [0069] 将所述prompt请求分段对应地填充在所述分镜模板中;
- [0070] 根据用户的触发将对应的所述prompt请求分段生成建议prompt用于代替所述prompt请求分段提交给大模型。
- [0071] 如图4所示,所述播放动态交互模块用于执行下述过程:
- [0072] S21、在生成视频结果播放时,实时接收用户的播放交互输入;
- [0073] S22、基于预设的播放交互规则或播放交互用例,根据所述用户的播放交互输入,基于时间戳,查找对应的视频参照帧和描述文本,实时修改或新增所述分镜模板上的对应参数,并提交给大模型;
- [0074] S23、大模型实时重新生成图片或视频内容后返回。
- [0075] 所述播放交互规则或所述播放交互用例由系统提供或由用户自行定义。
- [0076] 如图9所示,播放动态交互过程可以让用户和视频中的序列帧进行交互,通过播放交互输入,基于时间戳、索引对应的视频参照帧和描述文本,实时修改或新增所述分镜模板上的对应参数,实现画面内容交互,如基于鼠标交互模拟的视频画面镜头动作,或者是键盘快捷键对应的画面特效补充生成等,从而视频播放交互从静态序列帧播放交互拓展到生成内容内的沉浸式交互。
- [0077] 其中,用户的播放交互输入包括但不限于键盘输入、鼠标输入、音频输入、动作输入、视频输入等;所述播放交互规则或所述播放交互用例由系统提供或由用户自行定义。
- [0078] 所述播放交互规则以鼠标操作为例:
- [0079] 视频播放时,左键点击向左/右拖拽,运镜参数修改为推左/推右;
- [0080] 视频播放时,右键点击向左/右拖拽,运镜参数修改为摇左/摇右;右键点击向前/后拖拽,摇上/摇下;

[0081] 视频播放时，滚轮滚动操作，为变焦拉近；滚轮点击向前/向后，前进/后退。

[0082] 所述播放交互用例如可以是屏幕下雨或飘雪等。

[0083] 综上所述，本申请实施例中提供的技术方案，至少具有如下技术效果或优点：
为用户构建系统化的分镜模板，自动抽取prompt请求的实体配置，导入、编辑配置生成分镜模板，或直接导入系统外部的分镜模板，从而实现prompt请求的系统化、结构化编辑与管理，提升生成页面的交互体验；同时可支持prompt请求分段的单个生成或批量化生成，大大提高了图片和视频的生成效率；基于结构化分镜模板的信息动态编辑功能，用户可随时补充描述维度，保证大模型对图片或视频的生成质量；且在生成结果播放过程中，可针对用户的播放交互输入，实时修改或新增所述分镜模板上的对应参数，并提交给大模型，实现播放动态交互，大大提升了播放交互体验。

[0084] 虽然以上描述了本发明的具体实施方式，但是熟悉本技术领域的技术人员应当理解，我们所描述的具体的实施例只是说明性的，而不是用于对本发明的范围的限定，熟悉本领域的技术人员在依照本发明的精神所作的等效的修饰以及变化，都应当涵盖在本发明的权利要求所保护的范围内。

权利要求书

- [权利要求 1]

一种大模型交互的智能图片或视频的生成方法，其特征在于：包括图片或视频的生成过程，所述图片或视频的生成过程包括下述步骤：

S11、接收用户的原始prompt请求，抽取所述原始prompt请求中的描述信息并填充到系统内部的分镜模板中；或

直接导入原始prompt请求和系统外部的分镜模板，将系统外部的分镜模板映射到系统内部的分镜模板中；

S12、序列化、文字化所述分镜模板中的内容，得到至少一个prompt请求分段，每个prompt请求分段对应一图片或分镜视频；

S13、将prompt请求分段逐条或批量提交给大模型，由大模型生成图片或视频内容后返回。
- [权利要求 2]

根据权利要求1所述的一种大模型交互的智能图片或视频的生成方法，其特征在于：

所述分镜模板以表格的形式展示于生成交互页面中，且处于可二次编辑状态；

所述分镜模板的内容包括场景编号、镜头编号、画面参考、对象、景别、机位编号、拍摄角度、运镜方式、设备、镜头长度、声音、画面描述、剪辑与过渡方式、灯光、生成次数编号、拍摄时间；

所述系统外部的分镜模板为文本文件，表格文件，或json、yaml的markdown配置文件。
- [权利要求 3]

根据权利要求1所述的一种大模型交互的智能图片或视频的生成方法，其特征在于：所述prompt请求分段被对应地填充在所述分镜模板中，每个所述prompt请求分段还能根据用户的触发生成建议prompt，所述建议prompt代替所述prompt请求分段提交给大模型。
- [权利要求 4]

根据权利要求1所述的一种大模型交互的智能图片或视频的生成方法，其特征在于：还包括播放动态交互过程：

S21、在生成视频结果播放时，实时接收用户的播放交互输入；

- S22、基于预设的播放交互规则或播放交互用例，根据所述用户的播放交互输入，基于时间戳，查找对应的视频参照帧和描述文本，实时修改或新增所述分镜模板上的对应参数，并提交给大模型；
- S23、大模型实时重新生成图片或视频内容后返回。
- [权利要求 5] 根据权利要求4所述的一种大模型交互的智能图片或视频的生成方法，其特征在于：所述播放交互规则或所述播放交互用例由系统提供或由用户自行定义。
- [权利要求 6] 一种大模型交互的智能图片或视频的生成系统，其特征在于：包括模板管理模块，所述模板管理模块用于执行下述步骤：
- S11、接收用户的原始prompt请求，抽取所述原始prompt请求中的描述信息并填充到系统内部的分镜模板中；或
直接导入原始prompt请求和系统外部的分镜模板，将系统外部的分镜模板映射到系统内部的分镜模板中；
- S12、序列化、文字化所述分镜模板中的内容，得到至少一个prompt请求分段，每个prompt请求分段对应一图片或分镜视频；
- S13、将prompt请求分段逐条或批量提交给大模型，由大模型生成图片或视频内容。
- [权利要求 7] 根据权利要求6所述的一种大模型交互的智能图片或视频的生成系统，其特征在于：
- 所述分镜模板以表格的形式展示于生成交互页面中，且处于可二次编辑状态；
- 所述分镜模板的内容包括场景编号、镜头编号、画面参考、对象、景别、机位编号、拍摄角度、运镜方式、设备、镜头长度、声音、画面描述、剪辑与过渡方式、灯光、生成次数编号、拍摄时间；
- 所述系统外部的分镜模板为文本文件，表格文件，或json、yaml的markdown配置文件。
- [权利要求 8] 根据权利要求6所述的一种大模型交互的智能图片或视频的生成系统，其特征在于：所述模板管理模块还用于：

将所述prompt请求分段对应地填充在所述分镜模板中；
根据用户的触发将对应的所述prompt请求分段生成建议prompt用于代替所述prompt请求分段提交给大模型。

- [权利要求 9]

根据权利要求6所述的一种大模型交互的智能图片或视频的生成系统，其特征在于：还包括播放动态交互模块，所述播放动态交互模块用于执行下述过程：

S21、在生成视频结果播放时，实时接收用户的播放交互输入；

S22、基于预设的播放交互规则或播放交互用例，根据所述用户的播放交互输入，基于时间戳，查找对应的视频参照帧和描述文本，实时修改或新增所述分镜模板上的对应参数，并提交给大模型；

S23、大模型实时重新生成图片或视频内容后返回。
- [权利要求 10]

根据权利要求9所述的一种大模型交互的智能图片或视频的生成系统，其特征在于：所述播放交互规则或所述播放交互用例由系统提供或由用户自行定义。