



(12) 发明专利

(10) 授权公告号 CN 117745857 B

(45) 授权公告日 2025. 04. 08

(21) 申请号 202311755560.3

G06T 11/00 (2006.01)

(22) 申请日 2023.12.20

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 105874449 A, 2016.08.17

申请公布号 CN 117745857 A

CN 116188632 A, 2023.05.30

(43) 申请公布日 2024.03.22

审查员 庞洵

(73) 专利权人 北京百度网讯科技有限公司

地址 100085 北京市海淀区上地十街10号

百度大厦2层

(72) 发明人 戎康 宋雨鑫 张琦 刘芳龙

(74) 专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

专利代理师 王达佐 王艳春

(51) Int. Cl.

G06N 3/08 (2023.01)

G06N 3/0475 (2023.01)

权利要求书4页 说明书14页 附图4页

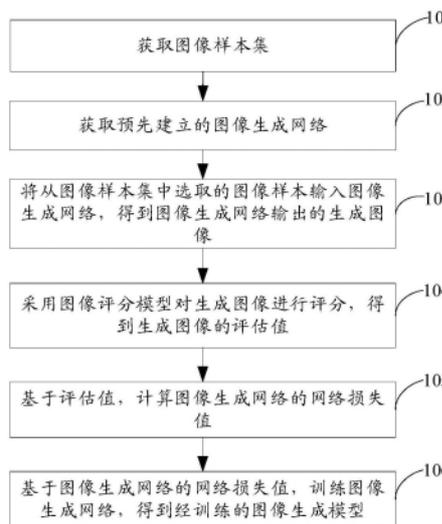
(54) 发明名称

图像生成模型训练方法和装置、图像处理方法和装置

(57) 摘要

本公开提供了一种图像生成模型训练方法和装置,涉及人工智能技术领域,具体涉及计算机视觉、深度学习、大模型等技术领域,可应用于人工智能的内容生成等场景。具体实现方案为:获取图像样本集;获取预先构建的图像生成网络,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型;将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像;采用图像评分模型对生成图像进行评分,得到生成图像的评估值;基于评估值,计算图像生成网络的网络损失值;基于评估值,计算图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型。

100



1. 一种图像生成模型训练方法,所述方法包括:

获取图像样本集,所述图像样本集包括至少一个图像样本;

获取预先构建的图像生成网络,所述图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,所述图文识别模块基于输入的图像,得到识别文本;所述大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;所述文本生成图像模型基于所述提示词文本,得到生成图像;

将从所述图像样本集中选取的图像样本输入所述图像生成网络,得到所述图像生成网络输出的生成图像;采用图像评分模型对所述生成图像进行评分,得到所述生成图像的评估值;基于所述评估值,计算所述图像生成网络的网络损失值;基于所述图像生成网络的网络损失值,训练所述图像生成网络,得到经训练的图像生成模型;

其中,所述图像评分模型通过训练多模态图文识别网络得到,所述多模态图文识别网络的输入包括第一图像样本、输出包括第一评分,所述第一图像样本包括由所述图像生成模型生成的第一图像和由人工对所述第一图像进行评分得到的所述第一图像的评分,以及所述第一图像的评分、所述第一评分和所述多模态图文识别网络的损失函数用于计算得到所述多模态图文识别网络的损失值。

2. 根据权利要求1所述的方法,其中,所述基于所述评估值,计算所述图像生成网络的网络损失值包括:

获取所述文本生成图像模型的损失函数;

基于选取的图像样本和所述文本生成图像模型的损失函数,计算所述文本生成图像模型的模型损失值;

基于所述评估值,调整所述模型损失值,得到所述网络损失值。

3. 根据权利要求1或2所述的方法,其中,所述基于所述图像生成网络的网络损失值,训练所述图像生成网络,得到经训练的图像生成模型,包括:

响应于所述图像生成网络的网络损失值满足训练完成条件,将所述图像生成网络作为图像生成模型。

4. 根据权利要求3所述的方法,其中,所述方法还包括:

响应于所述图像生成网络的网络损失值不满足训练完成条件,基于所述网络损失值,调整所述大语言模型的参数,并继续训练所述图像生成网络。

5. 根据权利要求1所述的方法,其中,所述多图像描述特征包括:元素特征、构图特征以及风格特征。

6. 根据权利要求1所述的方法,其中,所述多模态图文识别网络的训练步骤包括:

执行第一训练步骤:

将所述第一图像样本和获取的描述文本输入所述多模态图文识别网络,得到所述多模态图文识别网络输出的答案文本;

拼接所述第一图像样本、所述描述文本、所述答案文本以及评分文本,得到第一拼接信息,并将所述第一拼接信息输入所述多模态图文识别网络,得到所述第一评分;基于所述多模态图文识别网络的损失值,得到第一评分模型;

执行第二训练步骤:

将获取的第二图像样本、所述评分文本输入所述第一评分模型,得到所述第一评分模

型输出的第二评分；

基于所述第二评分,计算所述第一评分模型的损失值；

基于所述第一评分模型的损失值,得到图像评分模型。

7. 一种图像处理方法,所述方法包括:

获取待处理图像；

将所述待处理图像输入采用如权利要求1-6中任一项所述的方法生成的图像生成模型中,得到所述待处理图像的图像生成结果。

8. 根据权利要求7所述的方法,在将所述待处理图像输入图像生成模型之前,所述方法还包括:

检测所述待处理图像的尺寸是否为标准尺寸；

响应于所述待处理图像的尺寸不是标准尺寸,将所述待处理图像调整到标准尺寸。

9. 根据权利要求7所述的方法,所述图像生成模型包括:图文识别模块、大语言模型以及文本生成图像模型,所述将所述待处理图像输入采用如权利要求1-6中任一项所述的方法生成的图像生成模型中,输出所述待处理图像的图像生成结果包括:

将所述待处理图像输入所述图文识别模块,得到所述图文识别模块输出的识别文本；

将所述识别文本输入所述大语言模型,得到所述大语言模型输出的提示词文本；

将所述提示词文本输入所述文本生成图像模型,得到所述文本生成图像模型输出的生成图像。

10. 根据权利要求9所述的方法,所述方法还包括:

接收图像处理需求文本;其中,所述需求文本为对待生成的生成图像的需求；

在得到所述识别文本之后,将所述识别文本与所述图像处理需求文本进行拼接,得到第二拼接信息；

将所述第二拼接信息输入所述大语言模型,得到所述大语言模型输出的新的提示词文本；

将所述新的提示词文本输入所述文本生成图像模型,得到所述文本生成图像模型输出的新的生成图像。

11. 一种图像生成模型训练装置,所述装置包括:

集合获取单元,被配置成获取图像样本集,所述图像样本集包括至少一个图像样本；

网络获取单元,被配置成获取预先构建的图像生成网络,所述图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,所述图文识别模块基于输入的图像,得到识别文本;所述大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;所述文本生成图像模型基于所述提示词文本,得到生成图像；

样本输入单元,被配置成将从所述图像样本集中选取的图像样本输入所述图像生成网络,得到所述图像生成网络输出的生成图像；

评分单元,被配置成采用图像评分模型对所述生成图像进行评分,得到所述生成图像的评估值；

计算单元,被配置成基于所述评估值,计算所述图像生成网络的网络损失值；

模型得到单元,被配置成基于所述图像生成网络的网络损失值,训练所述图像生成网络,得到经训练的图像生成模型；

其中,所述图像评分模型通过训练多模态图文识别网络得到,所述多模态图文识别网络的输入包括第一图像样本、输出包括第一评分,所述第一图像样本包括由所述图像生成模型生成的第一图像和由人工对所述第一图像进行评分得到的所述第一图像的评分,以及所述第一图像的评分、所述第一评分和所述多模态图文识别网络的损失函数用于计算得到所述多模态图文识别网络的损失值。

12.根据权利要求11所述的装置,其中,所述计算单元进一步被配置成:获取所述文本生成图像模型的损失函数;基于选取的图像样本和所述文本生成图像模型的损失函数,计算所述文本生成图像模型的模型损失值;基于所述评估值,调整所述模型损失值,得到所述网络损失值。

13.根据权利要求11或12所述的装置,其中,所述模型得到单元进一步被配置成:响应于所述图像生成网络的网络损失值满足训练完成条件,将所述图像生成网络作为图像生成模型。

14.根据权利要求13所述的装置,其中,所述装置还包括:调整单元,被配置成响应于所述图像生成网络的网络损失值不满足训练完成条件,基于所述网络损失值,调整所述大语言模型的参数,并控制所述样本输入单元工作。

15.根据权利要求11所述的装置,其中,所述多图像描述特征包括:元素特征、构图特征以及风格特征。

16.根据权利要求11所述的装置,其中,所述多模态图文识别网络通过训练单元训练得到,所述训练单元被配置成:将所述第一图像样本和获取的描述文本输入所述多模态图文识别网络,得到所述多模态图文识别网络输出的答案文本;拼接所述第一图像样本、所述描述文本、所述答案文本以及评分文本,得到第一拼接信息,并将所述第一拼接信息输入所述多模态图文识别网络,得到所述第一评分;基于所述多模态图文识别网络的损失值,得到所述第一评分模型;将获取的第二图像样本、所述评分文本输入所述第一评分模型,得到所述第一评分模型输出的第二评分;基于所述第二评分,计算所述第一评分模型的损失值;基于所述第一评分模型的损失值,得到图像评分模型。

17.一种图像处理装置,所述装置包括:

图像获取单元,被配置成获取待处理图像;

结果得到单元,被配置成将所述待处理图像输入采用如权利要求11-16中任一项所述的装置生成的图像生成模型中,输出所述待处理图像的图像生成结果。

18.根据权利要求17所述的装置,所述装置还包括:检测单元,所述检测单元被配置成:检测所述待处理图像的尺寸是否为标准尺寸;响应于所述待处理图像的尺寸不是标准尺寸,将所述待处理图像调整到标准尺寸。

19.根据权利要求17所述的装置,其中,所述图像生成模型包括:图文识别模块、大语言模型以及文本生成图像模型,所述结果得到单元进一步被配置成:将所述待处理图像输入所述图文识别模块,得到所述图文识别模块输出的识别文本;将所述识别文本输入所述大语言模型,得到所述大语言模型输出的提示词文本;将所述提示词文本输入所述文本生成图像模型,得到所述文本生成图像模型输出的生成图像。

20.根据权利要求19所述的装置,所述装置还包括:

接收单元,被配置成接收图像处理需求文本;

文本输入单元,被配置成在得到所述识别文本之后,将所述识别文本与所述图像处理需求文本进行拼接,得到第二拼接信息;将所述第二拼接信息输入所述大语言模型,得到所述大语言模型输出的新的提示词文本;将所述新的提示词文本输入所述文本生成图像模型,得到所述文本生成图像模型输出的新的生成图像。

21.一种电子设备,其特征在于,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-10中任一项所述的方法。

22.一种存储有计算机指令的非瞬时计算机可读存储介质,其特征在于,所述计算机指令用于使所述计算机执行权利要求1-10中任一项所述方法。

23.一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现权利要求1-10中任一项所述的方法。

图像生成模型训练方法和装置、图像处理方法和装置

技术领域

[0001] 本公开涉及人工智能技术领域,具体涉及计算机视觉、深度学习、大模型等技术领域,可应用于人工智能的内容生成等场景,尤其涉及一种图像生成模型训练方法和装置、图像处理方法和装置、电子设备、计算机可读存储介质以及计算机程序产品。

背景技术

[0002] 随着SD(stablediffusion,稳定扩散)模型的提出,生成式图像模型已经展现出强大的能力,包括更加真实的场景,更加丰富的细节以及良好的指令跟随能力。

[0003] 对于广泛应用的深度学习文本到图像生成模型,将提示词输入深度学习文本到图像生成模型,几乎可以生成任何一种人类想象出的图像。

发明内容

[0004] 本公开提供了一种图像生成模型训练方法和装置、图像处理方法和装置、电子设备、计算机可读存储介质以及计算机程序产品。

[0005] 根据第一方面,提供了一种图像生成模型训练方法,该方法包括:获取图像样本集,图像样本集包括至少一个图像样本;获取预先构建的图像生成网络,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,图文识别模块基于输入的图像,得到识别文本;大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;文本生成图像模型基于提示词文本,得到生成图像;执行以下总训练步骤:将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像;采用图像评分模型对生成图像进行评分,得到生成图像的评估值;基于评估值,计算图像生成网络的网络损失值;基于图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型。

[0006] 根据第二方面,提供了一种图像处理方法,该方法包括:获取待处理图像;将待处理图像输入采用如第一方面任一实现方式描述的方法生成的图像生成模型中,得到待处理图像的图像生成结果。

[0007] 根据第三方面,提供了一种图像生成模型训练装置,该装置包括:集合获取单元,被配置成获取图像样本集,图像样本集包括至少一个图像样本;网络获取单元,被配置成获取预先构建的图像生成网络,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,图文识别模块基于输入的图像,得到识别文本;大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;文本生成图像模型基于提示词文本,得到生成图像;样本输入单元,被配置成将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像;评分单元,被配置成采用图像评分模型对生成图像进行评分,得到生成图像的评估值;计算单元,被配置成基于评估值,计算图像生成网络的网络损失值;模型得到单元,被配置成基于图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型。

[0008] 根据第四方面,又提供了一种图像处理装置,该装置包括:图像获取单元,被配置成获取待处理图像;结果得到单元,被配置成将待处理图像输入采用如第三方面任一实现方式描述的装置生成的图像生成模型中,输出待处理图像的图像生成结果。

[0009] 根据第五方面,提供了一种电子设备,该电子设备包括:至少一个处理器;以及与至少一个处理器通信连接的存储器,其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行如第一方面或第二方面任一实现方式描述的方法。

[0010] 根据第六方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,计算机指令用于使计算机执行如第一方面或第二方面任一实现方式描述的方法。

[0011] 根据第七方面,提供了一种计算机程序产品,包括计算机程序,计算机程序在被处理器执行时实现如第一方面或第二方面任一实现方式描述的方法。

[0012] 本公开的实施例提供的图像生成模型训练方法和装置,首先,获取图像样本集,图像样本集包括至少一个图像样本;其次,获取预先构建的图像生成网络,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,图文识别模块基于输入的图像,得到识别文本;大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;文本生成图像模型基于提示词文本,得到生成图像;再次,将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像;从次,采用图像评分模型对生成图像进行评分,得到生成图像的评估值;然后,基于评估值,计算图像生成网络的网络损失值;最后,基于图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型。本公开通过大语言模型对识别文本进行多图像描述特征的提示词文本的生成,提高了提示词对文本生成图像模型的指导性;在图像生成网络训练过程中,通过图像评分模型对生成图像进行评分,并基于评分得到的评估值得到图像生成模型,提高了图像生成模型训练的可靠性和准确性,提高了模型进行图像生成的效果。

[0013] 本公开的实施例提供的图像处理方法和装置,获取待处理图像,将待处理图像输入图像生成模型训练方法生成的图像生成模型中,得到待处理图像的图像生成结果。由此,采用包括大语言模型的图像生成模型生成图像生成结果,提高了图像生成结果得到的可靠性和准确性。

[0014] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0015] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0016] 图1是根据本公开图像生成模型训练方法的一个实施例的流程图;

[0017] 图2是本公开实施例中图像生成网络训练的一种结构示意图;

[0018] 图3是根据本公开图像处理方法的一个实施例的流程图;

[0019] 图4是本公开实施例中图像处理方法的另一个实施例的流程图;

[0020] 图5是根据本公开图像生成模型训练装置的一个实施例的结构示意图;

[0021] 图6是根据本公开图像处理装置的一个实施例的结构示意图;

[0022] 图7是用来实现本公开实施例的图像生成模型训练方法或图像处理方法的电子设

备的框图。

具体实施方式

[0023] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0024] 传统的风格图像生成过程包括:选择参考图片,使用原图片直接引导新的图片生成。将原始图片通过变换操作,得到图片内容的边缘分布图,大致的色值分布以及整体的构图等信息,一起输入文本到图像生成模型,由这些信息综合指导生成的图片。这张图片应该具有希望在生成图像中看到的风格或内容特征。

[0025] 现有方案需要将原始图片进行一系列预处理后输入原始生成模型,但是这些预处理并不一定能完整的体现图片的内容,尤其是图片风格等抽象特征,人文内容并不能很好的通过这种形式引导。在内容引导的场景中,模型可能无法精确地再现参考图片中的细节,导致生成的图像在内容上与原图有所偏差。或者过于跟随原有图片,没有更多的发散和延伸效果。生成结果极大地依赖于用户提供的文本提示和参考图像的质量。不精确或模糊的输入可能导致不理想的输出。

[0026] 基于此,本公开提出了一种图像生成模型训练方法,图1示出了根据本公开图像生成模型训练方法的一个实施例的流程100,上述图像生成模型训练方法包括以下步骤:

[0027] 步骤101,获取图像样本集。

[0028] 本实施例中,图像生成模型训练方法运行于其上的执行主体可以通过多种方式获取图像样本集,例如,执行主体可以通过有线连接方式或无线连接方式,从数据库服务器中获取存储于其中的图像样本集。再例如,用户可以通过与终端通信,获取终端所收集的图像样本集。

[0029] 在这里,图像样本集可以包括至少一个图像样本,图像样本中包括与待生成的生成图像相关的标注图像,标注图像中显示的内容包括各种对象、场景和风格;可选地,图像样本中还可以包括图像描述文本,在训练图像生成网络时,可以将图像样本中的图像输入图文识别模块,将图像描述文本与识别文本拼接,并将拼接后的文本输入大语言模型。

[0030] 本公开的技术方案中,所涉及的视频帧、视频序列的收集、存储、使用、加工、传输、提供和公开等处理,是在经授权后执行的,符合相关法律法规。

[0031] 步骤102,获取预先构建的图像生成网络。

[0032] 其中,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,图文识别模块基于输入的图像,得到识别文本;大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;文本生成图像模型基于提示词文本,得到生成图像。

[0033] 本实施例中,图文识别模块是一种将图像转化为文本的图像内容提取器,图文识别模块可以是多模态图文识别模型,将图像输入多模态图文识别模型,得到多模态图文识别模型输出的对图像中内容进行描述的识别文本。识别文本是以文本数据的形式表征图像的信息。多模态图文识别模型可以被训练为支持多种语言,从而跨越了语言障碍。

[0034] 本实施例中,大语言模型是一种基于深度学习的自然语言处理模型,其主要通过

学习大量文本数据,可以自动生成符合语言规则的语句、段落或者文章,大语言模型的核心思想就是利用深度神经网络学习自然语言的语法、语义等特征,从而能够预测下一个词汇的出现概率,并根据这些概率生成新的语句。

[0035] 本实施例中,将识别文本输入大语言模型,得到大语言模型输出的多图像描述特征的提示词文本,其中,图像描述特征是描述图像特征的文本,本公开的大语言模型得到的提示词文本中具有多个图像描述特征,通过多个图像描述特征可以更加全面的描述输入图文识别模块的输入的图像。

[0036] 本实施例中,文本生成图像模型是一种基于文本生成风格图像的模型,将提示词文本输入文本生成图像模型,可以向文本生成图像模块输入图像生成需求,从而为文本生成图像模块提供更多的目标图像的要求。

[0037] 本实施例中,文本生成图像模型可以是SD(stablediffusion,稳定扩散)模型,基于图文识别模块产生的识别文本比较详细和精准,同时通过大语言模型自动生成的提示词文本可能提供新的视角和创意,为文本生成图像模型的图像生成提供更多可能得空间,甚至激发模型使用者的创作灵感。

[0038] 步骤103,将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像。

[0039] 本实施例中,执行主体可以从步骤101中获取的图像样本集中选取图像样本,以及执行步骤103至步骤106的训练步骤,完成一次图像生成网络的迭代训练。其中,从图像样本集中选取视频帧的选取方式和选取数量在本申请中并不限制,并且图像生成网络的迭代训练的次数也并不限制。例如在一次迭代训练中可以是随机选取多个连续的图像集,该选取的图像集可以只有图像,也可以具有图像对应说明文本,通过选取的图像样本计算图像生成网络的网络损失值,调整图像生成网络的参数。

[0040] 步骤104,采用图像评分模型对生成图像进行评分,得到生成图像的评估值。

[0041] 本实施例中,图像评分模型是一种预先训练完成的对图像的内容、表现形式进行评分的模型,将图像输入图像评分模型,可以得到图像评分模型输出的评估值,而评估值则是表征图像内容的丰富程度、美感程度的具体表征。

[0042] 本实施例中,图像评分模型可以通过训练大模型得到的图像文本转换模型,大模型指具有大量参数的深度学习或机器学习模型,这些参数可以通过训练过程自动调整以捕获输入数据中的复杂关系。这类模型通常具有较深的网络结构和较多的神经元,以增加模型的表示能力和学习能力。

[0043] 具体如图2所示,将样本图像输入图文识别模块,图文识别模块输出识别文本,大语言模型基于识别文本,得到提示词文本,文本生成图像模型基于提示词文本,得到生成图像,图像评分模型获取生成图像,对生成图像进行评分,得到评估值,通过评估值对图像生成网络进行训练。

[0044] 步骤105,基于评估值,计算图像生成网络的网络损失值。

[0045] 本实施例中,图像生成网络的每次迭代训练时,均会从图像样本集中选取图像样本,并将选取的图像样本输入图像生成网络,基于预先为图像生成网络设置的损失函数和评估值,计算图像生成网络的网络损失值。

[0046] 本实施例中,图像生成网络的损失函数可以采用均方误差函数,均方误差函数是

图像生成网络的预测值(估计值)与真值差平方的期望,在图像生成网络迭代训练过程中,可以利用梯度下降算法最小化图像生成网络的损失函数,从而迭代地优化图像生成网络的网络参数。

[0047] 梯度的本意是一个向量,表示某一损失函数在该点处的方向导数沿着该方向取得最大值,即损失函数在该点处沿着该方向变化最快,变化率最大。在深度学习中,神经网络的主要任务是在学习时找到最优的网络参数(权重和偏置),这个最优的网络参数也就是损失函数最小时的参数。

[0048] 在图像生成网络训练过程,可以只为文本生成图像模型设计损失函数,通过文本生成图像模型的损失函数,计算损失值,基于损失值以及评估值,更新大语言模型的梯度,并调整大语言模型的参数,无需调整文本生成图像模型的参数,达到热插拔文本生成图像模型的目标。

[0049] 可选地,在图像生成网络训练过程中,还可以只为文本生成图像模型设计损失函数,通过文本生成图像模型的损失函数,计算损失值,基于损失值以及评估值,更新大语言模型以及文本生成图像模型的参数,达到文本生成图像模型与大语言模型均同时调整的目的。

[0050] 本实施例中,上述基于评估值,计算图像生成网络的网络损失值包括:计算图像生成网络的整体损失值;将整体损失值除以评估值,得到网络损失值。

[0051] 步骤106,基于图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型。

[0052] 本实施例中,图像生成模型是通过多次迭代训练,对图像生成网络进行调参后,得到的训练完成的图像生成网络,通过图像生成网络的网络损失值可以检测图像生成网络是否满足训练完成条件,在图像生成网络满足训练完成条件之后,得到图像生成模型。

[0053] 可选地,本实施例中,响应于图像生成网络不满足训练完成条件,则调整图像生成网络中的相关参数使得图像生成网络的网络损失值收敛,基于调整后的图像生成网络,继续执行上述训练步骤103-106。

[0054] 本可选实现方式中,在图像生成网络不满足训练完成条件时,调整图像生成网络的相关参数,有助于帮助图像生成网络的网络损失值收敛。

[0055] 本实施例提供的图像生成模型训练方法,可以自动化的生成提示词文本,大大减少了人工输入的时间和劳动,特别是对于大规模图像处理任务,提高了图像处理任务效率。并且自动生成的提示词文本可以提供标准化的图像描述,有助于统一沟通和搜索图像内容。

[0056] 本公开的实施例提供的图像生成模型训练方法,首先,获取图像样本集,图像样本集包括至少一个图像样本;其次,获取预先构建的图像生成网络,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,图文识别模块基于输入的图像,得到识别文本;大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;文本生成图像模型基于提示词文本,得到生成图像;再次,将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像;从次,采用图像评分模型对生成图像进行评分,得到生成图像的评估值;然后,基于评估值,计算图像生成网络的网络损失值;最后,基于图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型。本

公开通过大语言模型对识别文本进行多图像描述特征的提示词文本的生成,提高了提示词对文本生成图像模型的指导性;在图像生成网络训练过程中,通过图像评分模型对生成图像进行评分,并基于评分得到的评估值得到图像生成模型,提高了图像生成模型训练的可靠性和准确性。

[0057] 在本公开的一些可选实现方式中,上述基于评估值,计算图像生成网络的网络损失值包括:获取文本生成图像模型的损失函数;基于选取的图像样本和损失函数,计算文本生成图像模型的模型损失值;基于评估值,调整模型损失值,得到网络损失值。

[0058] 本可选实现方式中,以文本生成图像模型为主要调整网络,得到文本生成图像模型的损失函数。上述基于选取的图像样本和损失函数,计算文本生成图像模型损失值包括:基于选取的图像样本,得到文本生成图像模型的生成图像;基于损失函数,得到选取的图像样本与生成图像之间的差异值,将差异值作为模型损失值。

[0059] 上述基于评估值,调整模型损失值,得到网络损失值包括:将模型损失值除以评估值,得到网络损失值。

[0060] 本实施例提供的计算图像生成网络的网络损失值的方法,获取文本生成图像模型的损失函数,基于选取的图像样本和损失函数,计算文本生成图像模型的模型损失值;基于评估值,调整模型损失值,得到网络损失值,以文本生成图像模型为基准,得到模型损失值,在得到图像评分模型的评估值的基础上,调整模型损失值,得到网络损失值,为网络损失值的得到提供了一种可靠的实现方式。

[0061] 在本公开的一些可选实现方式中,上述基于图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型,包括:响应于图像生成网络的网络损失值满足训练完成条件,将图像生成网络作为图像生成模型。

[0062] 在本可选实现方式中,上述训练完成条件包括以下至少一项:图像生成网络的训练迭代次数达到预定迭代阈值,图像生成网络的网络损失值小于预定网络损失值阈值。其中,预定迭代阈值是基于图像生成网络的网络损失值得到的经验值。例如,图像生成网络的预定迭代阈值是5万次。图像生成网络的预定网络损失值阈值为0.01。

[0063] 本可选实现方式提供的得到图像生成模型的方法,在图像生成网络的网络损失值满足训练完成条件时,将图像生成网络作为图像生成模型,为图像生成模型的生成提供了一种可靠的实现方式。

[0064] 在本公开的一些可选实现方式中,上述图像生成模型训练方法还包括:响应于图像生成网络的网络损失值不满足训练完成条件,基于网络损失值,调整大语言模型的参数,并继续训练图像生成网络。

[0065] 本实施例中,上述继续执行总训练步骤是指:继续从图像样本集中选取图像样本,将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像;采用图像评分模型对生成图像进行评分,得到生成图像的评估值;基于评估值,计算图像生成网络的网络损失值;基于图像生成网络的网络损失值,得到图像生成模型。

[0066] 本实施提供的图像生成模型训练方法,在图像生成网络的网络损失值不满足训练完成条件时,调整大语言模型的参数,不改变文本生成图像模型的参数,可以将可插拔的文本生成图像模型应用于图像生成网络,提高了文本生成图像模型应用的灵活性。

[0067] 本实施例提供的图像生成模型训练方法,由于大语言模型的参数量较小,在图像

生成网络的网络损失值不满足训练完成条件时,只调整大语言模型的参数,不改变文本生成图像模型的参数,可以减少网络训练量,提高了图像生成网络的收敛效果。

[0068] 在本公开的一些可选实现方式中,上述多图像描述特征包括:元素特征、构图特征以及风格特征。

[0069] 本可选实现方式中,上述元素特征为描述待生成的生成图像中各个单元的特征文本,该单元可以是人、物、动物、景色,例如,识别文本中包括:一只公鸡,元素特征为:由复古锡和木材制成的公鸡。

[0070] 本可选实现方式中,上述构图特征为描述待生成的生成图像中各个单元的布局的特征文本。例如,识别文本中包括:位于图像中间的公鸡,则构图特征为白色背景下位于背景中心的公鸡。

[0071] 本可选实现方式中,上述风格特征为描述待生成的生成图像的风格特点的特征,例如,风格特征包括:复古风格、机械风格、漫画风格、油画风格等。

[0072] 可选地,上述多图像描述特征还可以包括:色调特征、情节特征等,其中,色调特征为描述待生成的生成图像的色调的特征,情节特征为描述待生成的生成图像的故事情节的特征。

[0073] 本可选实现方式提供的多图像描述特征,通过元素特征、构图特征以及风格特征对提示词文本进行限定,提高了提示词文本生成的丰富度,保证了图像生成模型的可靠性。

[0074] 在本公开的一些可选实现方式中,图像评分模型通过训练多模态图文识别网络得到,多模态图文识别网络的训练步骤包括:可以进行多次迭代的第一训练步骤,得到第一评分模块;然后可以进行多次迭代的第二训练步骤,第二训练步骤是在第一训练步骤之后进行的训练步骤;通过多次迭代第一训练步骤,可以得到第一评分模型;通过多次迭代第二训练步骤,可以得到图像评分模型。

[0075] 具体地,上述第一训练步骤包括:将获取的第一图像样本和描述文本输入多模态图文识别网络,得到多模态图文网络输出的答案文本;拼接第一图像样本、描述文本、答案文本以及评分文本,得到第一拼接信息,并将第一拼接信息输入多模态图文识别网络,得到多模态图文网络输出的第一评分;基于第一评分,计算多模态图文网络的损失值;基于多模态图文网络的损失值,得到第一评分模型。

[0076] 本实施例中,第一图像样本包括第一图像和第一图像的评分,其中,第一图像可以由图像生成模型生成的图像,第一图像的评分可以由人工对第一图像进行评分之后得到的评分,通过第一图像的评分、第一评分以及多模态图文网络的损失函数,可以计算得到多模态图文网络的损失值,响应于多模态图文网络的损失值满足训练完成条件,得到第一评分模型。

[0077] 本实施例中,描述文本是要求模型对第一图像样本中的第一图像进行描述的文本,例如,描述文本为:请以文本的形式描述第一图像中的主要内容。评分文本是要求模型对第一图像样本中的第一图像给出评分的文本,例如,评分文本为:请给出第一图像的评分。

[0078] 本实施例中,拼接第一图像样本、描述文本、答案文本以及评分文本得到第一拼接信息,并将第一拼接信息输入多模态图文识别网络,可以使多模态图文识别网络综合第一图像样本中第一图像、描述文本、答案文本以及评分文本,对第一图像样本中的第一图像给

出第一评分。

[0079] 本实施例中,答案文本是多模态图文识别网络在得到描述文本对第一图像样本中的第一图像的图像内容进行文本描述的文本,通过答案文本可以确定多模态图文识别网络对第一图像的描述的全面性。

[0080] 本实施例中,第一评分是多模态图文识别网络对第一图像样本中的第一图像进行评分之后的分值。其中,第一评分可以是一种对第一图像的美学效果、认知效果的评分。

[0081] 具体地,第二训练步骤包括:将获取的第二图像样本、所述评分文本输入所述第一评分模型,得到所述第一评分模型输出的第二评分;基于所述第二评分,计算所述第一评分模型的损失值;基于所述第一评分模型的损失值,得到图像评分模型。

[0082] 本实施例中,第二图像样本包括第二图像和第二图像的评分,其中,第二图像可以由图像生成模型生成的图像,第二图像的评分可以由人工对第二图像进行评分之后得到的评分,通过第二图像的评分、第二评分以及第一评分模型的损失函数(也是多模态图文网络的损失函数),可以计算得到第一评分模型的损失值,响应于第一评分模型的损失值满足训练完成条件,得到图像评分模型。

[0083] 本实施例中,通过多次迭代训练第二训练步骤,可以训练得到一种输入图像,输出为该图像的评分的图像评分模型,提高了图像评分模型使用的便利性。

[0084] 本实施例提供的训练多模态图文识别网络的方法,通过第一训练步骤可以使多模态图像识别网络充分理解图像以及图像内容的基础上,给出图像的分值;通过第二训练步骤可以使训练得到一种输入为图像,输出为分值的模型,提高了图像评分模型得到的可靠性。

[0085] 可选地,图像评分模型通过训练多模态图文识别网络得到,多模态图文识别网络的训练步骤包括:将获取的评分图像样本输入多模态图文识别网络,得到多模态图文识别网络输出的评分;响应于多模态图文识别网络满足训练完成条件,得到图像评分模型。

[0086] 进一步地,基于上述实施例提供的图像生成模型训练方法,本公开还提供了一种图像处理的一个实施例,本公开的图像处理方法结合了计算机视觉、深度学习等人工智能领域。

[0087] 参见图3,示出了根据本公开图像处理方法的一个实施例的流程300,本实施例提供的图像处理方法包括以下步骤:

[0088] 步骤301,获取待处理图像。

[0089] 在本实施例中,待处理图像可以包括人、物、景色等信息,通过图像生成模型对待处理图像进行处理,可以得到图像生成结果。图像处理方法的执行主体可以通过多种方式来获取待处理图像。例如,执行主体可以通过有线连接方式或无线连接方式,从数据库服务器中获取存储于其中的待处理图像。再例如,执行主体也可以实时接收终端或其他设备实时采集的待处理图像。

[0090] 步骤302,将待处理图像输入图像生成模型中,输出待处理图像的图像生成结果。

[0091] 本实施例中,执行主体可以将步骤301中获取的待处理图像输入图像生成模型中,从而得到待处理图像的图像生成结果。需要说明的是,图像生成结果包括生成图像,生成图像相对于待处理图像是一种新生成且进行了风格和/或内容变换之后的图像。

[0092] 本实施例中,图像生成模型可以是采用如上述图1实施例所描述的方法而训练得

到的,具体训练过程可以参见图1实施例的相关描述,在此不再赘述。

[0093] 本公开的实施例提供的图像处理方法,获取多个待处理图像;将待处理图像输入采用本实施例的图像生成模型训练方法生成的图像生成模型中,得到待处理图像的图像生成结果。由此,采用图像生成模型可以对待处理图像进行可靠的图像处理,提高了图像处理的有效性。

[0094] 在本公开的一些实施例中,上述图像处理方法包括:获取待处理图像;检测待处理图像的尺寸是否为标准尺寸;响应于待处理图像的尺寸不是标准尺寸,将待处理图像调整到标准尺寸。

[0095] 本实施例中,标准尺寸可以是与图像生成模型相适配的尺寸,例如,标准尺寸为448*448。

[0096] 本实施例中,可以通过图像处理工具中长度测量工具直接测量待处理图像的尺寸,在待处理图像的尺寸不是标准尺寸时,通过图像处理工具中图像裁剪工具或者图像缩放工具,对待处理图像进行处理,得到具有标准尺寸的待处理图像。

[0097] 本实施例提供的图像处理方法,在待处理图像的尺寸不是标准尺寸时,将待处理图像调整到标准尺寸,可以减少图像生成模型对图像处理的步骤,提高了图像生成结果得到的可靠性。

[0098] 在本公开的一些可选实现方式中,上述图像生成模型包括:图文识别模块、大语言模型以及文本生成图像模型,将待处理图像输入图像生成模型中,输出待处理图像的图像生成结果包括:将待处理图像输入图文识别模块,得到图文识别模块输出的识别文本;将识别文本输入大语言模型,得到大语言模型输出的提示词文本;将提示词文本输入文本生成图像模型,得到文本生成图像模型输出的生成图像。

[0099] 本实施例中,图文识别模块可以是多模态图文识别模型,而多模态图文识别模型是对多模态图文识别网络进行训练得到的模型,具体地,该多模态图文识别模型的训练过程如下:从图像样本集中获取图像样本,将该图像样本输入多模态图文识别网络,得到该多模态图文识别网络输出的文本;计算该多模态图文识别网络的损失值;响应于该多模态图像识别网络满足其训练完成条件,得到该多模态图文识别模型。

[0100] 如图4所示,将待处理图像D输入图文识别模块M1,得到图文识别模块M1输出的识别文本S,识别文本S的内容为“一只公鸡站在地上,有红色的鸡冠以及一个巨大的羽毛丰满的尾巴”;将识别文本S输入大语言模型M2,得到大语言模型M2输出的提示词文本T,提示词文本T的内容为“由复古锡和木材制成的公鸡,白色背景,神情骄傲,站姿威风,机械风格”;将提示词文本T输入文本生成图像模型M3,得到文本生成图像模型M3输出的生成图像W。

[0101] 本实施例提供的图像处理方法,在图像生成模型包括图文识别模块、大语言模型以及文本生成图像模型时,依次通过图文识别模块得到识别文本,通过大语言模型得到提示词文本,通过文本生成图像模型得到生成图像,由于大语言模型输出的多图像描述特征的提示词文本,提高了生成图像的准确度。

[0102] 在本公开的一些实施例中,上述图像处理方法还包括:接收图像处理需求文本;在得到识别文本之后,将识别文本与图像处理需求文本进行拼接,得到第二拼接信息;将第二拼接信息输入大语言模型,得到大语言模型输出的新的提示词文本;将新的提示词文本输入文本生成图像模型,得到文本生成图像模型输出的新的生成图像。

[0103] 本实施例中,图像处理需求文本可以是由用户输入的对待生成的生成图像的需求,通过图像处理需求文本可以提炼出用户的具体需求。

[0104] 本实施例提供的图像处理方法,获取待处理图像;接收图像处理需求文本;将待处理图像输入图文识别模块,得到图文识别模块输出的识别文本;将图像处理需求文本和识别文本进行拼接,得到第二拼接信息;将第二拼接信息输入大语言模型,得到大语言模型输出的新的提示词文本;将新的提示词文本输入文本生成图像模型,得到文本生成图像模型输出的新的生成图像。

[0105] 可选地,上述图像处理方法还可以包括:输出新的提示词文本,从而在用户交互环境中,模型可以提供即时的文本反馈,增强用户体验。模型可以根据用户的喜好和历史反馈定制化生成文本提示,还可以通过用户再次输入的图像处理需求文本以及业务场景定制化更加多样或者风格统一的文本。

[0106] 本实施例提供的图像处理方法,在接收到图像处理需求文本之后,将图像处理需求文本与识别文本进行拼接,得到第二拼接信息;将第二拼接信息输入大语言模型,得到大语言模型输出的新的提示词文本;将新的提示词文本输入文本生成图像模型,得到文本生成图像模型输出的新的生成图像,为此可以通过图像处理需求文本得到用户的图像处理,图像生成模型在融合用户处理需求的基础上,得到的新的生成图像,提高了生成图像生成的准确度。

[0107] 进一步参考图5,作为对上述各图所示方法的实现,本公开提供了图像生成模型训练装置的一个实施例,该装置实施例与图1所示的方法实施例相对应,该装置具体可应用于各种电子设备中。

[0108] 如图5所示,本实施例提供的图像生成模型训练装置500包括:集合获取单元501,网络获取单元502,样本输入单元503,评分单元504,计算单元505,模型得到单元506。其中,上述集合获取单元501,可以被配置成获取图像样本集,图像样本集包括至少一个图像样本。上述网络获取单元502,可以被配置成获取预先构建的图像生成网络,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,图文识别模块基于输入的图像,得到识别文本;大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;文本生成图像模型基于提示词文本,得到生成图像。上述样本输入单元503,可以被配置成将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像。上述评分单元504,可以被配置成采用图像评分模型对生成图像进行评分,得到生成图像的评估值。上述计算单元505,可以被配置成基于评估值,计算图像生成网络的网络损失值。上述模型得到单元506,可以被配置成基于图像生成网络的网络损失值,训练所述图像生成网络,得到经训练的图像生成模型。

[0109] 在本实施例中,图像生成模型训练装置500中:集合获取单元501,网络获取单元502,样本输入单元503,评分单元504,计算单元505,模型得到单元506的具体处理及其所带来的技术效果可分别参考图1对应实施例中的步骤101、步骤102、步骤103、步骤104、步骤105、步骤106的相关说明,在此不再赘述。

[0110] 在本实施例的一些可选的实现方式中,上述计算单元505进一步被配置成:获取文本生成图像模型的损失函数;基于选取的图像样本和损失函数,计算文本生成图像模型的模型损失值;基于评估值,调整模型损失值,得到网络损失值。

[0111] 在本实施例的一些可选的实现方式中,上述模型得到单元506进一步被配置成:响应于图像生成网络的网络损失值满足训练完成条件,将图像生成网络作为图像生成模型。

[0112] 在本实施例的一些可选的实现方式中,上述装置500还包括:调整单元(图中未示出),上述调整单元被配置成响应于图像生成网络的网络损失值不满足训练完成条件,基于网络损失值,调整大语言模型的参数,并控制样本输入单元503工作。

[0113] 在本实施例的一些可选实现方式中,上述多图像描述特征包括:元素特征、构图特征以及风格特征。

[0114] 在本实施例的一些可选实现方式中,上述图像评分模型通过训练多模态图文识别网络得到,多模态图文识别网络通过训练单元(图中未示出)训练得到,训练单元被配置成:将获取的第一图像样本和描述文本输入多模态图文识别网络,得到多模态图文网络输出的答案文本;拼接第一图像样本、描述文本、答案文本以及评分文本,得到第一拼接信息,并将第一拼接信息输入多模态图文识别网络,得到多模态图文网络输出的第一评分;基于第一评分,计算多模态图文网络的损失值;基于多模态图文网络的损失值,得到第一评分模型;将获取的第二图像样本、评分文本输入第一评分模型,得到第一评分模型输出的第二评分;基于第二评分,计算第一评分模型的损失值;基于第一评分模型的损失值,得到图像评分模型。

[0115] 本公开的实施例提供的图像生成模型训练装置,首先,集合获取单元501获取图像样本集,图像样本集包括至少一个图像样本;其次,网络获取单元502获取预先构建的图像生成网络,图像生成网络包括:依次连接的图文识别模块、大语言模型以及文本生成图像模型,图文识别模块基于输入的图像,得到识别文本;大语言模型基于识别文本,得到具有多图像描述特征的提示词文本;文本生成图像模型基于提示词文本,得到生成图像;再次,样本输入单元503将从图像样本集中选取的图像样本输入图像生成网络,得到图像生成网络输出的生成图像;再次,评分单元504采用图像评分模型对生成图像进行评分,得到生成图像的评估值;从次,计算单元505基于评估值,计算图像生成网络的网络损失值;最后,模型得到单元506基于图像生成网络的网络损失值,训练图像生成网络,得到经训练的图像生成模型。本公开通过大语言模型对识别文本进行多图像描述特征的提示词文本的生成,提高了提示词对文本生成图像模型的指导性;在图像生成网络训练过程中,通过图像评分模型对生成图像进行评分,并基于评分得到的评估值得到图像生成模型,提高了图像生成模型训练的可靠性和准确性。

[0116] 进一步参考图6,作为对上述各图所示方法的实现,本公开提供了图像处理装置的一个实施例,该装置实施例与图3所示的方法实施例相对应,该装置具体可应用于各种电子设备中。

[0117] 如图6所示,本实施例提供的图像处理装置600包括:图像获取单元601,结果得到单元602。其中,上述图像获取单元601,可以被配置成获取待处理图像。上述结果得到单元602,可以被配置成将待处理图像输入如上述图5实施例所描述的装置生成的图像生成模型中,输出待处理图像的图像生成结果。

[0118] 在本实施例中,图像处理装置600中:图像获取单元601,结果得到单元602的具体处理及其所带来的技术效果可分别参考图3对应实施例中的步骤301、步骤302的相关说明,在此不再赘述。

[0119] 在本实施例的一些可选实现方式中,图像处理装置600还包括:检测单元(图中未示出)。其中,上述检测单元被配置成:检测待处理图像的尺寸是否为标准尺寸;响应于待处理图像的尺寸不是标准尺寸,将待处理图像调整到标准尺寸。

[0120] 在本实施例的一些可选实现方式中,上述图像生成模型包括:图文识别模块、大语言模型以及文本生成图像模型,上述结果得到单元602进一步被配置成:将待处理图像输入图文识别模块,得到图文识别模块输出的识别文本;将识别文本输入大语言模型,得到大语言模型输出的提示词文本;将提示词文本输入文本生成图像模型,得到文本生成图像模型输出的生成图像。

[0121] 在本实施例的一些可选实现方式中,上述装置600还包括:接收单元(图中未示出)、文本输入单元(图中未示出)。上述接收单元,可被配置成接收图像处理需求文本。上述文本输入单元,可以被配置成在得到识别文本之后,将识别文本与图像处理需求文本进行拼接,得到第二拼接信息;将第二拼接信息输入大语言模型,得到大语言模型输出的新的提示词文本;将新的提示词文本输入文本生成图像模型,得到文本生成图像模型输出的新的生成图像。

[0122] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0123] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0124] 图7示出了可以用来实施本公开的实施例的示例电子设备700的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0125] 如图7所示,电子设备700包括计算单元701,其可以根据存储在只读存储器(ROM)702中的计算机程序或者从存储单元708加载到随机访问存储器(RAM)703中的计算机程序,来执行各种适当的动作和处理。在RAM 703中,还可存储电子设备700操作所需的各种程序和数据。计算单元701、ROM 702以及RAM703通过总线704彼此相连。输入/输出(I/O)接口705也连接至总线704。

[0126] 电子设备700中的多个部件连接至I/O接口705,包括:输入单元706,例如键盘、鼠标等;输出单元707,例如各种类型的显示器、扬声器等;存储单元708,例如磁盘、光盘等;以及通信单元709,例如网卡、调制解调器、无线通信收发机等。通信单元709允许电子设备700通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0127] 计算单元701可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元701的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元701执行上文所描述的各个方法和处理,例如图像生成模型训练方法或图像处理方法。例如,在一些实施例中,图像生成模型训练方法或图像处理方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存

储单元708。在一些实施例中,计算机程序的部分或者全部可以经由ROM 702和/或通信单元709而被载入和/或安装到电子设备700上。当计算机程序加载到RAM 703并由计算单元701执行时,可以执行上文描述的图像生成模型训练方法或图像处理方法的一个或多个步骤。备选地,在其他实施例中,计算单元701可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行图像生成模型训练方法或图像处理方法。

[0128] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0129] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程图像生成模型训练装置、图像处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0130] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0131] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0132] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网

(LAN)、广域网(WAN)和互联网。

[0133] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0134] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0135] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

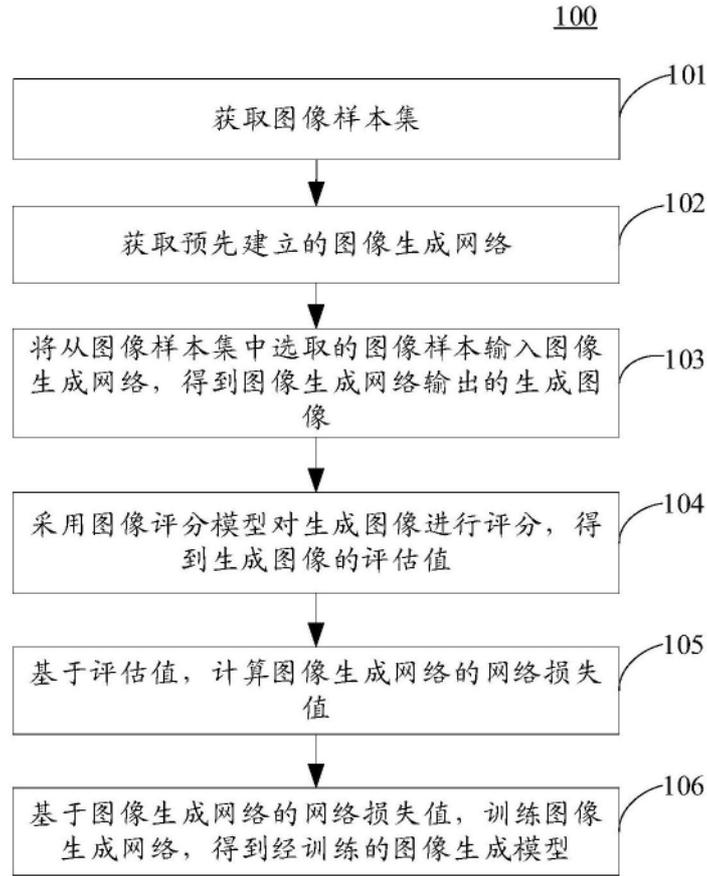


图1

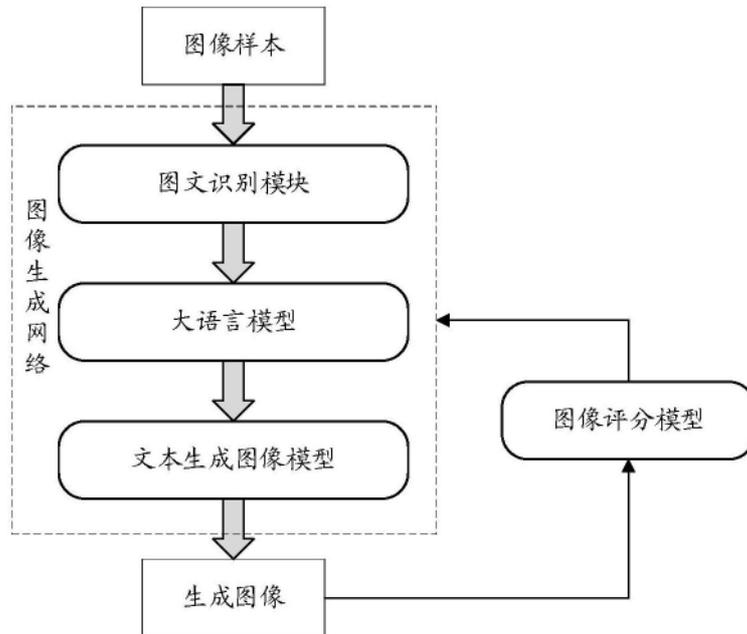


图2

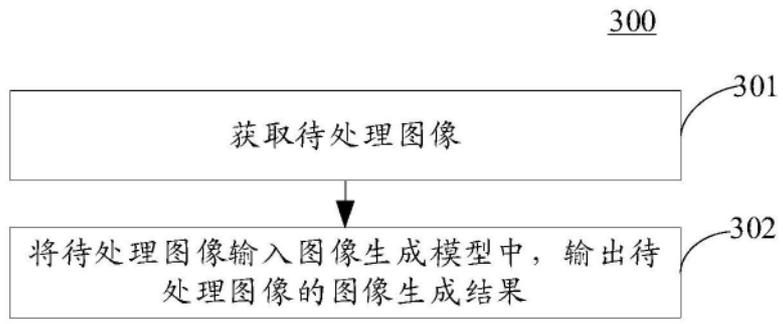


图3

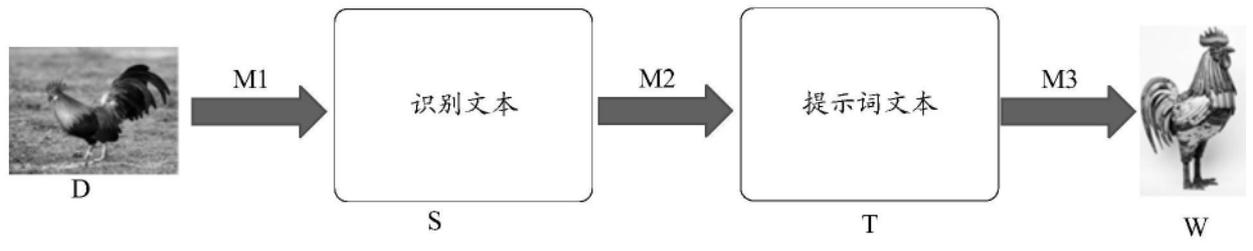


图4

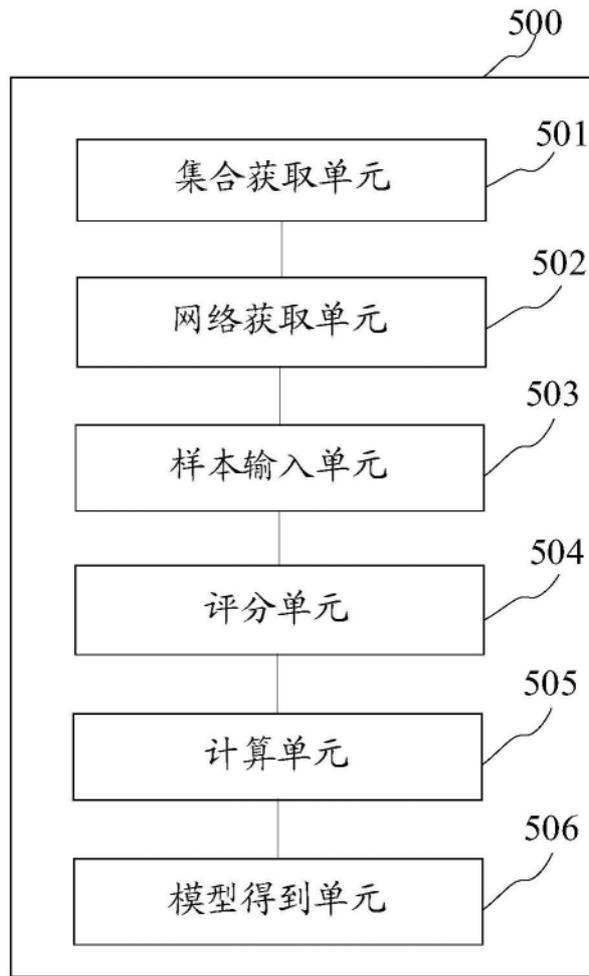


图5

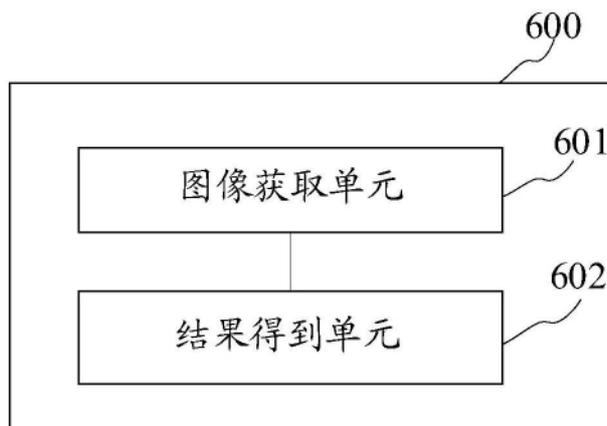


图6

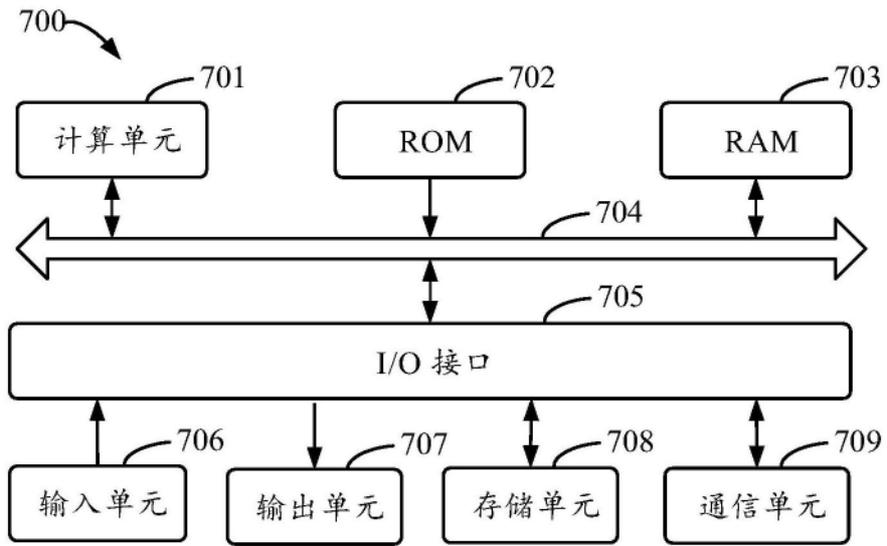


图7