



(12) 发明专利

(10) 授权公告号 CN 116664719 B

(45) 授权公告日 2023. 12. 29

(21) 申请号 202310941491.9

G06T 5/00 (2006.01)

(22) 申请日 2023.07.28

G06V 10/774 (2022.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 116664719 A

(56) 对比文件

CN 113850714 A, 2021.12.28

CN 114266840 A, 2022.04.01

US 2021264236 A1, 2021.08.26

US 2023153606 A1, 2023.05.18

CN 116188250 A, 2023.05.30

CN 110555896 A, 2019.12.10

CN 110599393 A, 2019.12.20

CN 114283050 A, 2022.04.05

CN 114904270 A, 2022.08.16

(43) 申请公布日 2023.08.29

(73) 专利权人 腾讯科技(深圳)有限公司

地址 518000 广东省深圳市南山区高新区

科技中一路腾讯大厦35层

(72) 发明人 郭卉

(74) 专利代理机构 广州嘉权专利商标事务所有

限公司 44205

专利代理师 余凯欢 谭英强

审查员 杨志忠

(51) Int. Cl.

G06T 11/20 (2006.01)

G06T 11/60 (2006.01)

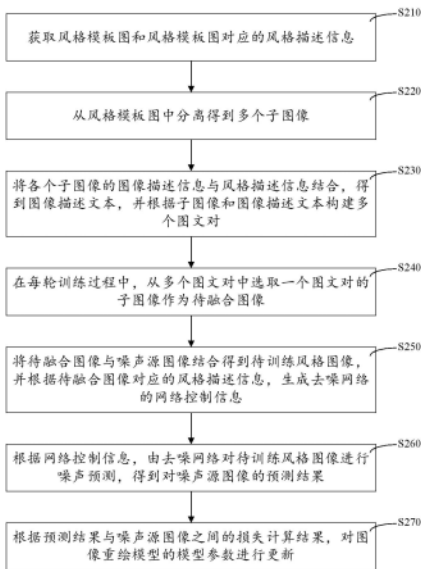
权利要求书3页 说明书29页 附图13页

(54) 发明名称

一种图像重绘模型训练方法、图像重绘方法及装置

(57) 摘要

本发明公开一种图像重绘模型训练方法、图像重绘方法及装置,该方法包括:获取风格模板图和风格描述信息;从风格模板图中分离得到子图像;将子图像的图像描述信息与风格描述信息结合得到图像描述文本,并根据子图像和图像描述文本构建多个图文对;从图文对中选取子图像作为待融合图像;将待融合图像与噪声源图像结合得到待训练风格图像,根据风格描述信息生成网络控制信息;根据网络控制信息,由去噪网络对待训练风格图像进行噪声预测得预测结果;根据预测结果与噪声源图像的损失计算结果,对模型参数进行更新。本发明使得少样本训练也能降低模型训练的过拟合效果,提高图像重绘准确性,可广泛应用于云计算、车联网等与互联网技术相关的其他技术领域。



1. 一种图像重绘模型训练方法,其特征在于,包括:

获取风格模板图和所述风格模板图对应的风格描述信息;

从所述风格模板图中分离得到多个子图像;

将各个所述子图像的图像描述信息与所述风格描述信息结合,得到图像描述文本,并根据所述子图像和所述图像描述文本构建多个图文对;其中,每个图文对中包含子图像和所述子图像对应的图像描述文本;

在每轮训练过程中,从多个所述图文对中选取一个图文对的所述子图像作为待融合图像;

将所述待融合图像与噪声源图像结合得到待训练风格图像,并根据所述待融合图像对应的风格描述信息,生成去噪网络的网络控制信息;

根据所述网络控制信息,由所述去噪网络对所述待训练风格图像进行噪声预测,得到对所述噪声源图像的预测结果;

根据所述预测结果与所述噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新。

2. 根据权利要求1所述的一种图像重绘模型训练方法,其特征在于,所述将各个所述子图像的图像描述信息与所述风格描述信息结合,得到图像描述文本,包括:

对各个所述子图像进行描述提取,得到各个所述子图像的图像描述信息;

对所述风格描述信息进行表征获取,得到所述风格描述信息对应的风格表征信息;

在每个所述子图像的图像描述信息前加入所述风格表征信息,得到每个所述子图像对应的图像描述文本。

3. 根据权利要求2所述的一种图像重绘模型训练方法,其特征在于,所述对所述风格描述信息进行表征获取,得到所述风格描述信息对应的风格表征信息,包括:

将所述风格描述信息与风格词库中的各个风格词进行查询匹配,确定所述风格描述信息对应的候选文本表征信息;

将所述候选文本表征信息与各个所述子图像的图像描述信息结合,确定所述风格描述信息对应的风格表征信息。

4. 根据权利要求3所述的一种图像重绘模型训练方法,其特征在于,所述方法还包括构建所述风格词库的步骤,该步骤包括:

收集多种风格词以及各种风格词对应的风格图像;

获取每种所述风格词的文本表征向量;

根据所述文本表征向量的所述风格图像,构建所述风格词库;其中,所述风格词库的索引项为各个所述风格词,每个所述索引项关联所述风格词的文本表征向量。

5. 根据权利要求3或4所述的一种图像重绘模型训练方法,其特征在于,所述将所述候选文本表征信息与各个所述子图像的图像描述信息结合,确定所述风格描述信息对应的风格表征信息,包括:

获取所述候选文本表征信息与所述图像描述信息之间的中点值,作为所述风格表征信息;

其中,所述风格表征信息用于表征所述子图像的图像约束信息和所述风格描述信息的文字约束信息。

6. 根据权利要求2-4任一项所述的一种图像重绘模型训练方法,其特征在于,所述去噪网络的网络控制信息由语义控制网络生成,所述方法还包括:

根据所述风格表征信息,对所述语义控制网络进行更新。

7. 根据权利要求6所述的一种图像重绘模型训练方法,其特征在于,所述根据所述风格表征信息,对所述语义控制网络进行更新,包括:

获取图像文本匹配模型的词汇表以及所述词汇表中各个词汇的词汇标识;其中,所述图像文本匹配模型用于根据输入的图像确定对应的文本描述信息;

对各个所述词汇标识进行文本表征推理,得到每个所述词汇标识对应的初始文本表征信息;

计算各个所述初始文本表征信息和所述风格表征信息之间的欧氏距离;

根据所述欧氏距离,从各个所述初始文本表征信息确定所述风格表征信息对应的风格词汇标识;

根据所述风格词汇标识,对所述语义控制网络进行更新。

8. 根据权利要求1所述的一种图像重绘模型训练方法,其特征在于,所述根据所述网络控制信息,由所述去噪网络对所述待训练风格图像进行噪声预测,得到对所述噪声源图像的预测结果,包括:

对所述待训练风格图像进行扩散采样处理,得到所述待训练风格图像的隐空间表征信息;

根据所述网络控制信息,由所述去噪网络对所述隐空间表征信息进行去噪处理,得到对所述噪声源图像的预测结果。

9. 根据权利要求8所述的一种图像重绘模型训练方法,其特征在于,所述根据所述网络控制信息,由所述去噪网络对所述隐空间表征信息进行去噪处理,得到对所述噪声源图像的预测结果,包括:

构建旁路注意力矩阵约束模块;

根据所述旁路注意力矩阵约束模块对所述风格描述信息进行前向计算,得到旁路注意力权重;

将所述旁路注意力权重融合到所述去噪网络中的主路注意力矩阵约束模块,对所述主路注意力矩阵约束模块的注意力矩阵进行更新;

通过所述主路注意力矩阵约束模块对所述隐空间表征信息进行去噪处理,得到对所述噪声源图像的预测结果。

10. 一种图像重绘方法,其特征在于,包括:

获取待重绘图像和所述待重绘图像对应的风格描述信息;

根据图像重绘模型对所述待重绘图像进行图像重绘;

其中,所述图像重绘模型根据权利要求1-9中任一项所述的图像重绘模型训练方法确定。

11. 一种图像重绘模型训练装置,其特征在于,包括:

第一模块,用于获取风格模板图和所述风格模板图对应的风格描述信息;

第二模块,用于从所述风格模板图中分离得到多个子图像;

第三模块,用于将各个所述子图像的图像描述信息与所述风格描述信息结合,得到图

像描述文本,并根据所述子图像和所述图像描述文本构建多个图文对;其中,每个图文对中
包含子图像和所述子图像对应的图像描述文本;

第四模块,用于在每轮训练过程中,从多个所述图文对中选取一个图文对的所述子图
像作为待融合图像;

第五模块,用于将所述待融合图像与噪声源图像结合得到待训练风格图像,并根据所
述待融合图像对应的风格描述信息,生成去噪网络的网络控制信息;

第六模块,用于根据所述网络控制信息,由所述去噪网络对所述待训练风格图像进行
噪声预测,得到对所述噪声源图像的预测结果;

第七模块,用于根据所述预测结果与所述噪声源图像之间的损失计算结果,对图像重
绘模型的模型参数进行更新。

12.一种图像重绘装置,其特征在于,包括:

第八模块,用于获取待重绘图像和所述待重绘图像对应的风格描述信息;

第九模块,用于根据图像重绘模型对所述待重绘图像进行图像重绘;

其中,所述图像重绘模型根据权利要求1-9中任一项所述的图像重绘模型训练方法确
定。

13.一种电子设备,其特征在于,包括处理器以及存储器;

所述存储器用于存储程序;

所述处理器执行所述程序实现如权利要求1至10中任一项所述的方法。

14.一种计算机可读存储介质,其特征在于,所述存储介质存储有程序,所述程序被处
理器执行实现如权利要求1至10中任一项所述的方法。

一种图像重绘模型训练方法、图像重绘方法及装置

技术领域

[0001] 本发明涉及图像处理技术领域,尤其是一种图像重绘模型训练方法、图像重绘方法及装置。

背景技术

[0002] 随着计算机技术的不断发展,相关技术可以利用人工智能技术对输入的文字需求进行处理,然后对指定的图像进行画风、背景等模态的微调,从而得到微调后的目标图像。在相关技术中,通常需要训练对应的图像重绘模型来实现上述图像重绘功能,但现实中可用于训练的不同风格的图像有时候很少,在训练样本小时模型容易过拟合,导致生成图像与训练图像完全相同。

发明内容

[0003] 有鉴于此,本发明实施例提供一种高效的图像重绘模型训练方法、图像重绘方法及装置,使得少样本训练也能降低模型训练过程的过拟合效果,提高图像重绘的准确性。

[0004] 本发明实施例的一方面提供了一种图像重绘模型训练方法,包括以下步骤:

[0005] 获取风格模板图和所述风格模板图对应的风格描述信息;

[0006] 从所述风格模板图中分离得到多个子图像;

[0007] 将各个所述子图像的图像描述信息与所述风格描述信息结合,得到图像描述文本,并根据所述子图像和所述图像描述文本构建多个图文对;其中,每个图文对中包含子图像和所述子图像对应的图像描述文本;

[0008] 在每轮训练过程中,从多个所述图文对中选取一个图文对的所述子图像作为待融合图像;

[0009] 将所述待融合图像与噪声源图像结合得到待训练风格图像,并根据所述待融合图像对应的风格描述信息,生成去噪网络的网络控制信息;

[0010] 根据所述网络控制信息,由所述去噪网络对所述待训练风格图像进行噪声预测,得到对所述噪声源图像的预测结果;

[0011] 根据所述预测结果与所述噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新。

[0012] 另一方面,本发明实施例还提供了一种图像重绘方法,包括以下步骤:

[0013] 获取待重绘图像和所述待重绘图像对应的风格描述信息;

[0014] 从所述待重绘图像中分离得到多个子图像;

[0015] 将各个所述子图像的图像描述信息与所述风格描述信息结合,得到图像描述文本;

[0016] 将所述待重绘图像与噪声源图像结合得到第一图像;

[0017] 根据所述图像描述文本,生成图像重绘模型的网络控制信息;

[0018] 根据所述网络控制信息,由所述图像重绘模型对所述第一图像进行图像重绘,得

到所述待重绘图像的重绘结果。

[0019] 另一方面,本发明实施例还提供了一种图像重绘方法,包括以下步骤:

[0020] 获取待重绘图像和所述待重绘图像对应的风格描述信息;

[0021] 根据图像重绘模型对所述待重绘图像进行图像重绘;

[0022] 其中,所述图像重绘模型根据上述的图像重绘模型训练方法确定。

[0023] 另一方面,本发明实施例还提供了一种图像重绘模型训练装置,包括:

[0024] 第一模块,用于获取风格模板图和所述风格模板图对应的风格描述信息;

[0025] 第二模块,用于从所述风格模板图中分离得到多个子图像;

[0026] 第三模块,用于将各个所述子图像的图像描述信息与所述风格描述信息结合,得到图像描述文本,并根据所述子图像和所述图像描述文本构建多个图文对;其中,每个图文对中包含子图像和所述子图像对应的图像描述文本;

[0027] 第四模块,用于在每轮训练过程中,从多个所述图文对中选取一个图文对的所述子图像作为待融合图像;

[0028] 第五模块,用于将所述待融合图像与噪声源图像结合得到待训练风格图像,并根据所述待融合图像对应的风格描述信息,生成去噪网络的网络控制信息;

[0029] 第六模块,用于根据所述网络控制信息,由所述去噪网络对所述待训练风格图像进行噪声预测,得到对所述噪声源图像的预测结果;

[0030] 第七模块,用于根据所述预测结果与所述噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新。

[0031] 可选地,本发明实施例中的图像重绘模型训练装置,第三模块包括:

[0032] 第一单元,用于对各个所述子图像进行描述提取,得到各个所述子图像的图像描述信息;

[0033] 第二单元,用于对所述风格描述信息进行表征获取,得到所述风格描述信息对应的风格表征信息;

[0034] 第三单元,用于在每个所述子图像的图像描述信息前加入所述风格表征信息,得到每个所述子图像对应的图像描述文本。

[0035] 可选地,本发明实施例中的图像重绘模型训练装置,第二单元包括:

[0036] 第一子单元,用于将所述风格描述信息与风格词库中的各个风格词进行查询匹配,确定所述风格描述信息对应的候选文本表征信息;

[0037] 第二子单元,用于将所述候选文本表征信息与各个所述子图像的图像描述信息结合,确定所述风格描述信息对应的风格表征信息。

[0038] 可选地,本发明实施例中的图像重绘模型训练装置还包括第七模块、第八模块以及第九模块,第七模块、第八模块以及第九模块应用于构建风格词库的步骤,其中:

[0039] 第七模块,用于收集多种风格词以及各种风格词对应的风格图像;

[0040] 第八模块,用于获取每种所述风格词的文本表征向量;

[0041] 第九模块,用于根据所述文本表征向量的所述风格图像,构建所述风格词库;其中,所述风格词库的索引项为各个所述风格词,每个所述索引项关联所述风格词的文本表征向量。

[0042] 可选地,本发明实施例中的图像重绘模型训练装置,第二子单元包括:

- [0043] 第一分部单元,用于获取所述候选文本表征信息与所述图像描述信息之间的中点值,作为所述风格表征信息;
- [0044] 其中,所述风格表征信息用于表征所述子图像的图像约束信息和所述风格描述信息的文字约束信息。
- [0045] 可选地,本发明实施例中的图像重绘模型训练装置中去噪网络的网络控制信息由语义控制网络生成,本发明实施例中的图像重绘模型训练装置还包括:
- [0046] 第十模块,用于根据所述风格表征信息,对所述语义控制网络进行更新。
- [0047] 可选地,本发明实施例中的图像重绘模型训练装置,第十模块包括:
- [0048] 第四单元,用于获取图像文本匹配模型的词汇表以及所述词汇表中各个词汇的词汇标识;其中,所述图像文本匹配模型用于根据输入的图像确定对应的文本描述信息;
- [0049] 第五单元,用于对各个所述词汇标识进行文本表征推理,得到每个所述词汇标识对应的初始文本表征信息;
- [0050] 第六单元,用于计算各个所述初始文本表征信息和所述风格表征信息之间的欧氏距离;
- [0051] 第七单元,用于根据所述欧氏距离,从各个所述初始文本表征信息确定所述风格表征信息对应的风格词汇标识;
- [0052] 第八单元,用于根据所述风格词汇标识,对所述语义控制网络进行更新。
- [0053] 可选地,本发明实施例中的图像重绘模型训练装置,第六模块包括:
- [0054] 第九单元,用于对所述待训练风格图像进行扩散采样处理,得到所述待训练风格图像的隐空间表征信息;
- [0055] 第十单元,用于根据所述网络控制信息,由所述去噪网络对所述隐空间表征信息进行去噪处理,得到对所述噪声源图像的预测结果。
- [0056] 可选地,本发明实施例中的图像重绘模型训练装置,第十单元包括:
- [0057] 第三子单元,用于构建旁路注意力矩阵约束模块;
- [0058] 第四子单元,用于根据所述旁路注意力矩阵约束模块对所述风格描述信息进行前向计算,得到旁路注意力权重;
- [0059] 第五子单元,用于将所述旁路注意力权重融合到所述去噪网络中的主路注意力矩阵约束模块,对所述主路注意力矩阵约束模块的注意力矩阵进行更新;
- [0060] 第六子单元,用于通过所述主路注意力矩阵约束模块对所述隐空间表征信息进行去噪处理,得到对所述噪声源图像的预测结果。
- [0061] 另一方面,本发明实施例还提供了一种图像重绘装置,包括:
- [0062] 第八模块,用于获取待重绘图像和所述待重绘图像对应的风格描述信息;
- [0063] 第九模块,用于根据图像重绘模型对所述待重绘图像进行图像重绘;
- [0064] 其中,所述图像重绘模型根据上述的图像重绘模型训练方法确定。
- [0065] 另一方面,本发明实施例还提供了一种电子设备,包括处理器以及存储器;
- [0066] 所述存储器用于存储程序;
- [0067] 所述处理器执行所述程序实现如前面的图像重绘模型训练方法或图像重绘方法。
- [0068] 另一方面,本发明实施例还提供了一种计算机可读存储介质,所述存储介质存储有程序,所述程序被处理器执行实现如前面的图像重绘模型训练方法或图像重绘方法。

[0069] 另一方面,本发明实施例还提供了一种计算机程序产品,所述计算机程序被处理器执行时实现如前面的图像重绘模型训练方法或图像重绘方法。

[0070] 本发明实施例首先获取风格模板图和风格模板图对应的风格描述信息,并从风格模板图中分离得到多个子图像,每个子图像可以用于后续多轮迭代训练的每轮输入图像,使得本发明实施例能够对样本量少(甚至只有一张)的风格模板图进行拆解,让最终进入隐空间去噪网络和图像重绘模型的可训练图像样本量增加,提高模型训练的准确性,防止过拟合现象;接着,本发明实施例将各个所述子图像的图像描述信息与所述风格描述信息结合,得到图像描述文本,本发明实施例对样本量少的风格描述信息进行拓展,使得最终进入隐空间去噪网络和图像重绘模型的可训练风格描述样本量增加,进一步提高模型训练的准确性;另外,本发明实施例对子图像的图像描述文本的构建进行优化,将原始的风格描述信息与子图像自动生成的图像描述信息相结合,产生更准确风格信息嵌入,使得模型更容易学习风格信息,提高模型的准确性的同时能减少因风格描述不准导致的欠拟合;在每轮训练过程中,本发明实施例从多个图文对中选取一个图文对的子图像作为待融合图像;将待融合图像与噪声源图像结合得到待训练风格图像,并根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息;最后根据网络控制信息,对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果;至此,本发明实施例可以根据预测结果与所述噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新,并且可以根据图像重绘模型对待重绘图像进行图像重绘。

附图说明

[0071] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0072] 图1是相关技术中通过稳定扩散生成模型的实现图像重绘的原理示意图;

[0073] 图2是本发明实施例中进行图像重绘模型训练的一种实施环境示意图;

[0074] 图3为本发明实施例中所提供的一种图像重绘模型训练方法的步骤流程示意图;

[0075] 图4为本发明实施例中多模态视觉语言预训练模型的原理示意图;

[0076] 图5为本发明实施例中风格表征信息获取的原理示意图;

[0077] 图6为本发明实施例中图像重绘模型训练的实现过程原理示意图;

[0078] 图7为本发明实施例中文本表征的步骤流程示意图;

[0079] 图8为本发明实施例中查询-键-值的结构示意图;

[0080] 图9为本发明实施例中查询-键-值中残差模块的结构示意图;

[0081] 图10为本发明实施例中查询-键-值中空间变换网络的结构示意图;

[0082] 图11为本发明实施例中旁路注意力矩阵约束模块与U型网格的结构示意图;

[0083] 图12为本发明实施例中所提供的一种图像重绘方法的步骤流程示意图;

[0084] 图13为本发明实施例中所提供的图像重绘模型推理过程示意图;

[0085] 图14为本发明实施例中所提供的图像重绘应用过程示意图;

[0086] 图15为本发明实施例中所提供的模板图像原图;

- [0087] 图16为本发明实施例中所提供的对图15的模板图像原图进行掩码遮盖的示意图；
- [0088] 图17为本发明另一实施例中所提供的一种图像重绘方法的步骤流程示意图；
- [0089] 图18为本发明实施例中所提供的一种图像重绘模型训练装置的模块化示意图；
- [0090] 图19为本发明实施例中所提供的一种图像重绘装置的模块化示意图；
- [0091] 图20为本发明实施例中所提供的一种电子设备的结构示意图。

具体实施方式

[0092] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0093] 应了解,在本发明实施例的描述中,多个(或多项)的含义是两个以上,大于、小于、超过等理解为不包括本数,以上、以下、以内等理解为包括本数。如果有描述到“第一”、“第二”等只是用于区分技术特征为目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量或者隐含指明所指示的技术特征的先后关系。

[0094] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本发明的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0095] 对本发明实施例进行进一步详细说明之前,对本发明实施例中涉及的名词和术语进行说明,本发明实施例中涉及的名词和术语适用于如下的解释:

[0096] 人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

[0097] 人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。本发明所示的包含图像采集组件的显示设备主要涉及其中的计算机视觉技术以及机器学习/深度学习、自动驾驶、智慧交通等方向。

[0098] 机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、示教学习等技术。

[0099] 深度学习,深度学习的概念源于人工神经网络的研究,含多隐层的多层感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特

征,以发现数据的分布式特征表示。深度学习是机器学习研究中的一个新的领域,其动机在于建立、模拟人脑进行分析学习的神经网络,通过模仿人脑的机制来解释数据,例如图像、声音和文本等。常用的深度学习模型包括:卷积神经网络(Convolutional Neural Networks,CNN)、循环神经网络(Recurrent Neural Network,RNN)、是长短期记忆网络(Long Short-Term Memory,LSTM)、深度神经网络(Deep Neural Network,DNN)、深度置信网(Deep Belief Nets,DBNs)等神经网络。数据在神经网络中的传播有两种方式,一种是沿着输入到输出的路径,被称为前向传播(Forward propagation),另一种是从输出返回到输入,被称为反向传播(Backpropagation)。在正向传播过程中,输入信息通过神经网络逐层处理并传向输出层,通过损失函数描述输出值与期望之间的误差,转入反向传播,逐层求出损失函数对各神经元的权重的偏导数,构成损失函数对权值向量的权重梯度数据,作为更新权重参数的依据,在不断更新权重参数的过程中完成神经网络的训练。

[0100] 损失函数(loss function),是将随机事件或其有关随机变量的取值映射为非负实数以表示该随机事件的“风险”或“损失”的函数。在应用中,损失函数通常作为学习准则与优化问题相联系,即通过最小化损失函数求解和评估模型。例如,在机器学习中,损失函数被用于模型的参数估计(parameteric estimation),基于损失函数得到的损失值可用来描述模型的预测值与实际值的差异程度。常见的损失函数有均方误差损失函数、支持向量机(Support Vector Machine,SVM)合页损失函数、交叉熵损失函数等。

[0101] 基于上述的理论基础,以及人工智能技术研究和进步,人工智能技术在多个领域展开研究和应用,例如常见的智能家居、智能穿戴设备、虚拟助理、智能音箱、智能营销、无人驾驶、自动驾驶、无人机、机器人、智能医疗、智能客服等,相信随着技术的发展,人工智能技术将在更多的领域得到应用,并发挥越来越重要的价值。

[0102] 为了便于理解本发明实施例,对本发明实施例可能涉及到的关键词进行解释:

[0103] 局部重绘(inpainting)模型:是一种计算机视觉模型,旨在通过从图像中删除缺失或损坏的区域并填补这些区域来恢复图像的完整性。其中,局部重绘模型可以用于修复被破坏的图像,去除不希望出现在图像中的对象,或从单个图像生成新的图像。

[0104] 交叉注意力控制(Cross Attention Control Module)模块:是一种在深度学习中使用的模块,用于在多个输入之间建立交叉注意力机制。其中,交叉注意力控制模块可以帮助模型在处理多个输入时,自动地学习输入之间的相关性,从而提高模型的性能。

[0105] 自反卷积(Self-Deconvolution Inpainting)模型:是一种用于图像修复的深度学习模型。图像修复是指通过填充缺失、损坏或噪声的像素来恢复原始图像的过程。SD-Inpainting模型基于自协方差矩阵(self-covariance matrix)的概念,利用深度卷积神经网络学习了图像中的统计结构,并通过自卷积操作从上下文信息中预测缺失的像素值。

[0106] 随着计算机技术的不断发展,相关技术可以利用人工智能技术对输入的文字需求进行处理,然后对指定的图像进行画风、背景等模态的微调,从而得到微调后的目标图像。在相关技术中,通常需要训练对应的图像重绘模型来实现上述图像重绘功能,但现实中可用于训练的不同风格的图像有时候很少,在训练样本小时模型容易过拟合,导致生成图像与训练图像完全相同。

[0107] 相关技术中,如图1所示,根据图像处理开源模型,如稳定扩散(stable-diffusion)模型,输入n张(n为20以上)的风格图像。然后,对图1中的图像去噪深度神经网络

络(Denoising U-Net)结构进行微调,把相应的描述词语与对应的训练图像作为图文对进行模型训练。相应地,在应用时,直接输入相关的风格描述,便能产生对应的图像结果。另外,在文本反转方法中,通过微调图1中的文本编码器 τ ,其余网络模块固定,从而避免前面的方法中的过拟合问题。相应地,该方法同样把风格词与训练图像构成训练图文对。然后,该方法对目标风格词的文本嵌入进行随机初始化,并通过模型训练得到一个与目标图像更接近的文本表征。其中,前面的方法中,第一种通过改变原始U-Net的参数,微调U-Net的方法,很容易让模型遗忘原模型训练好的参数,而直接记住现有有限训练集,从而导致过拟合问题,例如只能生成与训练集相近的图像,无法对不同的提示词产生对应的图像,如输入风格词和其他文本描述提示词后,最终只能产生训练图像。而第二种文本反转的方法虽然可以抗过拟合,但容易出现欠拟合现象,即生成图像风格与目标图像不一样。另外,相关技术中都采用单一风格词来微调生成模型,在训练样本数据量有限的情况下,容易出现模型对训练任务不明确,无法归纳出训练的“风格词”文本,无法确定应该与图像中的物体、环境还是风格进行学习,导致生成的结果常常带上无关的物体,例如,当训练图像上有一台空中飞船,则该风格下总出现该空中飞船,即模型任务“风格词”对应了画面中的飞船。

[0108] 为此,本发明实施例提供的方案涉及人工智能的深度学习等技术,基于构建的深度学习模型,通过对子图像的图像描述文本的构建进行优化,将原始的风格描述信息与子图像自动生成的图像描述信息相结合,产生更准确风格信息嵌入,使得模型更容易学习风格信息,提高模型的准确性的同时能减少因风格描述不准导致的欠拟合。同时,在每轮训练过程中,本发明实施例从多个图文对中选取一个图文对的子图像作为待融合图像,并将待融合图像与噪声源图像结合得到待训练风格图像,然后根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息。最后根据网络控制信息,对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果,从而能够根据图像重绘模型对待重绘图像进行图像重绘。

[0109] 可以理解的是,本发明实施例提供的图像重绘模型训练方法,是能够应用于任何一种具备数据处理计算能力计算机设备,而这一计算机设备可以是各类终端或是服务器。当本发明实施例中的计算机设备是服务器时,该服务器是独立的物理服务器,或者,是多个物理服务器构成的服务器集群或者分布式系统,或者,是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN(Content Delivery Network,内容分发网络)、以及大数据和人工智能平台等基础云计算服务的云服务器。可选地,该终端是智能手机、平板电脑、笔记本电脑以及台式计算机等,但也并不局限于此。

[0110] 需要进一步补充说明的是,本发明实施例中所涉及终端包括但不限于智能手机、电脑、智能语音交互设备、智能家电、车载终端、飞行器等。本发明实施例可应用于各种场景,包括但不限于云技术、人工智能、智慧交通、辅助驾驶等。

[0111] 在一些可能实现方式中,能够实现本发明实施例所提供的图像重绘模型训练方法或者图像重绘方法的计算机程序,可被部署在一个计算机设备上执行,或者在位于一个地点的多个计算机设备上执行,又或者,在分布在多个地点且通过通信网络互连的多个计算机设备上执行,分布在多个地点且通过通信网络互连的多个计算机设备能够组成区块链系统。

[0112] 基于多个计算机设备能够组成区块链系统,本发明实施例中实现图像重绘方法的

计算机设备可以是区块链系统中的节点。该节点中存储中有深度学习模型,如图像重绘模型,该深度学习模型获取待重绘图像和待重绘图像对应的风格描述信息,进而根据图像重绘模型对待重绘图像进行图像重绘。该节点或者该区块链中的其他设备对应的节点还能够存储图像重绘的重绘结果、风格描述信息以及在图像重绘过程中得到其它中间特征数据等等。

[0113] 如图2所示,是本发明实施例提供的一种实施环境示意图。参照图2,该实施环境包括至少一个终端101和服务器102。终端101和服务器102之间可以通过无线或者有线的方式进行网络连接,完成数据传输交换。终端101可以设置有模型训练、模型训练参数的有监督调整以适应于下游任务要求等功能,或者终端101通过与服务器102交互通信,基于服务器102提供对应的程序逻辑实现适应于下游任务要求的相关功能。可选地,服务器102为该程序逻辑的后台服务器或者为提供云计算以及云存储等服务的云服务器。该终端101可以根据获取的特征样本进行各种图像重绘模型的训练。预训练好进而构建得到的图像重绘模型可以直接应用于各种图像重绘(包括但不限于不同类型的图像重绘,例如不同风格图像的重绘、图像修复等)的下游任务中,还可以进一步根据下游任务的需求,基于重绘结果进行相应数据处理(包括但不限于对相关的重绘图像进行数据展示或存储等处理内容)。其中,终端101为电脑、手机、平板电脑或者其它终端。

[0114] 示例性地,服务器可以从终端处获取风格模板图和风格模板图对应的风格描述信息进行图像重绘模型的训练,然后执行如下处理以构建得到训练好的图像重绘模型:a. 从风格模板图中分离得到多个子图像;b. 将各个子图像的图像描述信息与风格描述信息结合,得到图像描述文本,并根据子图像和图像描述文本构建多个图文对;其中,每个图文对中包含子图像和子图像对应的图像描述文本;c. 在每轮训练过程中,从多个图文对中选取一个图文对的子图像作为待融合图像;d. 将待融合图像与噪声源图像结合得到待训练风格图像,并根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息;e. 根据网络控制信息,由去噪网络对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果;f. 根据预测结果与噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新。服务器可以将预训练好的图像重绘模型发送至终端,终端基于该预训练好的图像重绘模型执行各种图像重绘的下游任务。

[0115] 该模型训练方法可以应用于各种场景中。

[0116] 示例性地,在目标图像重绘场景下:在获得了目标对象授权的情况下,获取目标对象提供的待重绘图像和待重绘图像对应的风格描述信息,通过终端101发送给服务器102,服务器102采用前述方法获得预训练好的图像重绘模型并发送给终端101,这样,终端101就可以利用该图像重绘模型进行图像重绘,进而根据图像重绘模型对待重绘图像进行图像重绘。可以理解,上述的应用场景仅起到示例作用,在实际应用中,除了可以利用服务器对图像重绘模型进行预训练之外,还可以利用其它具备模型训练能力的设备如终端,预训练并构建得到图像重绘模型;此外,也可以采用除终端设备外的其他设备承载预训练好的图像重绘模型,在此不对本发明实施例提供的模型训练方法的应用场景做任何限定。

[0117] 在图2所示的实施环境的基础上,本发明实施例提供了一种图像的重绘场景。在该场景中,服务器通过与终端的数据交互,服务器可以通过终端上搭载的目标应用获取目标对象的待重绘图像和待重绘图像对应的风格描述信息,如目标对象需要进行重绘的图像以

及相应的重绘风格描述,进一步根据服务器预存的图像重绘模型对待重绘图像进行重绘,从而实现对待重绘图像的图像重绘,并提高了图像重绘的准确性。示例性地,目标应用可以为进行图像应用处理平台,也可以为多媒体应用(如影视应用、短视频应用或音乐应用等等)、社交应用、游戏应用和导航应用等应用中设置的带有图像处理功能的网页链接、小程序或应用插件。

[0118] 需要说明的是,在本发明的各个具体实现方式中,当涉及到需要获取对象的特征数据、对象行为数据,对象历史数据以及对象位置信息等与对象身份或特性相关的数据进行相关处理时,都会先获得对象的许可或者授权,而且,对这些数据的收集、使用和处理等,都会遵守相关国家和地区的相关法律法规和标准。此外,当本发明实施例需要获取对象的敏感个人信息时,会通过弹窗或者跳转到确认页面等方式获得对象的单独许可或者单独同意,在明确获得对象的单独许可或者单独同意之后,再获取用于使本发明实施例能够正常运行的必要的对象相关数据。

[0119] 如图3所示,是本发明实施例提供的一种图像重绘模型训练方法的步骤流程图,该图像重绘模型训练方法的执行主体可以是前述的任意一种计算机设备。参照图3,该方法包括以下步骤:

[0120] S210:获取风格模板图和风格模板图对应的风格描述信息。

[0121] 本发明实施例中,风格模板图是指用于描述和展示图像风格的模板图,不同风格模板图可以包含各种不同的设计元素,如颜色、形状、纹理、图案等,以及各种不同的风格特征,如抽象、现实、流行、哥特等。相应地,本发明实施例中风格描述信息是指风格模板图的相关描述词或风格词,例如极简主义、表现主义、流行艺术、新艺术风格等。本发明实施例在获取不同的风格模板图的同时,获取各个风格模板图对应的风格描述信息。示例性地,在一风格模板图获取的应用场景中,本发明实施例获取若干张风格模板图,如20张不同风格的风格模板图。相应地,本发明实施例在获取风格模板图的同时,还获取与该风格模板图对应的风格描述信息,从而得到相关的风格模板图,以及与其匹配的风格描述信息。

[0122] S220:从风格模板图中分离得到多个子图像。

[0123] 本发明实施例中,子图像是指从风格模板图中提取出的图像片段。其中,本发明实施例中子图像可以是风格模板图中的任意区域,其可以包括一个对象、一个特定的区域或者一个局部特征。示例性地,以一子图像构建场景为例,本发明实施例中子图像可以从原始的风格模板图中截取某些对象(如图中的动物对象或者物体对象等),从而分离得到多个子图像。或者,本发明实施例中还可以对风格模板图中不同区域进行分离,如以特定的长宽比例将原始风格模板图划分为多个不同的图像片段,从而得到多个子图像。又或者,本发明实施例中还可以对风格模板图中的局部特征进行分离,如对风格目标图中风格特征较为明显的区域进行分离,从而构建得到相应的子图像。

[0124] S230:将各个子图像的图像描述信息与风格描述信息结合,得到图像描述文本,并根据子图像和图像描述文本构建多个图文对。其中,每个图文对中包含子图像和子图像对应的图像描述文本。

[0125] 本发明实施例中图像描述信息是指子图像对应的特征描述信息,如子图像的特征信息、风格信息或者对象信息等。本发明实施例中图像描述文本是指通过自然语言描述的图像信息文本,包括图像的相关特征、风格类型等描述。另外,本发明实施例中图文对是指

文字和图像的融合,即子图像与其对应的图像描述文本的融合,从而将子图像与相应的图像描述文本相匹配。相应地,本发明实施例中每个图文对中都包含有子图像以及该子图像所对应的图像描述文本。示例性地,以一图像修复应用场景为例,本发明实施例中需要对一些已损坏的图像进行修复,因此可以通过图像重绘的方式对损坏的图像进行修复。相应地,在图像重绘模型训练过程中,本发明实施例对风格模板图进行分离,得到的相应的子图像,以及子图像对应的图像描述信息。接着,本发明实施例将各个子图像对应的图像描述信息与相应的风格描述信息进行结合,得到各个子图像的图像描述文本。进一步地,本发明实施例将子图像与相应的图像描述文本进行配对,从而构建得到相应的图文对。容易理解的是,本发明实施例的图文对中包含有子图像以及该子图像对应的图像描述文本,从而通过将图像和描述文本结合在一起的方式,能够更直观地传达信息,增强模型训练的准确性。

[0126] S240:在每轮训练过程中,从多个图文对中选取一个图文对的子图像作为待融合图像。

[0127] 本发明实施例中每轮训练,即每轮迭代,全量图像数据在模型中均被训练完一次称为一轮迭代。其中,本发明实施例对全量数据一共进行M轮(如10轮)迭代。容易理解的是,本发明实施例中,由于训练机器的显存资源有限,因此在每一轮迭代中,不能把全量样本一次性输入到模型中进行训练,需要对所有样本进行分批次(batch)训练。其中,每批次样本输入到模型中进行前向计算、后向计算以及模型参数更新等。相应地,本发明实施例中待融合图像是指从多个图文对中选出的训练图像样本。示例性地,以一游戏开发应用场景为例,在一些游戏开发应用过程中,可以通过图像重绘的方式对游戏场景或游戏角色进行设计和绘制,以提升有效画面的视觉效果,使得游戏场景或游戏角色模型风格更统一和逼真。相应地,在图像重绘模型训练过程中,本发明实施例在每轮训练过程中,从构建的多个图文对中随机选取一个图文对,并将该图文对中的子图像作为待融合图像,通过该待融合图像进行该轮次的模型训练。

[0128] 需要说明的是,本发明实施例在第一轮的第一个批次训练前需要对相关参数进行初始化。例如,对生成模型(VAE)、text_encoder、U-Net、新增的旁路模块采用已训练好的模型参数,如stable-diffusion v1-5,并且text_encoder、旁路模块在训练中更新。示例性地,本发明实施例中初始化采用0.0004的学习率,以后每5轮学习后学习率变为原来的0.1倍,共训练10轮。

[0129] S250:将待融合图像与噪声源图像结合得到待训练风格图像,并根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息。

[0130] 本发明实施例中噪声源图像是指图像中存在噪声的图像。其中,图像中存在的噪声可以表现为颗粒状噪点、条纹状噪声或其他形式的噪声干扰。相应地,去噪网络的网络控制信息是指用于控制和调整去噪网络的参数或行为的信息,例如,去噪网络的输入信号、网络结构、训练数据等。本发明实施例通过将待融合图像与噪声源图像结合,得到待训练风格图像。然后根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息。示例性地,以一新媒体应用场景为例,在一些宣发海报中,常常需要对海报进行相关调整,如海报风格、特定区域替换等,因此,需要对相关的图像进行重绘。相应地,本发明实施例在进行图像重绘模型训练过程中,从多个图文对中选取一个图文对的子图像作为待融合图像后,生成噪声源图像。其中,本发明实施例中噪声源图像可以通过随机抽取一个种子i来生成一张

噪声图,即噪声源图像。需要说明的是,本发明实施例中生成的噪声图可以是一个随机的噪声图像。接着,本发明实施例将待融合图像与该噪声源图像相加,从而得到待训练风格图像。同时,本发明实施例根据待融合图像,从相应的图文对中得到对应的风格描述信息,并根据风格描述信息生成得到去噪网络的网络控制信息。例如,本发明实施例中将待融合图像对应的风格描述信息经过文本编码器(text encoder),得到相应的文本表征信息,即网络控制信息。相应地,本发明实施例通过将文本表征信息输入生成模型中进行网络的控制。其中,本发明实施例中文本表征信息被作为键值(KV),隐空间表征作为查询(Q)。

[0131] S260:根据网络控制信息,由去噪网络对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果。

[0132] 本发明实施例中去噪网络是指用于去除图像数据中噪声的神经网络模型,本发明实施例通过去噪网络对待训练风格图像中的噪声进行预测。本发明实施例中预测结果是指去噪网络对待训练风格图像上的噪声进行预测,以预测得到对应的噪声图像。示例性地,以一产品包装应用场景为例,可以通过图像重绘的方式,使得产品包装更符合目标对象的需求。相应地,本发明实施例根据生成的去噪网络的网络控制信息,通过去噪网络预测待训练风格图像中的噪声。例如,本发明实施例中将待融合图像对应的风格描述信息经过文本编码器(text encoder),得到相应的网络控制信息后,将网络控制信息输入到生成模型中。接着,在键值(KV)的约束下,本发明实施例通过去噪网络模型,如denoising Unet模型,进行前向计算,以对当前时刻的隐空间表征(ZT)进行去噪操作,从而得到前一时刻的隐空间表征。例如,对于输入的是当前时刻ZT,则得到前一时刻ZT-1,而对于输入的是ZT-1,则得到ZT-2。然后,经过若干次去噪操作后,得到Z0的预测值,即噪声源图像的预测值。容易理解的是,本发明实施例通过解码器将Z0的预测值进行转换,从而将预测值转换成为预测的噪声图,即得到噪声源图像的预测结果。

[0133] S270:根据预测结果与噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新。

[0134] 本发明实施例中损失计算结果是指衡量模型预测值与真实值之间的差异的指标。其中,本发明实施例中损失计算可以通过相关的损失函数进行计算,如均方误差函数、交叉熵损失函数等。相应地,本发明实施例中模型参数是指控制模型的行为和性能的参数,通过对数据进行学习和优化从而找到相应的最佳参数值。本发明实施例通过对预测结果与噪声源图像进行损失计算,以统计该批次样本的总损失,得到相应的损失计算结果。接着,本发明实施例通过损失计算结果对图像重绘模型的模型参数进行更新。示例性地,以一图片编辑应用场景为例,在一些图片编辑的应用程序中,需要对一些图片进行修改,如对图像进行风格转换或替换图像中的特定区域的物体等。因此可以通过图像重绘的方式对相关的图片进行编辑。相应地,本发明实施例在对图像重绘模型训练过程中,首先计算预测图与原始输入噪声图像的损失,即预测结果与噪声源图像之间的损失,并统计该批次样本的总损失,得到相应的损失计算结果。接着,本发明实施例通过SGD随机梯度下降方法,将损失反向回传到模型中得到文本表征模块(如clip文本分支)、旁路模块的梯度,并更新图像重绘模型的模型参数。其中,随机梯度下降(Stochastic Gradient Descent,SGD)是梯度下降算法的一个扩展。由于批量梯度下降法在更新每一个参数时,都需要所有的训练样本,所以训练过程会随着样本数量的加大而变得异常的缓慢,而随机梯度下降法能够缓解批量梯度下降法这

一弊端。相应地,随机梯度下降是通过每个样本来迭代更新一次。

[0135] 需要说明的是,本发明实施例通过基于均方差 (Mean Squared Error, MSE) 作为L2损失函数,通过对预测结果与噪声源图像进行均方误差的计算来衡量模型的性能。其中,L2损失函数通过计算预测值与真实值之间的差异的平方和来衡量模型的拟合程度。相应地,本发明实施例中图像的预测损失计算公式如下式(1)所示:

$$[0136] \quad MSE = \sum_{i=1}^n (y_i - y_i^p)^2 \quad (1)$$

[0137] 其中,式中 n 为样本数量, y_i 为噪声源图像中各点像素值, y_i^p 为预测结果中的预测像素。

[0138] 为了能够提高图像描述的准确性,以提高图像重绘模型训练的准确性,在一些可行的实施例中,在将各个子图像的图像描述信息与风格描述信息结合,得到图像描述文本这一步骤中,可以包括步骤S310至步骤S330:

[0139] S310:对各个子图像进行描述提取,得到各个子图像的图像描述信息。

[0140] 本发明实施例中描述提取是指对子图像的内容进行文本描述提取。相应地,本发明实施例中图像描述信息是指对个子图像的图像内容的描述文本,如“空中飞船”等。本发明实施例在对风格模板图进行分离得到多个子图像后,为了提高图像描述的准确性,对各个子图像进行描述提取,进而得到各个子图像相应的图像描述信息。示例性地,以一子图像描述提取应用场景为例,本发明实施例通过深度学习模型对各个子图像的图像描述信息进行提取。例如,本发明实施例通过多模态视觉语言预训练 (Blender-based Language-Image Pre-training, BLIP) 模型对各个子图像进行描述提取。其中,本发明实施例中BLIP模型是一种通过联合训练图像和文本数据来学习视觉和语言表示的模型。因此,本发明实施例中BLIP模型可以理解 and 生成自然语言,处理与图像相关的任务,如提取图像的描述信息。相应地,本发明实施例首先将各个子图像输入图像编码器 (image encoder),提取图像的语义信息,通过提取图像中的特征来获取图像的视觉信息。接着,本发明实施例将提取到的信息通过交叉注意力机制模块和文字生成模块,以生成与子图像对应的语义描述,并通过交叉注意力机制模块控制生成的文本信息与图像的语义描述更加接近。需要说明的是,参照图4,本发明实施例中图像编码器 (image encoder) 为视觉转换器 (transformer) 模型分支,由 N 个自注意力机制 (self-attention) 以及前馈网络 (feed forward) 组成的注意力结构堆叠而成。其中,该模块基于在ImageNet数据集上进行预训练的ViT (Vision Transformer) 模型。相应地,本发明实施例中产生图像描述的模块,如图4中右侧结构,是一个以图像为基础的文本解码器 (Image-grounded text decoder),即图像驱动文本解码器,其通过因果自注意力层 (causal self-attention layer) 替代编码器中的双向自注意力层。另外,本发明实施例通过 [Decode] 标记 (解码标记) 来表示一个序列的开始。

[0141] S320:对风格描述信息进行表征获取,得到风格描述信息对应的风格表征信息。

[0142] 本发明实施例中风格描述信息是指与获取的风格模板图对应的描述词或风格词。相应地,本发明实施例中风格表征信息是指从风格描述信息中提取得到的高层次的抽象特征信息,能够进一步地反映图像的风格信息。本发明实施例通过对风格描述信息进行表征获取,以得到对应的风格表征信息,从而提高图像描述的准确性。示例性地,在一视频图像处理应用场景中,常常需要对视频中一些帧图像进行修改或增加相关的图像。此时,可以通过图像重绘的方式将相关的图像重绘至目标类型的图像。相应地,在图像重绘模型训练过

程中,本发明实施例通过预设的机器学习算法学习风格描述信息的特征表示,从原始的风格描述信息中学习得到相关风格描述的高层次特征,使得数据更易于处理和解释,从而得到风格描述信息对应的风格表征信息。

[0143] S330:在每个子图像的图像描述信息前加入风格表征信息,得到每个子图像对应的图像描述文本。

[0144] 本发明实施例中图像描述文本由图像描述信息和风格表征信息构成,通过在各子图像对应的图像描述信息前加入风格表征信息,从而构建得到各个子图像对应的图像描述文本。示例性地,以一图像描述文本构建应用场景为例,本发明实施例中提取到的某一子图像的图像描述信息为“空中飞船”,而通过对风格描述信息进行表征获取,得到对应的风格表征信息为“A风格”,则本发明实施例在图像描述信息“空中飞船”前加入风格表征信息“A风格”,从而构建得到该子图像对应的图像描述文本“A风格,空中飞船”。需要说明的是,本发明实施例在实际应用中图像描述文本采用英文文本,以便于模型应用。

[0145] 为了使得模型更容易学习得到相关的风格信息,在一些可行的实施例中,在对风格描述信息进行表征获取,得到风格描述信息对应的风格表征信息这一步骤中,可以包括步骤S410至步骤S420:

[0146] S410:将风格描述信息与风格词库中的各个风格词进行查询匹配,确定风格描述信息对应的候选文本表征信息。

[0147] 本发明实施例中风格词是指描述图像风格的词汇,如流行风格、现实风格、抽象风格等。相应地,风格词库是预先构建的词汇库,其存储有相关的风格描述的风格词。本发明实施例中候选文本表征信息是指候选的图像描述的相关风格词。本发明实施例通过将风格描述信息与风格词库中的各个风格词进行查询匹配,以确定风格描述信息所对应的候选文本表征信息。示例性地,以一图片处理应用场景为例,当需要获取风格描述信息对应的风格表征信息时,本发明实施例首先将风格描述信息与相应的风格词汇库中的各个风格词进行匹配,以匹配得到相关的风格词,如匹配到“B风格”、“C风格”以及“D风格”,即根据风格描述信息查询得到“B风格”、“C风格”以及“D风格”与图像的描述相关,则本发明实施例将这些风格词作为候选的词汇,构建得到风格描述信息对应的候选文本表征信息。

[0148] S420:将候选文本表征信息与各个子图像的图像描述信息结合,确定风格描述信息对应的风格表征信息。

[0149] 本发明实施例通过将候选文本表征信息与各个子图像的图像描述信息进行结合,从而确定各个子图像对应的风格表征信息,即确定风格描述信息对应的风格表征信息。示例性地,以一游戏开发应用场景为例,在游戏场景建模过程中,常常需要对一些场景进行重绘和修复,以提高游戏场景开发的效率,并提高游戏场景的逼真度和视觉效果。相应地,在进行图像重绘模型训练过程中,本发明实施例首先获取相关的游戏场景风格图像以及相应的风格描述信息,并将游戏场景风格图像分离得到多个子图像。同时,本发明实施例将游戏场景风格图像的风格描述信息与风格词库中的风格词进行匹配,以确定相应的候选文本表征信息。另外,本发明实施例将候选文本表征信息与分离得到的各个子图像进行结合,从而确定风格描述信息对应的风格表征信息,实现对图像风格描述信息的微调,通过词汇库中的风格词与风格模板图进行联合初始化,能够使得在整个生成模型词汇空间中找到与待训练风格词汇更接近的词作为初始化描述,通过初始化的词汇与其所属性质(物体、风格等)

更接近,使得图像重绘模型能够更容易地学习图像中的风格信息,提高模型学习的准确性。

[0150] 为了提高风格词查询匹配的效率和准确性,在一些可行的实施例中,本发明实施例提供的图像重绘模型训练方法还包括风格词库的步骤,可以包括步骤S510至步骤S530:

[0151] S510:收集多种风格词以及各种风格词对应的风格图像。

[0152] 本发明实施例中风格图像是指不同风格的样本图像。相应的,本发明实施例中收集的风格词是指描述相关风格图像的图像风格的词汇。本发明实施例在构建风格词库时,首先收集多种风格词,以及各种风格词对应的风格图像。示例性地,以一风格词库构建场景为例,本发明实施例中通过互联网对相关的风格词进行收集。同时,而本发明实施例还收集各个风格词对应的模板风格图像。例如,当本发明实施例收集到风格词“Q风格类型”,则例针对该风格词“Q风格类型”,本发明实施收集若干张(如20张)对应的“Q风格类型”的风格图像。需要说明的是,对于收集到的一些风格词,当收集不到相应数量的风格图像时,本发明实施例通过从生成模型中输入对应的风格词和一些其他描述词,例如“Y风格,田野和稻草人”或“Y风格,花园”等,以产生具有相应风格的图像。

[0153] S520:获取每种风格词的文本表征向量。

[0154] 本发明实施例中文本表征向量是将风格词文本转化为向量形式的表示,用于表示风格词的语义和特征。其中,本发明实施例中文本表征向量是通过将风格词映射到向量空间中的点来实现的,将风格词转化为固定维度的向量,从而捕捉到风格词的语义信息和特征。示例性地,以一文本表征向量应用场景为例,本发明实施例通过CLIP文本表征对获取到的各个风格词进行文本表征向量提取。其中,CLIP(Contrastive Language-Image Pre-training)文本表征是一种用于将文本和图像统一表示的模型。其通过对大量的图像和文本数据进行预训练,从而学习到图像和文本之间的关联性。相应地,CLIP文本表征的目标是将图像和文本映射到一个共享的特征空间中,使得相似的图像和文本在该空间中的表示也相似。本发明实施例通过CLIP文本表征获取每种风格词的文本表征向量的方式,以将文本和图像统一表示为一个固定长度的向量,从而能够将该文本表征向量用于后续风格词库的构建。

[0155] S530:根据文本表征向量的风格图像,构建风格词库。其中,风格词库的索引项为各个风格词,每个索引项关联风格词的文本表征向量。

[0156] 本发明实施例中索引项是指风格词库中用于快速查找和访问数据的数据结构,通过该索引项能够查询得到相应的数据信息。本发明实施例通过设置索引项,以加快风格词查询的速度。具体地,本发明实施例根据获取到的文本表征向量的风格图像,构建风格词库。相应地,本发明实施例中将各个风格词作为风格词库的索引项,并且每个索引下关联各个风格词库对应的文本表征向量,从而构建得到风格词库。示例性地,以一风格词库构建应用场景为例,本发明实施例在通过CLIP文本表征对获取到的各个风格词进行文本表征向量提取,得到文本表征向量后,本发明实施例首先根据文本表征向量对应的风格图像构建相关的风格词库。同时,本发明实施例构建的风格词库中将各个风格词作为索引,并且将各个风格词对应的文本表征向量关联至相应的索引下。例如,本发明实施例中风格词“K风格”,其相应的文本表征向量为“k”,则本发明实施例将该风格词“K风格”作为索引项,并将其相应的文本表征向量“k”关联至该索引下,从而构建得到相应的风格词库。

[0157] 为了能够提高风格表征信息的准确性,从而提高图像重绘模型训练的准确性,在

一些可行的实施例中,在将候选文本表征信息与各个子图像的图像描述信息结合,确定风格描述信息对应的风格表征信息这一步骤中,可以包括步骤S610:

[0158] S610:获取候选文本表征信息与图像描述信息之间的中点值,作为风格表征信息。

[0159] 本发明实施例中风格表征信息用于表征子图像的图像约束信息和风格描述信息的文字约束信息。相应地,本发明实施例中的中点值是指候选文本表征信息与图像描述信息之间的距离中点。本发明实施例将候选文本表征信息与图像描述信息之间的中点值作为风格描述信息对应的风格表征信息。示例性地,参照图5,在一风格表征信息获取应用场景中,本发明实施例首先对获取的风格模板图进行分割,得到多个子图像。接着,对各个子图像进行描述提取,从而得到各个子图像对应的图像描述信息,并对图像描述信息进行表征获取,得到风格表征信息。进一步地,本发明实施例对风格图像全图,即风格模板图,提取风格描述信息,并在风格词库中进行风格词的查询匹配,得到相应的候选文本表征信息。接着,本发明实施例将候选文本特征信息与图像描述信息之间的中点值作为风格表征信息。例如,当以获得风格图像特征信息,即风格表征信息,以及最接近的参考风格词特征,即候选文本表征信息,则本发明实施例取所有向量点的中点作为风格表征信息,如四个向量点a、b、c、d的中点为 $(a + b + c + d)/4$ 。其中,本发明实施例中获取到的风格词经过上述初始化后,在文本编码器的输出中能产生目标嵌入(emb_target)表征,即该目标嵌入表征融合了由图像约束的目标生成风格信息(图像约束信息),以及由文字约束的相关生成风格信息(文字约束信息)。容易理解的是,如图5所示,其中已经获得风格图像特征信息为二维坐标系中的圆点,最接近的参考风格词特征为图中的五边形点,取所有点的中点,即三角形点,作为初始化embedding表征目标,从而实现风格表征初始化。需要说明的是,图5中“Q风格”是通过查询风格词库中的风格词,如“Y风格”、“U风格”、“P风格”、“R风格”等,从而得到的最接近的风格词。

[0160] 为了实现对语义控制网络的更新,在一些可行的实施例中,本发明实施例提供的图像重绘模型训练方法还可以包括步骤S710:

[0161] S710:根据风格表征信息,对语义控制网络进行更新。

[0162] 本发明实施例中语义控制网络是用于生成去噪网络的网络控制信息的神经网络。本发明实施例通过风格表征信息对语义控制网络进行更新。具体地,参照图6,本发明实施例将文本text输入训练(conditioning)模块中的 τ 结构中,以通过分词器(tokenizer)将输入的文本进行分解,得到token单元。其中,本发明实施例中 τ 结构为CLIP文本模型的文本输入分支。相应地,本发明实施例中token单元可以是一个单词、字符或子词。同时,本发明实施例中分词器在分解得到相应的token单元后,为每个token单元分配一个唯一的标记ID(token_IDs)。接着,本发明实施例对相应的token_IDs提取文本嵌入信息,从而得到相应的文本表征,即网络控制信息。需要说明的是,在初始得到风格表征信息后,U-net以及 τ 结构都会被更新。其中, τ 结构的工作过程是先产生标记ID,再产生嵌入向量(embedding),微调时嵌入向量提取(即新增图的transformer模块)会被微调使之更适应风格图像生成所需要的表征要求。在本实施例得到风格表征信息的过程中,以图6的Y风格为例,基于Y风格风格模板图,对输入的文本描述进行对应的结合,得到“Y风格,两个人坐在草地上”的图像文本描述,后续则基于此文本描述进行相应图像的生成。但本发明实施例基于更优的文本表征初始化,将文本表征初始化为嵌入向量(embedding),因此需要反向找到该初始化的标记

ID(token_IDs),以允许生成模型进行相应的微调训练过程来训练更新 τ 结构。示例性的,以一语义控制网络更新应用场景为例,参照图7,本发明实施例首先把相关文本词映射到词汇表中,即对比语言-图像预训练(CLIP)文本模型的分词(tokenizer)过程。例如,对于一个词汇,如“an apple”,首先查找其中的“an”对应的词典id,如1000,然后查找“apple”对应的id,如3500。接着,本发明实施例通过基于transformer的文本表征抽取结构对每一个词汇进行抽取,产生若干个相应的嵌入向量,如 $k \times 768$ 个。其中,本发明实施例中 k 为该文本词对应的词汇表id数量,一个词在词汇表中可能对应多个词汇id,因为某个词可能有多个不同的含义,此时需要更多的词汇id数量来明确该词属于哪个词汇的意思,因此需要多于1个词汇。容易理解的是,通过上述步骤得到的目标嵌入(emb_target)表征为期望的初始化结果,如图7中的输出文本表征。另外,本发明实施例还需要对得到的目标嵌入表征进行tokens反查,通过反向查找得到该目标嵌入表征的标记ID(token_IDs),以使得生成模型能够按照上述的微调训练过程更新 τ 结构,实现对语义控制网络的更新。

[0163] 示例性地,本发明实施例通过CLIP文本模型提取文本表征的代码中,可以包括“self.transformer=CLIPTextModel.from_pretrained(version).eval()”、“self.device=device”、“batch_encoding=self.tokenizer(prompts,truncationl,max_lengthself.max_length,return_lengthl,return_overflowing_tokens=0,padding=”max_length,return_tensors=”pt”)”、“tokens=batch_encoding[”input_ids”].to(self.device)”以及“return self.transformer(input_ids=tokens).last_hidden_state)”等代码语句。其中,在本发明实施例的代码应用实施例中,本发明实施例首先定义一个类,如CLIPTextEmbedder,其可以继承至nn.Module类。接着,本发明实施例通过import语句引入相应的库和模块,如list、nn、CLIPTokenizer和CLIPTextModel等。进一步地,本发明实施例定义构造函数__init__,以初始化类的实例。其中,本发明实施例构造函数接受三个参数version、device和max_length,并将其设置为类的属性。例如,本发明实施例中version参数表示CLIP模型的版本,默认为“openai/clip-vit-large-patch14”;device参数表示模型运行所在的设备,默认为“cuda:0”;max_length参数表示输入文本的最大长度,默认为77。接着,本发明实施例通过调用父类nn.Module的构造函数,以确保类的初始化正常进行。然后,本发明实施例创建一个CLIPTokenizer对象,并调用from_pretrained方法来加载预训练的CLIPTokenizer模型,并将其赋值给self.tokenizer属性。参数version指定了要加载的预训练模型的版本。进一步地,本发明实施例创建一个CLIPTextModel对象,并调用from_pretrained方法来加载预训练的CLIPTextModel模型,并将其赋值给self.transformer属性。其中,eval()方法将模型设置为评估模式,即在推理时不进行梯度计算。接着,本发明实施例通过代码“self.device=device”将构造函数传入的device参数赋值给self.device属性,表示模型所在的设备,并通过“self.max_length = max_length”将构造函数传入的max_length参数赋值给self.max_length属性,表示输入文本的最大长度。然后,本发明实施例调用self.tokenizer的方法,对输入的文本进行处理和编码,并使用CLIPTokenizer的方法将文本编码为模型所需的input_ids张量,包含了文本的标记化表示。进一步地,本发明实施例从batch_encoding字典中获取编码后的input_ids张量,并将其移动到指定的设备上,然后赋值给tokens。最后,本发明实施例调用self.transformer的方法,传入tokens作为输入,并返回模型的输出结果last_hidden_

state,即文本嵌入的表示(文本表征)。

[0164] 为了能够提高token查找的准确性和效率,从而实现语义控制网络的更新,在一些可行的实施例中,在根据风格表征信息,对语义控制网络进行更新这一步骤中,可以包括步骤S810至步骤S850:

[0165] S810:获取图像文本匹配模型的词汇表以及词汇表中各个词汇的词汇标识。其中,图像文本匹配模型用于根据输入的图像确定对应的文本描述信息。

[0166] 本发明实施例中图像文本匹配模型是指将图像与相应的文本进行匹配的模型。其中,图像文本匹配模型通过会学习图像和文本之间的语义关联,从而判断给定图像和文本是否相关。本发明实施例通过将图像输入图像匹配模型,以根据输入的图像确定对应的文本描述信息。相应地,本发明实施例中词汇表是相关图像描述文本词汇的集合,其中每一个词汇为一个token。另外,本发明实施例中词汇标识是指词汇表中各个词汇对应的标记ID(token_IDs)。其中,由于同一个词汇可能存在多种不同的含义,因此本发明实施例的词汇表中一个词汇可能对应多个词汇id,即词汇标识。示例性地,以一词汇表构建应用场景为例,在进行token反查找过程中,本发明实施例首先构建相应的词汇表。其中,本发明实施例中图像文本匹配模型包括CLIP文本模型。本发明实施例首先获取相关图像描述的所有词汇,即获取全部tokens,如49152个,构建得到词汇表。同时,本发明实施例根据获取的词汇,获取相应词汇token对应的词汇标识,即标记ID,将词汇标识与相应的词汇进行关联,从而完成CLIP文本模型(图像文本匹配模型)的词汇表构建。需要说明的是,本发明实施例中token是CLIP文本模型训练的一个词汇表(共49152个词汇)。其中,词汇与词的概念不同,如happiness可以表示幸福、开心,而当表示开心时可以与enjoying共享一个词汇,当表示幸福时可以与pleasure共享一个词汇,从而将文本句子映射到一个token列表。相应地,本发明实施例中并非一个词对应一个token,即不是一一对应的,但token与embedding是一一对应的,即一个token对应一个embedding。

[0167] S820:对各个词汇标识进行文本表征推理,得到每个词汇标识对应的初始文本表征信息。

[0168] 本发明实施例中文本表征推理是指通过图像文本匹配模型对词汇标识进行预测和推理的过程。例如,本发明实施例通过CLIP文本模型对各个词汇标识进行文本表征推理,以初始化得到相应的文本表征信息,即初始文本表征信息。其中,本发明实施例中初始文本表征信息是指各个词汇标识的嵌入向量表示(embedding)。示例性地,在一游戏开发应用场景中,当构建得到相应的词汇表后,本发明实施例对词汇表中的各个词汇标识进行CLIP文本模型的文本表征推理,即对词汇表中的所有token_IDs进行CLIP文本模型推理,以得到初始化的文本表征矩阵emb_all。其中,该文本表征矩阵由各个词汇标识对应的初始文本表征信息构成,该文本表征矩阵emb_all是一个N*K的向量,N是指token的数量,K是指每个token的特征维度数,例如当词汇表中全部词汇的数量为49152,每个词汇的特征维度为768,则得到的文本表征矩阵emb_all为一个49152*768的向量矩阵。需要说明的是,本发明实施例中clip文本模型中的transformer模块包括clip Text Transformer模型,其通过该模型的推理过程,根据输入的token,输出相应的嵌入向量表示embedding。另外,本发明实施例中得到文本表征矩阵emb_all的输出结果前,需要对其进行正则化处理,以使得向量的数值范围在预设区间内,从而提高模型的稳定性以及性能。

[0169] S830:计算各个初始文本表征信息和风格表征信息之间的欧氏距离。

[0170] 本发明实施例中风格表征信息是指前面步骤中通过对风格描述信息进行表征获取到的风格表征信息,即从风格描述信息中提取得到的高层次的抽象特征信息。相应地,本发明实施例中欧氏距离是一种距离定义,表示在m维空间中两个点之间的真实距离,能够衡量多维空间中两个点之间的绝对距离。本发明实施例中通过欧氏距离计算文本表征矩阵emb_all中各个初始文本表征和风格表征信息之间的距离,以通过计算得到的欧氏距离确定对应的风格词汇。示例性地,以一欧氏距离计算场景为例,本发明实施例中通过欧氏距离(L2距离)算法,如 $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$,其中 (x_1, y_1) 为各个初始文本表征信息对应的向量, (x_2, y_2) 为风格表征信息对应的向量。例如,当文本表征矩阵emb_all为一个49152*768的向量矩阵,则本发明实施例通过欧氏距离算法对各个词汇对应的初始文本表征信息与风格表征信息之间的距离进行计算,从而计算得到各个词汇与风格表征信息之间的距离,得到49152个距离数据。

[0171] S840:根据欧氏距离,从各个初始文本表征信息确定风格表征信息对应的风格词汇标识。

[0172] 本发明实施例中风格词汇标识是指通过欧氏距离确定的风格表征信息对应的词汇表中词汇的词汇标识。本发明实施例通过计算得到的各个初始文本表征信息与风格表征信息的欧氏距离,确定风格表征信息对应的风格词汇标识。示例性地,以一风格词汇标识获取应用场景为例,本发明实施例将计算得到的各个距离,按距离大到小,或小到大,进行排序,并取其中距离最小的预设数量的词汇,得到相应的风格词汇标识,即相应词汇对应的token_IDs。其中,本发明实施例中预设数量为预设对风格词约束,如预设数量的词汇可以为4个,即将距离最小的4个词汇来表达新的风格词。需要说明的是,本发明实施例中还可以通过设置距离阈值的方式,确定风格表征信息对应的风格词汇标识。例如,当确定某一词汇的初始文本表征信息与风格表征信息之间的欧氏距离小于设置的距离阈值,则确定该词汇为相应的风格词汇。

[0173] S850:根据风格词汇标识,对语义控制网络进行更新。

[0174] 本发明实施例根据确定的风格表征信息对应的风格词汇标识,对语义控制网络进行更新,从而实现对语义控制网络结构的训练更新。示例性地,在一语义控制网络更新应用场景中,本发明实施例通过反向查找得到对应的风格词汇标识的方式,确定对应的初始化词汇标识,从而使得语义控制网络在进行训练时,能够正常进行更新优化,实现对语义控制网络的更新。

[0175] 为了提高模型噪声预测的准确性,在一些可行的实施例中,在根据网络控制信息,由去噪网络对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果这一步骤中,可以包括步骤S910至步骤S920:

[0176] S910:对待训练风格图像进行扩散采样处理,得到待训练风格图像的隐空间表征信息。

[0177] 本发明实施例中扩散采样处理是通过减少图像中的像素数量来降低图像的细节和清晰度,从而减小图像的尺寸的处理方法,可以用于图像压缩、图像缩放和图像降噪等应用中。相应地,本发明实施例中隐空间表征信息是指在机器学习或深度信息中,通过学习模型自动地从原始数据中提取出的一组潜在变量或特征表示。其中,这些潜在变量或特征表

示可以捕捉到数据中的关键信息和结构,从而能够更好地描述和表示数据。本发明实施例通过对待训练风格图像进行扩散采样处理的方式,以得到待训练风格图像的隐空间表征信息。示例性的,以一噪声源图像预测应用场景为例,本发明实施例对相关的图像,如训练的图文对中的图像,添加噪声后,进行编码,如通过变分自编码器进行编码,得到相应的编码结果。其中,本发明实施例中变分自编码器(Variational Autoencoder)是一种生成模型,其能够从输入数据中学习潜在变量,并生成新的样本。接着,本发明实施例将相应的编码结果经过扩散(diffusion)过程,从而得到T时刻的隐空间表征信息。其中,本发明实施例中diffusion扩散过程的扩散抽样用于将噪声图像VAE编码的特征映射到T时刻的隐空间表征。相应地,通过后续对图像的去噪过程学习产生噪声表征的拟合,从而通过原图减去噪声表征得到真实需要的图像表征,并经过相应的解码器D得到真实需要的图像,即预测结果。

[0178] S920:根据网络控制信息,由去噪网络对隐空间表征信息进行去噪处理,得到对噪声源图像的预测结果。

[0179] 本发明实施例中去噪处理是指减少图像中的噪声的过程。相应地,本发明实施例中网络控制信息是指根据待融合图像对应的风格描述信息生成的文本表征。本发明实施例根据去噪网络生成的网络控制信息,通过去噪网络对隐空间表征信息进行去噪处理,从而实现了对噪声源图像的预测,得到预测结果。示例性地,在一去噪应用场景中,本发明实施例将经过CLIP文本模型得到的文本表征(embedding)作为去噪网络的控制输入,即将网络控制信息作为U-net网络的交叉注意力查询-键-值(QKV)结构的控制输入。同时,本发明实施例将获取得到的T时刻的隐空间表征,通过去噪网络(denoising U-Net)进行T次去噪操作,以还原得到图像的原特征,即未加噪声的原图特征。相应地,本发明实施例将还原得到的特征经过相应的解码过程,如VAE解码器,从而解码得到相应的原图像,即得到对噪声源图像的预测结果。

[0180] 需要说明的是,参照图8、图9以及图10,本发明实施例中一个查询-键-值(QKV)结构中包含了多个堆叠的残差块(resBlock)结构和空间变换网络(Spatial Transformer Networks, STNs)结构。其中,本发明实施例U型网络(U-Net)中的QKV结构含有两个空间变换网络结构,每个空间变换网络结构为一个QKV过程,在第一个QKV过程中KV与输入的Q相同(Q为上一网络结构的输出),在第二个QKV过程中KV为用于控制生成的文本嵌入(embedding)约束。

[0181] 为了提高模型噪声预测的可靠性以及稳定性,在一些可行的实施例中,在根据网络控制信息,由去噪网络对隐空间表征信息进行去噪处理,得到对噪声源图像的预测结果这一步骤中,可以包括步骤S1010至步骤S1040:

[0182] S1010:构建旁路注意力矩阵约束模块。

[0183] 本发明实施例中旁路注意力矩阵约束模块是一个进行新风格信息学习的KV旁路约束模块,以防止过拟合问题。示例性地,在一旁路注意力矩阵约束模块构建场景中,参照图11,本发明实施例中旁路注意力矩阵约束模块中的QKV结构与图中U-net模块中的QKV结构相同,并且预训练参数也相同。但是,本发明实施例在训练过程中旁路注意力矩阵约束模块的参数会被更新,而U-net模块的参数则不更新。相应地,本发明实施例中旁路注意力矩阵约束模块的QKV结构中的查询模块(Q)与主路中的查询模块(Q)一致,即与U-net模块中的Q一致,为某时刻的隐空间表征,如 $[T, T-1, T-2, \dots, 1]$ 时刻的表征。同时,本发明实施例中旁

路注意力矩阵约束模块中包括有若干个QKV结构,且每个QKV结构与U-net结构中相应的QKV结构相连接。本发明实施例中旁路注意力矩阵约束模块的QKV结构中KV是纯风格表征,通过QKV结构后得到行的风格表征,并将旁路注意力约束模块的输出与U-net模块的QKV结构输入进行融合。

[0184] S1020:根据旁路注意力矩阵约束模块对风格描述信息进行前向计算,得到旁路注意力权重。

[0185] 本发明实施例中前向计算是一种神经网络的计算方式,是指从输入层到输出层的一次完整计算过程。其中,在前向计算中,神经网络根据接收的输入数据进行一系列的计算和传递,并将数据逐层地传递给下一层,得到输出结果。相应地,在每一层中,神经网络对输入数据进行线性变换和非线性激活操作,以提取和转换数据的特征。另外,在前向计算过程中,神经网络的参数(权重和偏置)会被用来调整输入数据的权重,以便更好地拟合训练数据和预测未知数据。本发明通过旁路注意力矩阵约束模块对风格描述信息进行前向计算,以得到旁路注意力权重。示例性地,在一旁路注意力矩阵约束模块应用场景中,如图11所示,本发明实施例的旁路注意力矩阵约束模块中设置有多个QKV结构,如查询-键-值7、查询-键-值8以及查询-键-值9。其中,本发明实施例中查询-键-值7、查询-键-值8以及查询-键-值9依次连接,并且查询-键-值7与U型网络(U-net)模块中的查询-键-值1连接,查询-键-值8与U-net模块中的查询-键-值2连接,查询-键-值9与U-net模块中的查询-键-值3连接。当旁路注意力矩阵约束模块接收到相应的风格描述信息输入时,本发明实施例中旁路注意力矩阵约束模块根据接收到的风格描述信息依次通过查询-键-值7、查询-键-值8以及查询-键-值9,进行前向计算,从而得到旁路注意力权重。

[0186] S1030:将旁路注意力权重融合到去噪网络中的主路注意力矩阵约束模块,对主路注意力矩阵约束模块的注意力矩阵进行更新。

[0187] 本发明实施例中主路注意力矩阵约束模块是指与旁路注意力矩阵约束模块相连接的生成模型结构,如U型网络(U-net)模块。本发明实施例通过将生成的旁路注意力权重融合到去噪网络中的主路注意力矩阵约束模块,从而对主路注意力矩阵约束模块的注意力矩阵进行更新。示例性地,以一主路注意力矩阵模块更新的应用场景为例,本发明实施例的旁路注意力矩阵约束模块在接收到相应的风格描述信息时,将风格描述信息输入旁路注意力矩阵约束模块中相应的QKV结构中,进行前向计算,得到旁路注意力权重,同时本发明实施例还将相应的旁路注意力权重输入到主路注意力矩阵约束模块中相应的QKV结构中进行加权。例如,参照图11,本发明实施例中旁路注意力矩阵约束模块QKV7(查询-键-值7)与主路注意力矩阵约束模块中的QKV1(查询-键-值1)连接,旁路注意力矩阵约束模块QKV8(查询-键-值8)与主路注意力矩阵约束模块中的QKV2(查询-键-值2)连接,则QKV7结构输出除了输入到QKV8中进行网络的前向计算外,还按预设的加权方式,与融合得到QKV2的输入数据,如根据式子 $out_QKV1 + a * out_QKV7 = input_QKV2$ 产生QKV2的输入。其中, out_QKV1 为QKV1的输出数据, out_QKV7 为QKV7的输出数据, a 为比例常数。本发明实施例通过将旁路注意力矩阵约束模块的输出加权到主路注意力矩阵约束模块的方式,从而实现对主路注意力矩阵约束模块阻力矩阵的更新。

[0188] S1040:通过主路注意力矩阵约束模块对隐空间表征信息进行去噪处理,得到对噪声源图像的预测结果。

[0189] 本发明实施例中主路注意力矩阵约束模块是进行旁路注意力权重融合,更新后的主路注意力矩阵约束模块。相应地,本发明实施例通过该主路注意力矩阵约束模块对扩散采样处理,得到的待训练风格图像相应的隐空间表征信息进行去噪处理,从而得到对噪声源图像的预测结果。示例性地,在一去噪处理应用场景中,本发明实施例将CLIP文本模型处理得到的文本表征作为主路注意力矩阵约束模块中QKV结构的控制输入。同时,本发明实施例将获取得到的隐空间表征信息,通过更新后的主路注意力矩阵约束模块进行去噪处理,从而还原得到未加噪声的原图特征。接着,本发明实施例通过对预测得到的噪声源图像特征进行解码,从而得到预测的噪声源图像,即得到噪声源图像的预测结果。

[0190] 示例性地,以视频平台应用场景为例,对本申请技术方案中图像重绘模型训练方法的完整实施过程说明如下:

[0191] 在本发明实施例中,相关的视频平台通常需要对视频进行编辑处理,如通过对视频流中的图像帧进行风格转换或对图像中的某一区域进行修改替换。因此,可以通过训练图像重绘模型的方式,以通过训练好的图像重绘模型执行相关的视频编辑处理,从而能够提高图像重绘的准确性,进而提高视频重绘的准确性和效率。在视频平台的图像重绘模型训练过程中,本发明实施例首先获取相关视频的风格模板图以及风格模板图对应的风格描述信息。例如,本发明实施例中不同的风格模板图中包含的设计元素以及风格特征不同,如现实风格的模板图、抽象风格的模板图或漫画风格模板图等。相应地,在获取不同风格风格模板图的同时,本发明实施例还获取各种风格目标图对应的风格描述信息,如现实风格、抽象风格以及漫画风格等,从而得到相应的风格模板图以及与其匹配的风格描述信息。接着,本发明实施例从获取得到的风格模板图中分离得到多个子图像。本发明实施例通过对各个风格模板图进行裁剪,如对左中右、上中下九个位置进行裁剪,从而得到多个子图像。另外,本发明实施例还可以从风格模板图中截取相应的目标对象,如图中的相关物体对象等,从而分离得到多个子图像,或者还可以对风格模板图中特定的局部特征进行分离,从而得到相应的子图像。

[0192] 进一步地,本发明实施例对各个子图像进行提取描述,以得到各个子图像的图像描述信息。例如,本发明实施例通过多模态视觉语言预训练(BLIP)模型对各个子图像进行描述提取,从而得到相应的图像描述信息。接着,本发明实施例将风格描述信息与风格词库中的各个风格词进行查询匹配,以确定风格描述信息对应的候选文本表征信息。然后,本发明实施例将获取得到的候选文本表征信息与图像描述信息之间的中点值作为风格表征信息。需要说明的是,本发明实施例中风格表征信息用于表征子图像的图像约束信息和风格描述信息的文字约束信息。另外,本发明实施例中还需要构建风格词库。相应地,本发明实施例收集多种风格词以及各种风格词对应的风格图像。例如,本发明实施例可以通过相应的互联网收集相应的图像风格描述词汇,即风格词,并且本发明实施例还收集相关风格词对应的风格图像,如对每个风格词收集若干张相应的风格图像。进一步地,本发明实施例获取每种风格词的文本表征向量,如通过CLIP文本模型对获取得到的各种风格词进行文本表征向量提取,以将文本和图像统一表示为一个固定长度的向量,以便于后续将该文本表征向量用于风格词库的构建中。接着,本发明实施例根据文本表征向量的风格图像,构建得到风格词库。需要说明的是,本发明实施例将风格词作为风格词库的索引,并且各个风格词与其对应的文本表征向量相关联。接着,本发明实施例在每个子图像的图像描述信息前加入

风格表征信息,从而得到每个子图像对应的图像描述文本。例如,当提取得到的某一子图像的图像描述信息为“小汽车”,而通过对风格描述信息进行表征获取,得到对应的风格表征信息为“F风格”,则本发明实施例在图像描述信息“小汽车”前加入风格表征信息“F风格”,从而构建得到该子图像对应的图像描述文本“F风格,小汽车”。

[0193] 同时,本发明实施例将各个子图像与图像描述文本相结合,构建得到多个图文对。其中,本发明实施例中每个图文对中包含子图像和子图像对应的图像描述文本。例如,本发明实施例将子图像与相应的图像描述文本进行配对,从而构建得到相应的图文对。接着,在每一轮训练过程中,本发明实施例从多个图文对中选取一个图文对的子图像作为待融合图像,并将待融合图像与噪声源图像进行结合,得到待训练风格图像。同时,本发明实施例根据待融合图像对应的风格描述信息生成去噪网络的网络控制信息。容易理解的是,由于训练机器的显存资源有限,难以将全量样本一次性输入到模型中进行训练,因此需要对所有样本进行分批次(batch)训练,本发明实施例在每轮训练过程中,随机地从多个图文对中选取一个图文对,将该图文对中的子图像作为待融合图像,进行该轮次的模型训练。相应地,本发明实施例可以通过随机抽取一个种子来生成一张噪声图,并将待融合图像与该噪声源图像相加,从而得到待训练风格图像。同时,本发明实施例将待融合图像对应的风格描述信息经过文本编码器(text encoder),得到相应的网络控制信息。接着,本发明实施例对待训练风格图像进行扩散采样处理,得到待训练风格图像的隐空间表征信息。同时,本发明实施例通过构建旁路注意力矩阵约束模块,以根据旁路注意力矩阵约束模块对风格描述信息进行前向计算,得到旁路注意力权重。然后,本发明实施例将旁路注意力权重融合到去噪网络中的主路注意力矩阵约束模块,从而对主路注意力矩阵约束模块的注意力矩阵进行更新。进一步地,本发明实施例通主路注意力矩阵约束模块对隐空间表征信息进行去噪,从而得到噪声源图像的预测结果。

[0194] 进一步地,本发明实施例根据噪声源图像的预测结果与噪声源图像之间的损失计算结果,对图像重绘模型的参数进行更新。例如,本发明实施例通过计算预测结果与噪声源图像之间的损失,并统计该批次样本的总损失,从而得到相应的损失计算结果,同时,本发明实施例通过SGD随机梯度下降方法,将损失反向回传到模型中得到文本表征模块、旁路模块(旁路注意力矩阵约束模块)的梯度,并更新图像重绘模型的模型参数。另外,本发明实施例中去噪网络的网络控制信息有语义控制网络生成。相应地,本发明实施例通过风格表征信息对语义开网络进行更新。具体地,本发明实施例首先获取图像文本匹配模型的词汇表以及词汇表中各个词汇对应的词汇标识。需要说明的是,本发明实施例通过将相应的图像输入图像文本匹配模型以确定对应的文本描述信息。接着,本发明实施例对各个词汇标识进行文本表征推理,从而得到各个词汇标识对应的初始文本表征信息。例如,本发明实施例通过CLIP文本模型对各个词汇标识进行文本表征推理,得到相应的初始文本表征信息。然后,本发明实施例通过计算各个初始文本表征信息与风格表征信息之间的欧氏距离,以根据得到的欧氏距离从各个初始文本表征信息中确定风格表征信息对应的风格词汇标识。例如,将计算得到的各个欧氏距离进行排序,并取最小的若干个初始文本表征信息,从而得到对应的风格词汇标识。接着,本发明实施例根据得到的风格词汇标识对语义控制网络进行更新。

[0195] 如图12所示,本发明实施例还公开了一种图像重绘方法,该方法可以包括但不限于

于步骤T001至T006:

[0196] T001:获取待重绘图像和待重绘图像对应的风格描述信息。

[0197] 本发明实施例中待重绘图像是指需要进行重绘的目标图像。相应地,本发明实施例中风格描述信息是指对该待重绘图像进行相应重绘的风格描述文本,如相关的描述词或风格词。示例性地,在一图像重绘应用场景中,本发明实施例首先获取需要进行重绘的图像,即待重绘图像。相应地,本发明实施例还获取该待重绘图像对应重绘描述信息,即风格描述信息。例如,本发明实施例中待重绘图像对应的风格描述信息可以为“一只老鼠飞在天空中,K风格”。

[0198] T002:从待重绘图像中分离得到多个子图像。

[0199] 本发明实施例中子图像是通过对待重绘图像进行分离提取得到的图像片段。本发明实施例在获取得到待重绘图像后,通过对待重绘图像进行分离的方式,以得到多个子图像。示例性的地,在一待重绘图像分离应用场景中,本发明实施例通过对待重绘图像进行预设比例的裁剪,如按照预设的长宽度对待重绘图像进行裁剪,从而将待重绘图像分离为多个子图像。或者,本发明实施例可以对待重绘图像中特定的对象进行截取,如图像中的物品、背景等,从而从带重绘图像中分离得到多个子图像。又或者,本发明实施例还可以通过对待重绘图像中的局部特征进行分离,以构建得到相应的子图像,如提取待重绘图像中风格特征较为明显的局部区域。

[0200] T003:将各个子图像的图像描述信息与风格描述信息结合,得到图像描述文本。

[0201] 本发明实施例中通过将各个子图像的图像描述信息与风格描述信息进行结合,从而构建得到图像描述文本。示例性地,以一图像描述文本构建场景为例,本发明实施例首先通过对各个子图像进行描述提取得到各个子图像的图像描述信息,如通过多模态视觉语言预训练(BLIP)模型进行描述提取。然后,本发明实施例将风格描述信息与风格词库中的风格词进行匹配,以确定风格描述信息对应的候选文本表征信息,并将候选文本表征信息与各个子图像的图像描述信息相结合,从而确定风格描述信息对应的风格表征信息。进一步地,本发明实施例通过在每个子图像的图像描述信息前增加风格表征信息,从而得到每个子图像对应的图像描述文本。

[0202] T004:将待重绘图像与噪声源图像结合得到第一图像。

[0203] 本发明实施例中第一图像是指对待重绘图像叠加了噪声的图像。本发明实施例通过将待重绘图像与噪声源图像进行结合的方式,构建得到第一图像,本发明实施例通过噪声图的叠加的方式,以增加图像的细节、纹理和真实感。其中,本发明实施例中噪声源图像是一种随机生成的图像,其包含了随机的像素值,能够模拟真实世界中的噪声和细节。示例性地,以一待重绘图像与噪声源图像结合的应用场景为例,本发明实施例中首先通过随机抽取一个种子来生成一张噪声图,即噪声源图像。需要说明的是,本发明实施例中噪声源图像可以是一个随机的噪声图像。接着,本发明实施例将待重绘图像与该噪声源图像相加,通过图像相加的方式,从而得到第一图像。

[0204] T005:根据图像描述文本,生成图像重绘模型的网络控制信息。

[0205] 本发明实施例中网络控制信息是指图像重绘模型的输入控制信息。本发明实施例通过图像描述文本生成图像重绘模型的网络控制信息。示例性的,在一网络控制信息生成场景中,本发明实施例将待重绘图像对应的风格描述信息经过文本编码器(text

encoder),得到相应的文本表征信息,即网络控制信息。相应地,本发明实施例通过将文本表征信息输入图像重绘模型中进行网络的控制。其中,本发明实施例的图像重绘模型中文本表征信息被作为键值(KV),隐空间表征作为查询(Q)。

[0206] T006:根据网络控制信息,由图像重绘模型对第一图像进行图像重绘,得到待重绘图像的重绘结果。

[0207] 本发明实施例中重绘结果是指重绘待重绘图像得到的重绘图像。相应地,本发明实施例中图像重绘模型是指通过上述图像重绘模型训练方法预先训练得到的模型。本发明实施例根据网络控制信息,通过图像重绘模型对第一图像进行图像重绘,从而得到相应的重绘图像,即待重绘图像的重绘结果。示例性地,以一图像重绘模型的应用场景为例,本发明实施例将经过文本编码器(text encoder)得到的文本表征信息输入图像重绘模型。其中,本发明实施例中输入图像重绘模型已经预先训练完成。接着,本发明实施例中图像重绘模型根据输入的文本表征信息(网络控制信息)以及第一图像,进行相应的图像重绘,从而得到重绘后的图像,即得到重绘结果。

[0208] 示例性地,参照图13和图14,在本发明实施例的图像重绘方法应用过程中,以U风格作为风格模板图,在本实施例中,可以采用结构化多深层次生成对抗网络修复(SD-Inpainting)模型,该模型在输入图像的位置上与sd-v1-5模型有所差异,但其余结构相同。其中,SD-Inpainting模型和sd-v1-5-ema模型共享基本相同的模型结构和推理过程(仅左侧噪声图输入的通道数不同,为4和9),其他参数相同。因此,在进行旁路模块微调的过程中,可以将微调好的模型直接迁移应用到具有相同U-net结构的SD-Inpainting重绘模型中。另外,该SD-Inpainting模型可以根据给定的输入重绘目标图像和掩码(mask),对图像中需要修复的部分进行重绘。其中,输入的掩码中需要重绘的部分被标记为1,不需要重绘的部分被标记为0。本发明实施例中该模型通过学习输入图像的上下文信息和给定的掩码,以及可能的文本控制信息,来生成一张与输入相对应的重绘图像,且重绘的图像会填补掩码标记的区域,使其与周围的图像内容保持一致,实现对图像的重绘。需要说明的是,本发明实施例中目标对象可以输入某张图像,并画出图中需要进行重绘的区域。然后,系统模型通过对该图像进行相应的新风格微调后,产生新的微调模块,在上述SD-Inpainting重绘模型下支持目标对象的任意指令重绘。

[0209] 示例性地,参照图14、图15以及图16,本发明实施例的图像重绘方法应用过程中,本发明实施例的图像重绘方法可以应用于基于单模板图的重绘创作。例如,由目标对象提供某张模版图,如图15所示,对模版图中特定物体(如图15中的大飞船)用掩码遮盖,如图16所示。其中,图16中M所示,即为掩码遮盖区域。参照图14,目标对象输入任意想在该物体位置生成的目标(如“U场景中的飞船改成大老鼠”,则输入:U风格的大老鼠),模型通过新训练的文本表征以及旁路模块生成有效结果。此时通过单张模版图(如图15)进行生成模型微调,微调后的模型进行重绘生成。本发明实施例在目标对象提供模版图后,经过一定时间的加载模型(其中包含模型微调时间),即可提供模版下的重绘能力。

[0210] 如图17所示,本发明实施例还公开了一种图像重绘方法,该方法可以包括但不限于步骤T007至T008:

[0211] T007:获取待重绘图像和待重绘图像对应的风格描述信息。

[0212] 本发明实施例中,待重绘图像为目标对象所需要进行重绘的图像。相应地,获取待

重绘图像可以包括照片、视频帧、绘本图像等。其中,本发明实施例中待重绘图像可以通过目标对象的移动终端、PC端上的应用程序的获取。示例性地,在一图片编辑应用场景中,本发明实施例首先通过搭载在目标对象移动终端上的图片编辑应用程序获取相应的待重绘图像,另外,本发明实施例在获取待重绘图像的同时,还获取其对应的风格描述信息,如相关描述词、风格词以及语句等。

[0213] T008:根据图像重绘模型对待重绘图像进行图像重绘。

[0214] 本发明实施例中,图像重绘模型根据前述的图像重绘模型训练方法确定。本发明实施例通过预训练完成构建的图像重绘模型作为目标模型,用于对待重绘图像进行图像重绘。本发明实施例通过图像重绘模型进行图像重绘的方式,能够有效提高图像重绘的准确性。并且,本发明实施例在图像重绘模型进行训练的过程中,通过从风格模板图中分离得到多个子图像,且每个子图像可以用于后续多轮迭代训练的每轮输入图像,从而使得本发明实施例能够对样本量少(甚至只有一张)的风格模板图进行拆解,让最终进入隐空间去噪网络和图像重绘模型的可训练图像样本量增加,提高模型训练的准确性,防止过拟合现象。同时,本发明实施例还通过对样本量少的风格描述信息进行拓展,使得最终进入隐空间去噪网络和图像重绘模型的可训练风格描述样本量增加,进一步提高了模型训练的准确性。另外,本发明实施例对子图像的图像描述文本的构建进行优化,通过将原始的风格描述信息与子图像自动生成的图像描述信息相结合的方式,能够产生更准确风格信息嵌入,使得模型更容易学习风格信息,在提高模型的准确性的同时,能减少因风格描述不准导致的欠拟合。示例性地,以一游戏开发应用场景为例,本发明实施例通过获取游戏开发目标对象的相关待重绘图像和待重绘图像对应的风格描述信息,将待重绘图像和风格描述信息输入图像重绘模型进行图像重绘。容易理解的是,本发明实施例中对待重绘图像进行的重绘可以为对图像中的特定区域进行相应风格描述的填充,如将待重绘图像中的某个区域中以风格描述信息中的相关风格进行重绘,从而得到一幅符合游戏开发目标对象需求的重绘图像。或者,本发明实施例中还可以根据相关的风格描述信息,通过图像重绘模型将待重绘图像重绘成预期风格的图像,从而实现对相关游戏应用场景的构建。

[0215] 示例性地,以教学应用场景为例,对本发明实施例中图像重绘方法的完整过程进行描述如下:

[0216] 在一搭载了待重绘图像和风格描述信息获取接口的服务器上执行本发明提出的图像重绘方法,首先,通过与前述服务器数据交互的终端基于待重绘图像和风格描述信息获取接口获取由教学目标对象或其他教学执行对象输入的待重绘图像以及待重绘图像对应的风格描述信息,进而调用服务器中预存的训练完成构建的图像重绘模型,其中,图像重绘模型根据前述的图像重绘模型训练方法确定。

[0217] 通过图像重绘模型,基于获取的待重绘图像以及待重绘图像对应的风格描述信息进行图像重绘。进而,基于重绘得到的重绘图像进行相关的教学展示。例如,在一些地理教学场景中,需要对相关的地理演变场景进行演变展示,则本发明实施例可以将相关的待演变图像(待重绘图像),以及相应的演变描述信息(风格描述信息)输入到图像重绘模型,从而得到相应的重绘结果,即演变结果,提高了图像重绘的准确性,并提高了教学效果。

[0218] 如图18所示,本发明实施例还提供了一种图像重绘模型训练装置,该装置包括:

[0219] 第一模块1110,用于获取风格模板图和风格模板图对应的风格描述信息。

[0220] 第二模块1120,用于从风格模板图中分离得到多个子图像。

[0221] 第三模块1130,用于将各个子图像的图像描述信息与风格描述信息结合,得到图像描述文本,并根据子图像和图像描述文本构建多个图文对。其中,每个图文对中包含子图像和子图像对应的图像描述文本。

[0222] 第四模块1140,用于在每轮训练过程中,从多个图文对中选取一个图文对的子图像作为待融合图像。

[0223] 第五模块1150,用于将待融合图像与噪声源图像结合得到待训练风格图像,并根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息。

[0224] 第六模块1160,用于根据网络控制信息,由去噪网络对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果。

[0225] 第七模块1170,用于根据预测结果与噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新。

[0226] 可以理解的是,上面所示的图像重绘模型训练方法实施例中的内容均适用于本图像重绘模型训练装置实施例中,本图像重绘模型训练装置实施例所具体实现的功能与上面所示的图像重绘模型训练方法实施例相同,并且达到的有益效果与如上面所示的图像重绘模型训练方法实施例所达到的有益效果也相同。

[0227] 结合附图18,对本发明提供的图像重绘模型训练装置的具体实施过程进行描述:首先,第一模块1110获取风格模板图和风格模板图对应的风格描述信息。然后,第二模块1120从风格模板图中分离得到多个子图像。其中,本发明实施例通过对样本量少(甚至只有一张)的风格模板图进行拆解的方式,以使得最终进入隐空间去噪网络和图像重绘模型的可训练图像样本量增加,提高模型训练的准确性,防止过拟合现象。接着,第三模块1130将各个子图像的图像描述信息与风格描述信息结合,得到图像描述文本,并根据子图像和图像描述文本构建多个图文对。其中,本发明实施例中每个图文对中包含子图像和子图像对应的图像描述文本。本发明实施例通过将子图像描述信息与风格描述信息结合的方式,对样本量少的风格描述信息进行拓展,从而使得进入隐空间去噪网络和图像重绘模型的可训练风格描述样本量增加,进而提高模型训练的准确性,并且通过将原始的风格描述信息与子图像自动生成的图像描述信息相结合的方式,产生更准确的风格信息嵌入,使得模型更容易学习风格信息,从而在提高模型的准确性的同时,能减少因风格描述不准导致的欠拟合。进一步地,第四模块1140在每轮训练过程中,从多个图文对中选取一个图文对的子图像作为待融合图像。接着,第五模块1150将待融合图像与噪声源图像结合得到待训练风格图像,并根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息。最后,第六模块1160根据网络控制信息,由去噪网络对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果。相应地,第七模块1170能够根据预测结果与噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新。同时,本发明实施例可以根据图像重绘模型对待重绘图像进行图像重绘,使得少样本训练也能降低模型训练过程的过拟合效果,提高图像重绘的准确性。

[0228] 如图19所示,本发明实施例还提供了一种图像重绘装置,该装置包括:

[0229] 第八模块1210,用于获取待重绘图像和待重绘图像对应的风格描述信息。

[0230] 第九模块1220,用于根据图像重绘模型对待重绘图像进行图像重绘。

[0231] 其中,图像重绘模型根据上述的图像重绘模型训练方法确定。

[0232] 可以理解的是,上面所示的图像重绘方法实施例中的内容均适用于本图像重绘装置实施例中,本图像重绘装置实施例所具体实现的功能与上面所示的图像重绘方法实施例相同,并且达到的有益效果与如上面所示的图像重绘方法实施例所达到的有益效果也相同。

[0233] 结合附图19,对本发明提供的图像重绘装置的具体实施过程进行描述:首先,第八模块1210获取待重绘图像以及待重绘图像对应的风格描述信息。接着,第九模块1220根据图像重绘模型对待重绘图像进行图像重绘。具体地,图像重绘模型可以根据上述实施例中任意一个图像重绘模型训练装置确定,使得少样本训练也能降低模型训练过程的过拟合效果。本发明实施例基于图像重绘模型来进行图像重绘,提高了图像重绘的准确性。

[0234] 如图20所示,本发明实施例还提供了一种电子设备,该电子设备包括处理器1310以及存储器1320;存储器1320存储有程序;处理器1310执行程序以执行前述的图像重绘模型训练方法或图像重绘方法;该电子设备具有搭载并运行本发明实施例提供的业务数据处理的软件系统的功能,例如,个人计算机(Personal Computer,PC)、手机、智能手机、个人数字助手(Personal Digital Assistant,PDA)、可穿戴设备、掌上电脑PPC(Pocket PC)、平板电脑、车载终端等。

[0235] 可以理解的是,上面所示的图像重绘模型训练方法或图像重绘方法实施例中的内容均适用于本电子设备实施例中,本电子设备实施例所具体实现的功能与上面所示的图像重绘模型训练方法或图像重绘方法实施例相同,并且达到的有益效果与如上面所示的图像重绘模型训练方法或图像重绘方法实施例所达到的有益效果也相同。

[0236] 本发明实施例还提供了一种计算机可读存储介质,存储介质存储有程序,所述程序被处理器执行实现前述的图像重绘模型训练方法或图像重绘方法。与此同时,本发明实施例还提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器可以从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行前述的图像重绘模型训练方法或图像重绘方法。

[0237] 可以理解的是,上面所示的图像重绘模型训练方法或图像重绘方法实施例中的内容均适用于本计算机可读存储介质实施例中,本计算机可读存储介质实施例所具体实现的功能与上面所示的图像重绘模型训练方法或图像重绘方法实施例相同,并且达到的有益效果与如上面所示的图像重绘模型训练方法或图像重绘方法实施例所达到的有益效果也相同。

[0238] 本发明实施例还公开了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器可以从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行图3、图12、图17所示的方法。

[0239] 可以理解的是,上面所示的图像重绘模型训练方法或图像重绘方法实施例中的内容均适用于本计算机程序产品或计算机程序实施例中,本计算机程序产品或计算机程序实施例所具体实现的功能与上面所示的图像重绘模型训练方法或图像重绘方法实施例相同,并且达到的有益效果与如上面所示的图像重绘模型训练方法或图像重绘方法实施例所达

到的有益效果也相同。

[0240] 在一些可选的实施例中,在方框图中提到的功能/操作可以不按照操作示图提到的顺序发生。例如,取决于所涉及的功能/操作,连续示出的两个方框实际上可以被大体上同时地执行或所述方框有时能以相反顺序被执行。此外,在本发明的流程图中所呈现和描述的实施例以示例的方式被提供,目的在于提供对技术更全面的理解。所公开的方法不限于本文所呈现的操作和逻辑流程。可选的实施例是可预期的,其中各种操作的顺序被改变以及其中被描述为较大操作的一部分的子操作被独立地执行。

[0241] 从上述说明书中所提供的实施例,可以清楚地得知,本发明技术方案存在着至少以下的有益效果:

[0242] 本发明的技术方案通过获取风格模板图和风格模板图对应的风格描述信息,并从风格模板图中分离得到多个子图像,其中,每个子图像可以用于后续多轮迭代训练的每轮输入图像,从而使得本发明实施例能够对样本量少(甚至只有一张)的风格模板图进行拆解,让最终进入隐空间去噪网络和图像重绘模型的可训练图像样本量增加,提高模型训练的准确性,防止过拟合现象。接着,本发明实施例将各个子图像的图像描述信息与风格描述信息结合,得到图像描述文本,本发明实施例通过对样本量少的风格描述信息进行拓展的方式,使得最终进入隐空间去噪网络和图像重绘模型的可训练风格描述样本量增加,进而提高模型训练的准确性。另外,本发明实施例对子图像的图像描述文本的构建进行优化,以将原始的风格描述信息与子图像自动生成的图像描述信息相结合,从而产生更准确风格信息嵌入,使得模型更容易学习风格信息,在提高模型的准确性的同时,缓解因风格描述不准导致的欠拟合问题。在每轮训练过程中,本发明实施例从多个图文对中选取一个图文对的子图像作为待融合图像。然后,本发明实施例将待融合图像与噪声源图像结合得到待训练风格图像,并根据待融合图像对应的风格描述信息,生成去噪网络的网络控制信息。最后,本发明实施例根据网络控制信息,对待训练风格图像进行噪声预测,得到对噪声源图像的预测结果。至此,本发明实施例可以根据预测结果与所述噪声源图像之间的损失计算结果,对图像重绘模型的模型参数进行更新,并且可以根据图像重绘模型对待重绘图像进行图像重绘,实现对图像重绘准确性的提高。

[0243] 在一些可选的实施例中,在方框图中提到的功能/操作可以不按照操作示图提到的顺序发生。例如,取决于所涉及的功能/操作,连续示出的两个方框实际上可以被大体上同时地执行或所述方框有时能以相反顺序被执行。此外,在本发明的流程图中所呈现和描述的实施例以示例的方式被提供,目的在于提供对技术更全面的理解。所公开的方法不限于本文所呈现的操作和逻辑流程。可选的实施例是可预期的,其中各种操作的顺序被改变以及其中被描述为较大操作的一部分的子操作被独立地执行。

[0244] 此外,虽然在功能性模块的背景下描述了本发明,但可以理解的是,除非另有相反说明,所述的功能和/或特征中的一个或多个可以被集成在单个物理装置和/或软件模块中,或者一个或多个功能和/或特征可以在单独的物理装置或软件模块中被实现。还可以理解的是,有关每个模块的实际实现的详细讨论对于理解本发明是不必要的。更确切地说,考虑到在本文中公开的装置中各种功能模块的属性、功能和内部关系的情况下,在工程师的常规技术内将会了解该模块的实际实现。因此,本领域技术人员运用普通技术就能够在无需过度试验的情况下实现在权利要求书中所阐明的本发明。还可以理解的是,所公开的特

定概念仅仅是说明性的,并不意在限制本发明的范围,本发明的范围由所附权利要求书及其等同方案的全部范围来决定。

[0245] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0246] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤,例如,可以被认为是用于实现逻辑功能的可执行指令的定序列列表,可以具体实现在任何计算机可读介质中,以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用,或结合这些指令执行系统、装置或设备而使用。就本说明书而言,“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。

[0247] 计算机可读介质的更具体的示例(非穷尽性列表)包括以下:具有一个或多个布线的电连接部(电子装置)、便携式计算机盘盒(磁装置)、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或闪速存储器)、光纤装置以及便携式光盘只读存储器(CDROM)。另外,计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质,因为可以例如通过对纸或其他介质进行光学扫描,接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序,然后将其存储在计算机存储器中。

[0248] 可以理解,本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如,如果用硬件来实现,和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或他们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列(PGA),现场可编程门阵列(FPGA)等。

[0249] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何一个或多个实施例或示例中以合适的方式结合。

[0250] 尽管已经示出和描述了本发明的实施例,本领域的普通技术人员可以理解:在不脱离本发明的原理和宗旨的情况下可以对这些实施例进行多种变化、修改、替换和变型,本发明的范围由权利要求及其等同物限定。

[0251] 以上是对本发明的较佳实施进行了具体说明,但本发明并不限于所述实施例,熟悉本领域的技术人员在不违背本发明精神的前提下还可做出种种的等同变形或替换,这些等同的变形或替换均包含在本发明权利要求所限定的范围内。

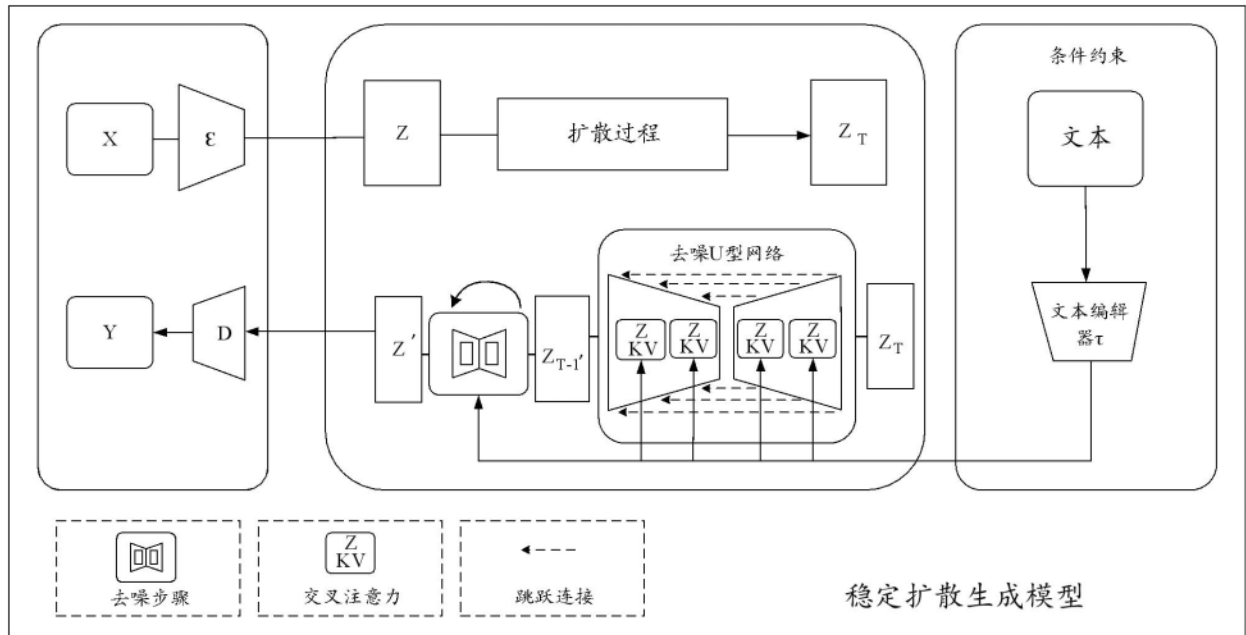


图1

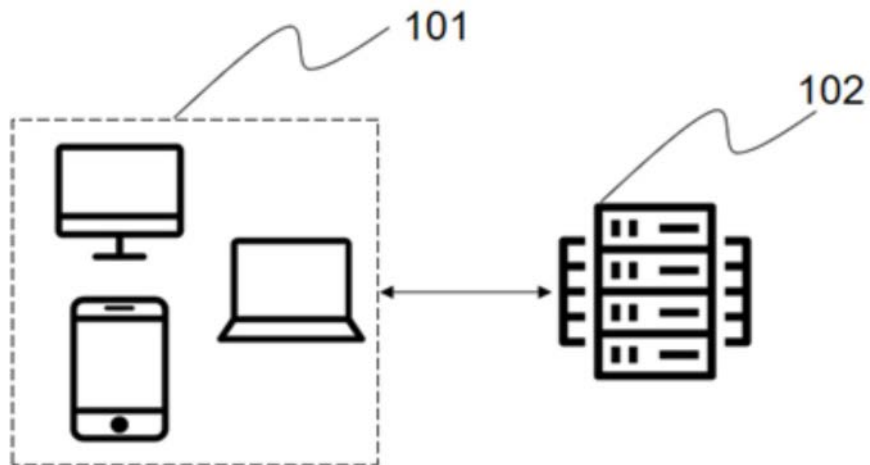


图2

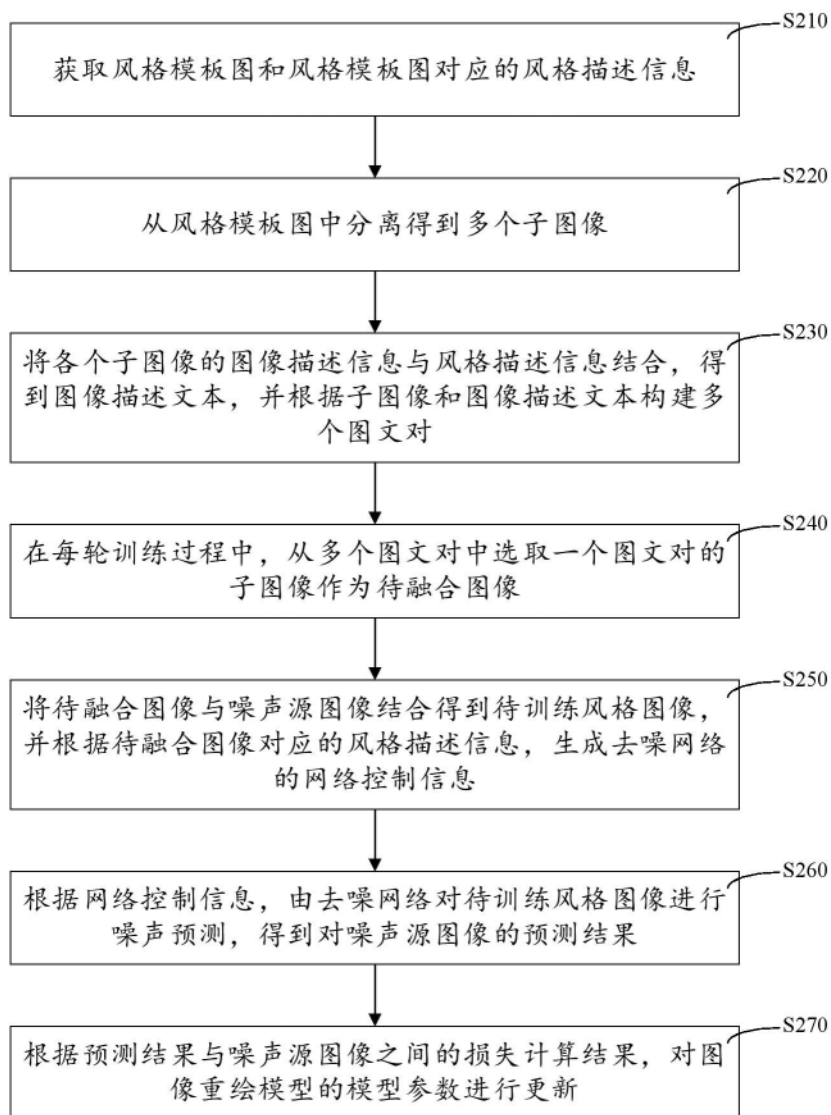


图3

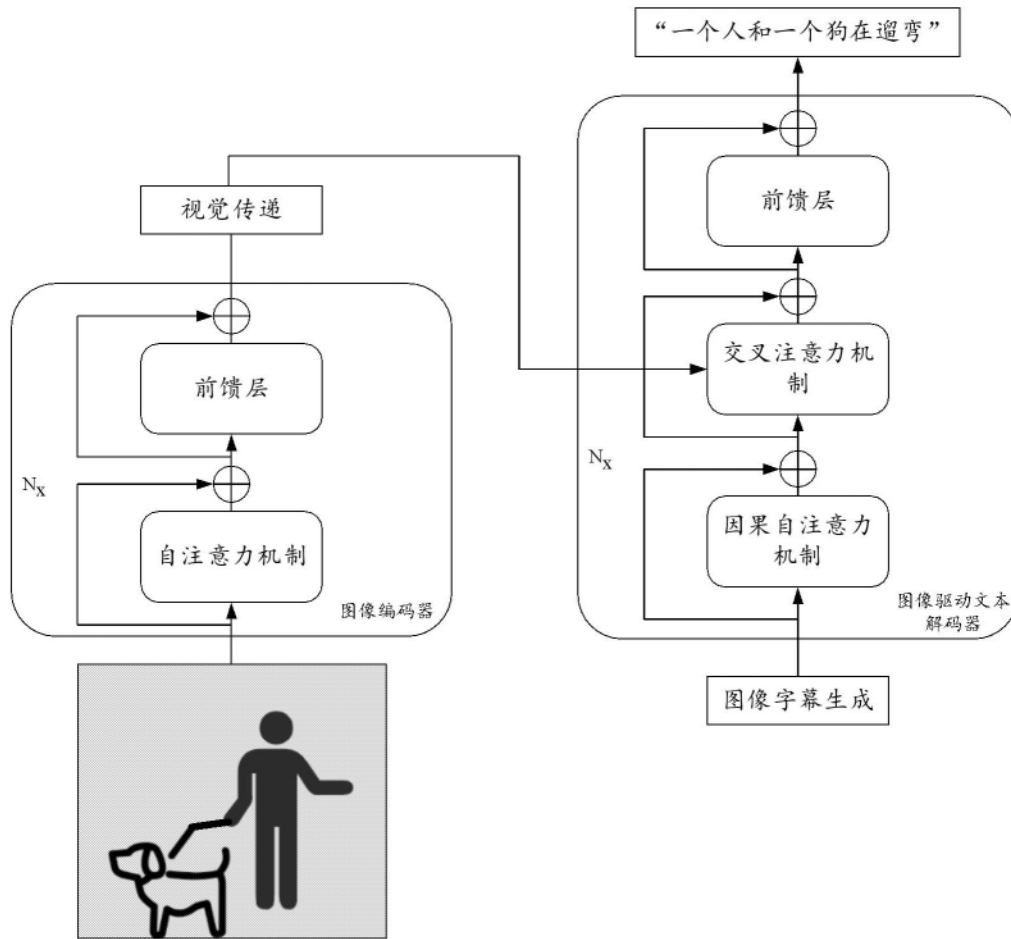


图4

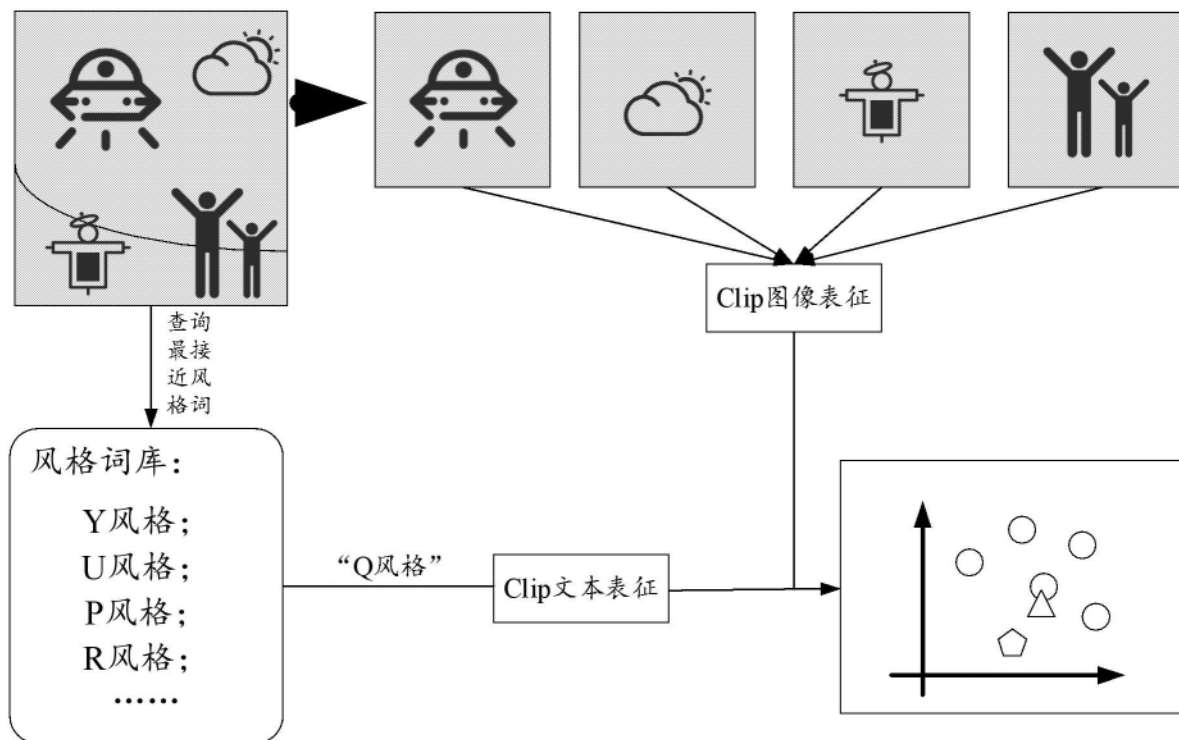


图5

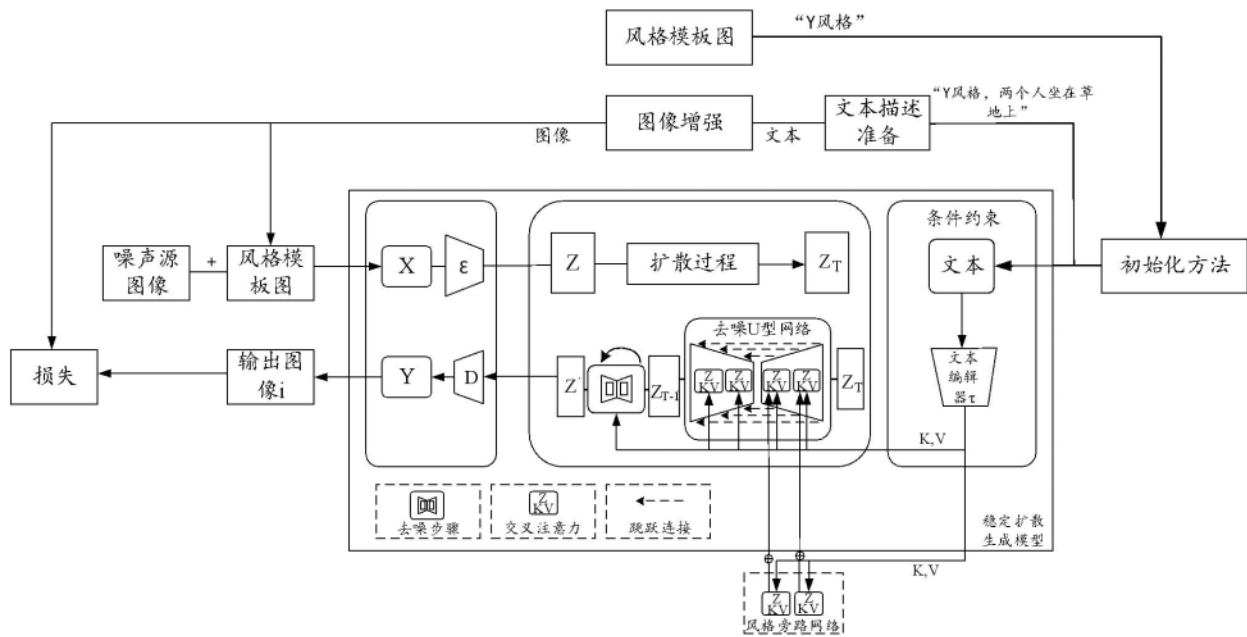


图6

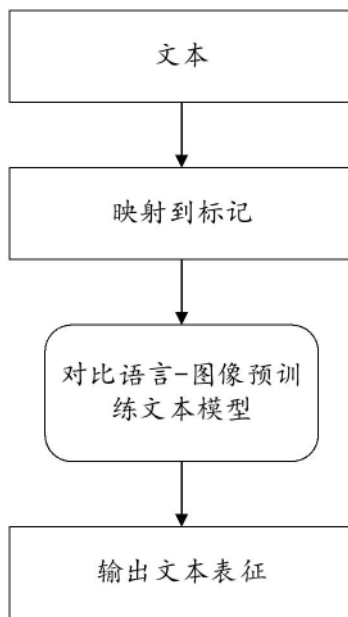


图7

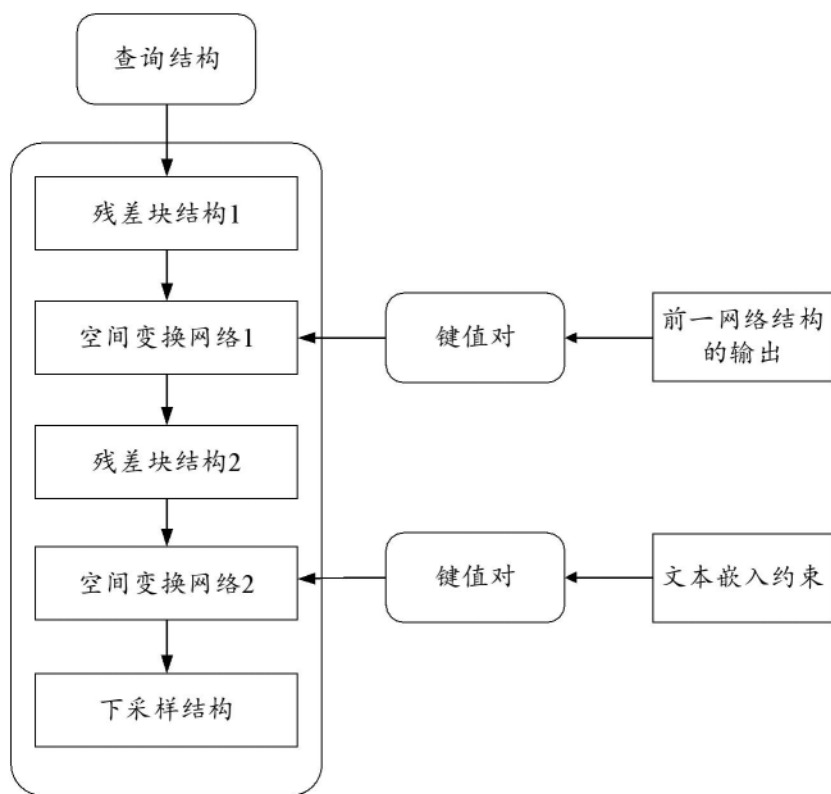


图8

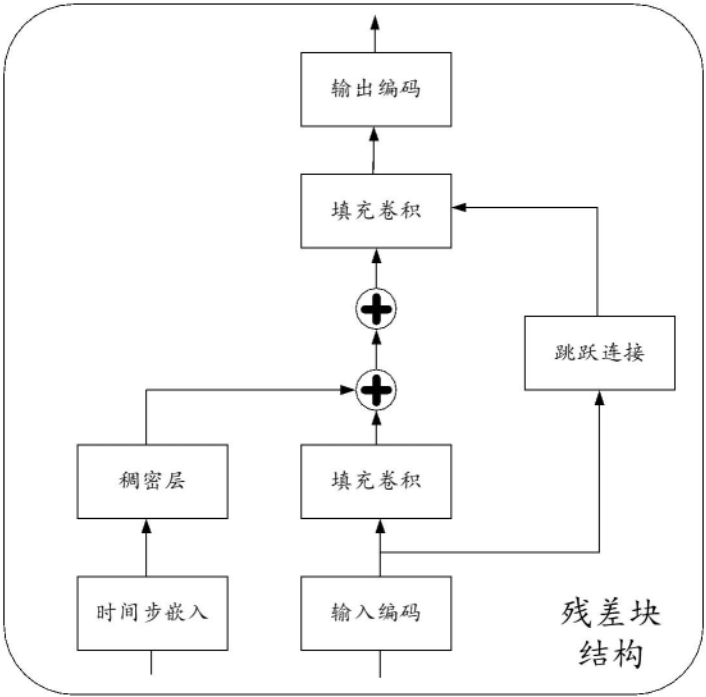


图9

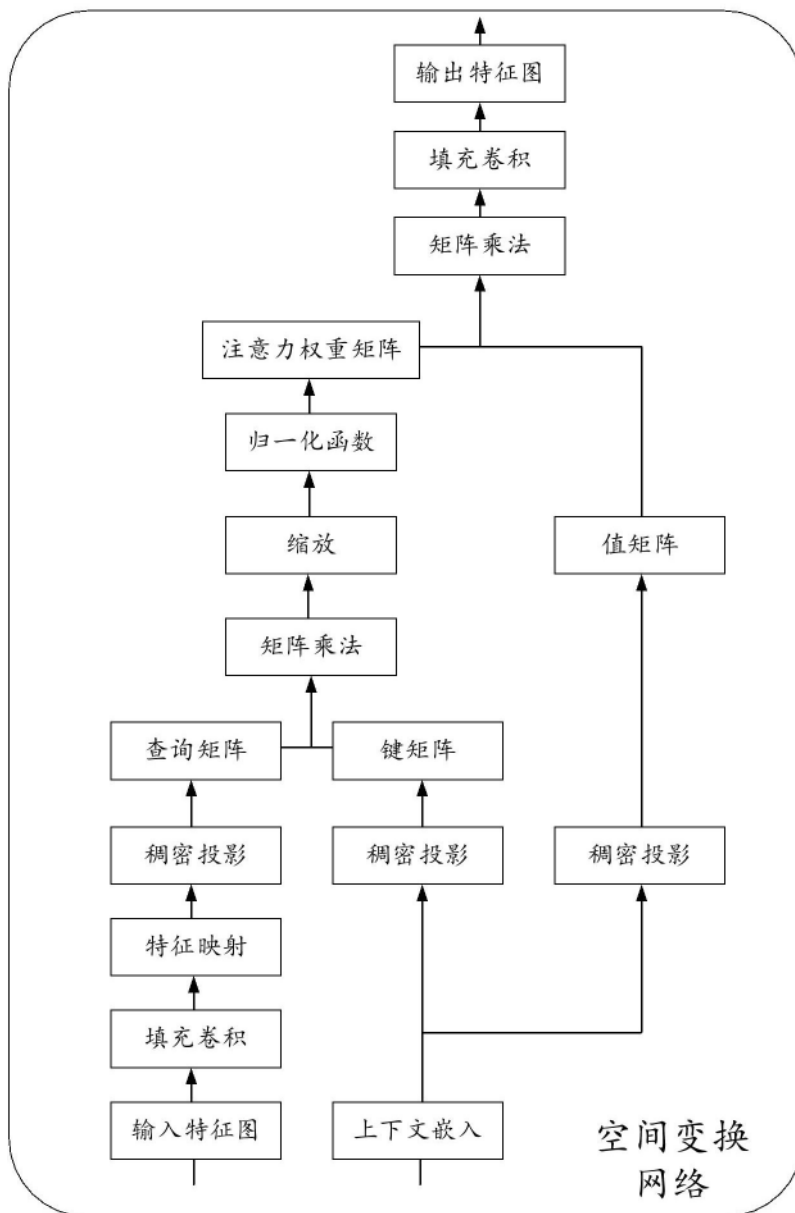


图10

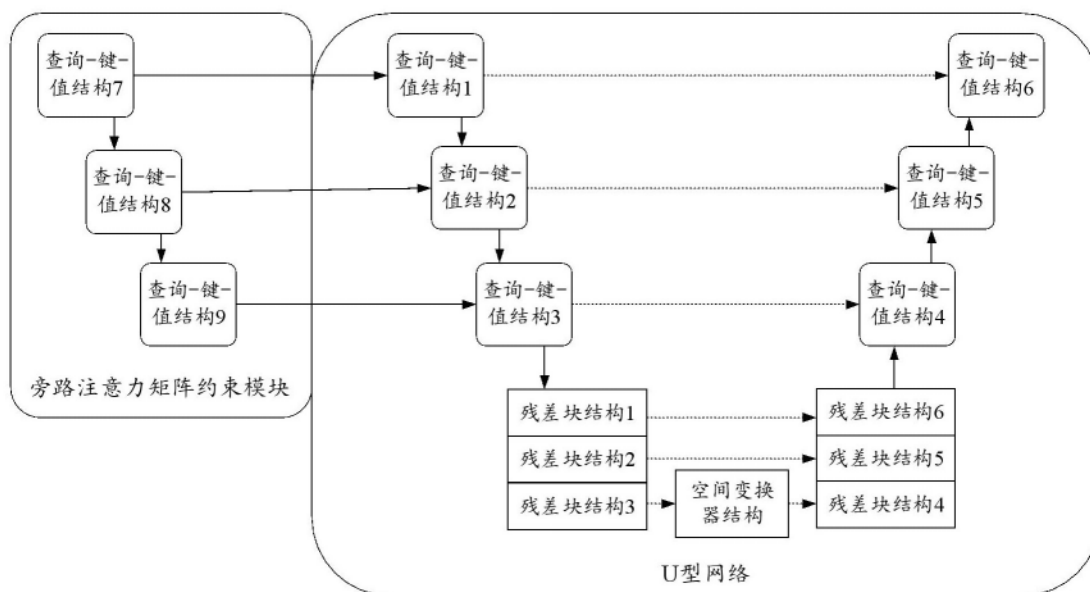


图11

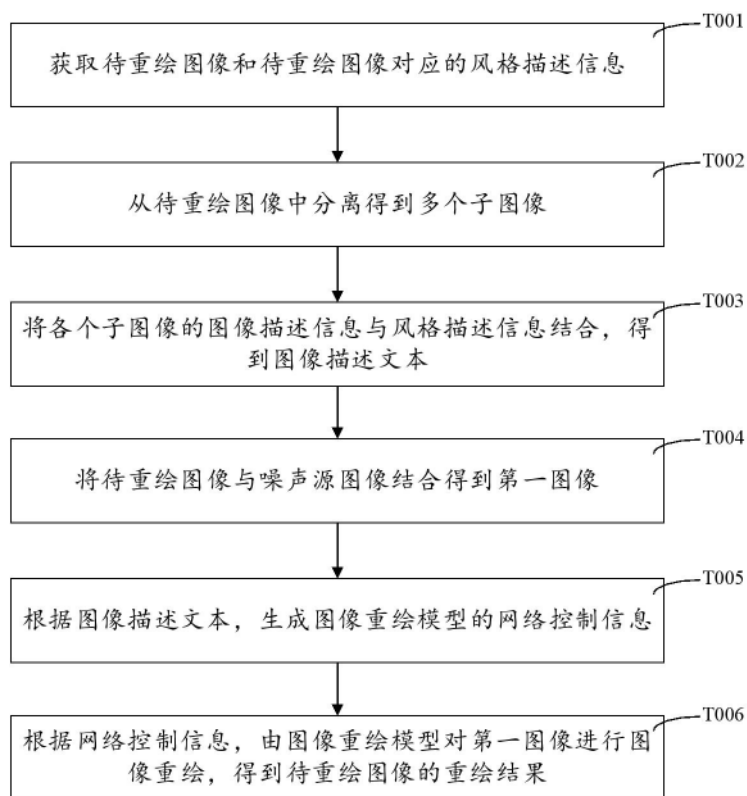


图12

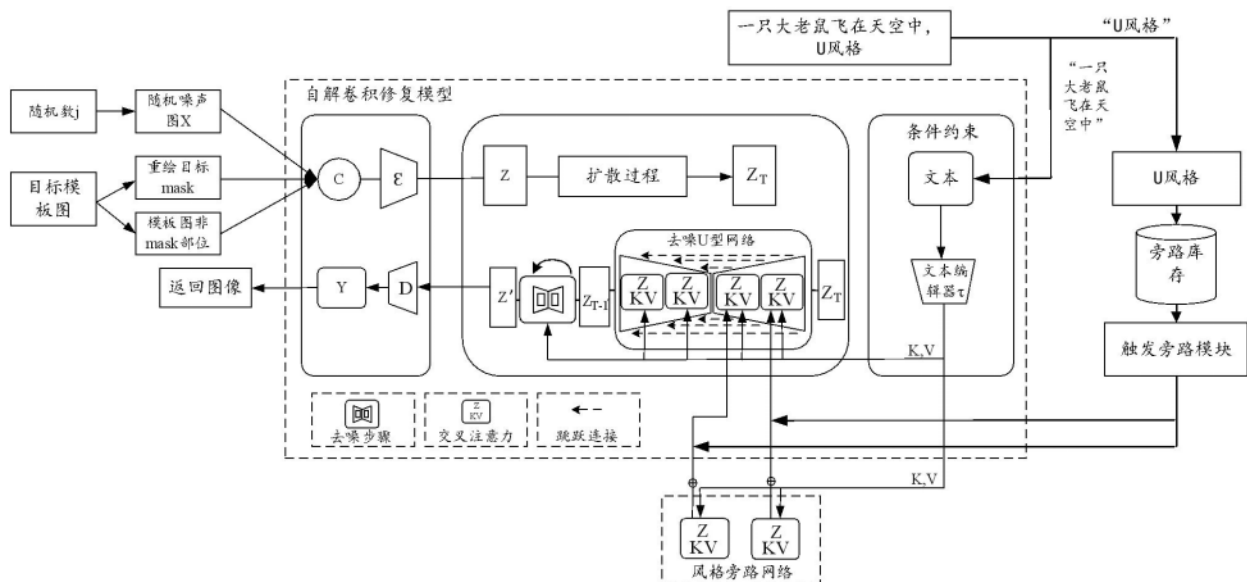


图13

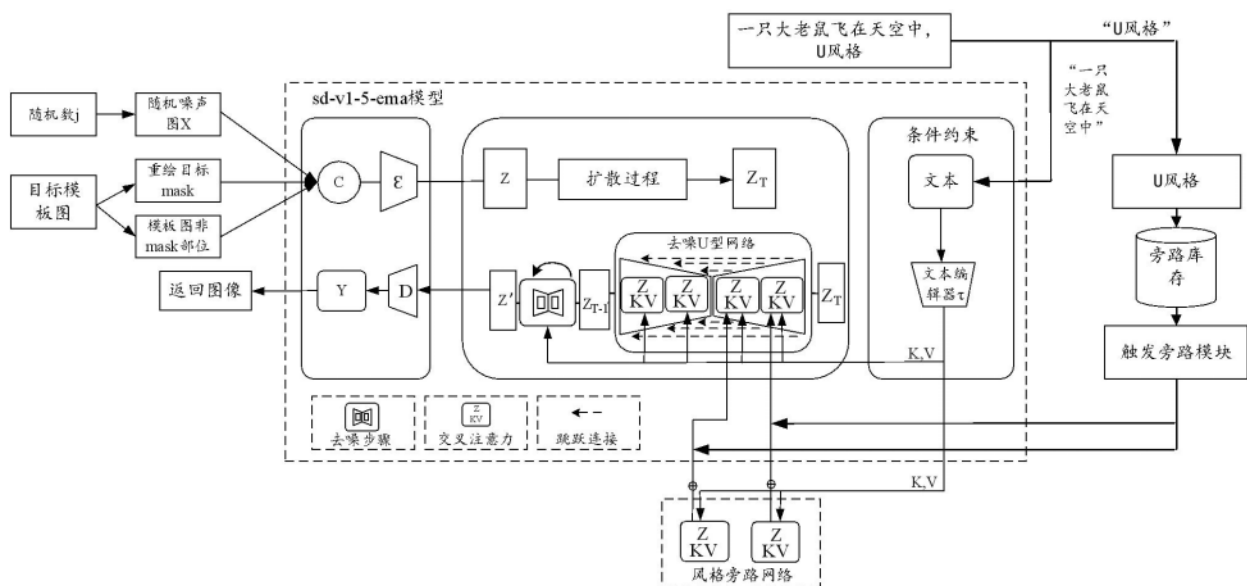


图14

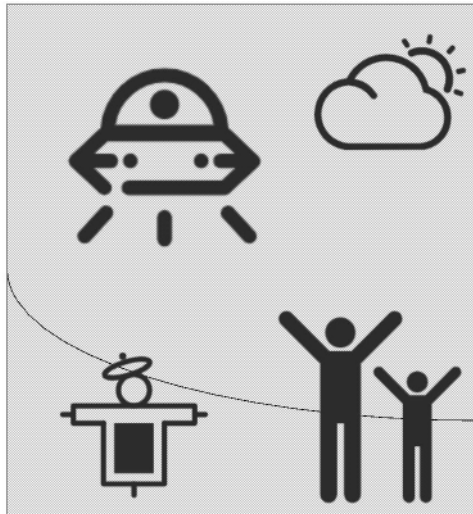


图15

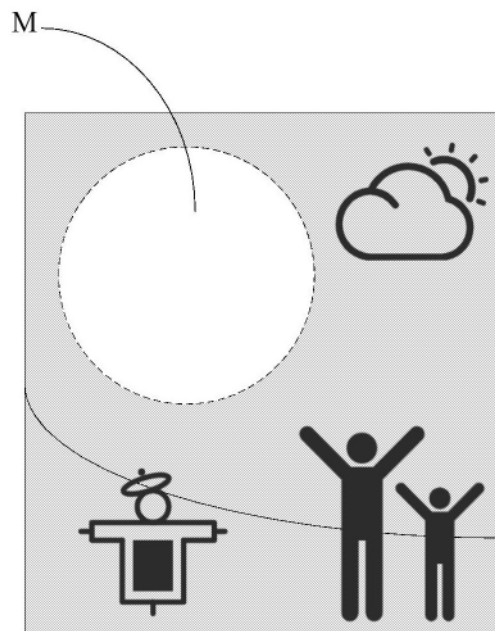


图16

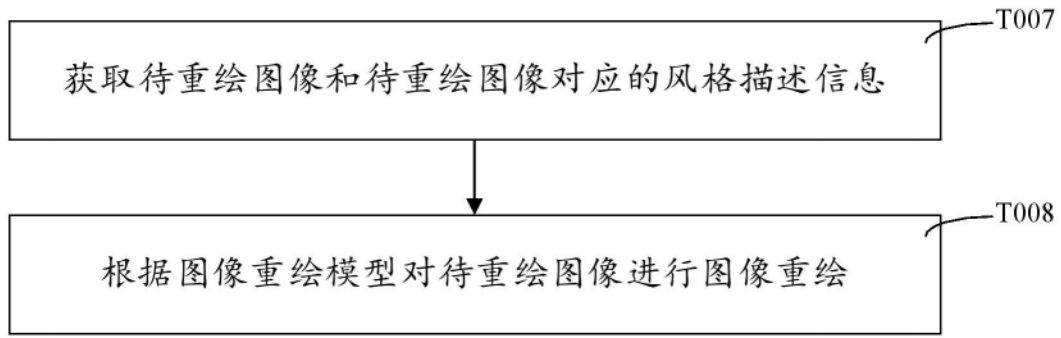


图17

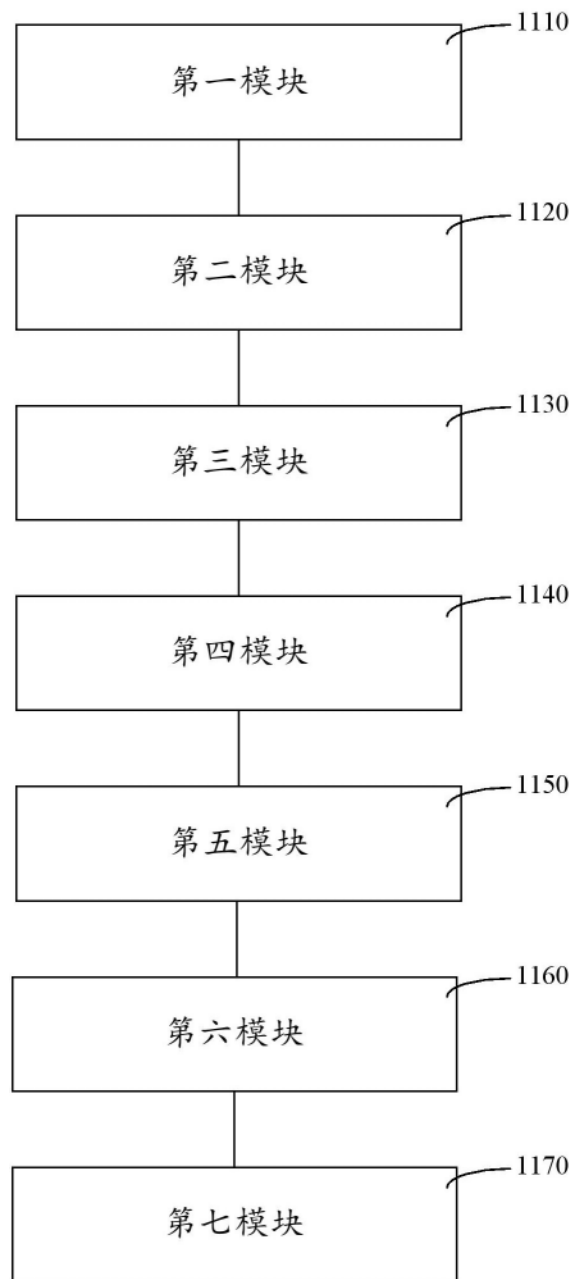


图18



图19

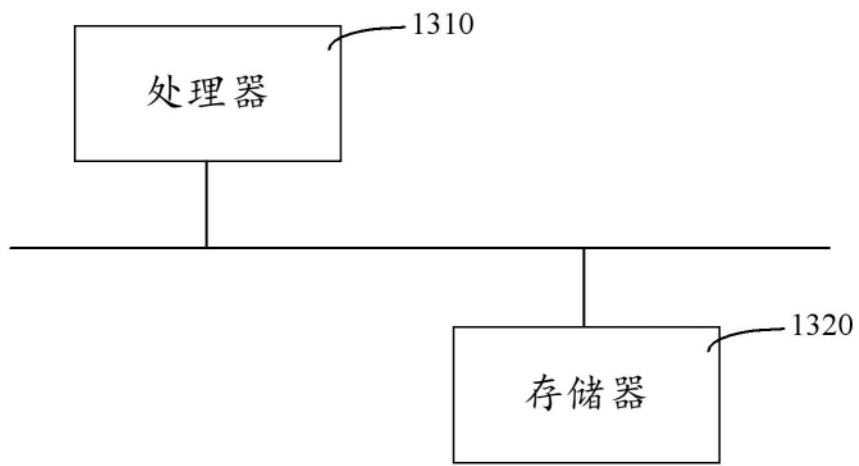


图20