



(12) 发明专利

(10) 授权公告号 CN 114091662 B

(45) 授权公告日 2024.05.14

(21) 申请号 202111425908.3

G06V 10/774 (2022.01)

(22) 申请日 2021.11.26

G06V 10/26 (2022.01)

(65) 同一申请的已公布的文献号

G06N 3/0464 (2023.01)

申请公布号 CN 114091662 A

G06N 3/08 (2023.01)

(43) 申请公布日 2022.02.25

(56) 对比文件

(73) 专利权人 广东伊莱特生活电器有限公司

WO 2021088935 A1, 2021.05.14

地址 528425 广东省中山市东凤镇民乐社

CN 112669215 A, 2021.04.16

区民安三街61号

CN 111275784 A, 2020.06.12

专利权人 广东伊莱特电器有限公司

CN 110706302 A, 2020.01.17

(72) 发明人 杨广良 李成武 曹刚 万景峰

CN 111275780 A, 2020.06.12

(74) 专利代理机构 广州瑞之凡知识产权代理事

CN 111598712 A, 2020.08.28

务所(普通合伙) 44514

CN 113140020 A, 2021.07.20

专利代理师 廖夏林

CN 113298816 A, 2021.08.24

莫建文等. 结合皮尔逊重构的文本到图像生成模型. 桂林电子科技大学学报. 2020, (第01期), 全文.

(51) Int. Cl.

G06V 10/82 (2022.01)

G06F 40/30 (2020.01)

审查员 赵亚丽

权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种文本图像生成方法、装置及电子设备

(57) 摘要

本申请实施例提供一种文本图像生成方法、装置及电子设备,其中,该方法包括:获取文本和文本对应的原始图像;根据文本获得文本特征向量;将文本特征向量输入到多级生成对抗模型中,获得初始文本图像;将初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图;根据初始文本图像空间信息二值图及原始图像空间信息二值图获得文本图像。实施本申请实施例,可以快速、准确获得文字对应的图片,不需要浪费大量人力资源。



1. 一种文本图像生成方法,其特征在于,所述方法包括:

获取文本和所述文本对应的原始图像;

根据所述文本获得文本特征向量;所述根据所述文本获得文本特征向量的步骤,包括:根据所述原始图像和所述文本获得所述文本与所述原始图像的关联度;根据所述关联度获得所述文本特征向量;所述根据所述原始图像和所述文本获得所述文本与所述原始图像的关联度的步骤,包括:获取所述原始图像的目标区域;获取所述文本中的词汇数据;根据所述目标区域和所述词汇数据获得所述文本与所述原始图像的关联度;所述根据所述关联度获得文本特征向量的步骤,包括:获得所述词汇数据的嵌入向量;将所述嵌入向量和所述关联度输入多级生成对抗模型,获得所述文本特征向量;

将所述文本特征向量输入到多级生成对抗模型中,获得初始文本图像;

将所述初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图;

根据所述初始文本图像空间信息二值图及所述原始图像空间信息二值图获得文本图像;所述根据所述初始文本图像空间信息二值图及原始图像空间信息二值图获得文本图像的步骤,包括:将所述初始文本图像空间信息二值图及所述原始图像空间信息二值图输入到空间约束判别模型中,分别获得初始文本图像二值图向量和原始生成图像二值图向量;根据所述初始文本图像二值图向量和所述原始生成图像二值图向量获得文本图像;

所述空间约束判别模型对应的空间约束公式为:

$$\sum_i ||\phi(f_{t,i}) - \phi(f_{g,i})||^2;$$

其中, $f_{t,i}$ 为初始文本图像二值图向量, $f_{g,i}$ 为原始生成图像二值图向量;

所述多级生成对抗模型的鉴别器损失函数为:

$$L_D = E_{(I,t) \sim P_{data}} [\log D(I, \phi_t)] + E_{s \sim PG, t \sim P_{data}} [\log (I - D(G(s, c), \phi_t))] + \sum_i ||\phi(I) - \phi(G(s, c))||^2;$$

其中, $E_{(I,t) \sim P_{data}}$ 为原始图像的均方误差, $\phi_t$ 为文本的特征向量, $E_{s \sim PG, t \sim P_{data}}$ 为初始文本图像的均方误差, $G()$ 表示生成器的输出结果, $D()$ 表示鉴别器的输出结果, $I$ 为原始图像, $s$ 为文本, $c$ 为噪声, $L_D$ 为鉴别器损失函数;

所述多级生成对抗模型的生成器损失函数为:

$$L_G = E_{s \sim PG, t \sim P_{data}} [\log (I - D(G(s, c), \phi_t))];$$

其中, $E_{s \sim PG, t \sim P_{data}}$ 为初始文本图像的均方误差, $G()$ 表示生成器的输出结果, $D()$ 表示鉴别器的输出结果, $I$ 为原始图像, $s$ 为文本, $c$ 为噪声, $\phi_t$ 为文本的特征向量, $L_G$ 为生成器损失函数。

2. 一种文本图像生成装置,其特征在于,应用如权利要求1所述的文本图像生成方法,所述装置包括:

获取模块,用于获取文本和所述文本对应的原始图像;

转换模块,用于根据所述文本获得文本特征向量;

生成模块,用于将所述文本特征向量输入到多级生成对抗模型中,获得初始文本图像;

语义分割模块,用于将所述初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图;

图像生成模块,用于根据所述初始文本图像空间信息二值图及所述原始图像空间信息

二值图获得文本图像。

3.一种电子设备,其特征在于,包括存储器及处理器,所述存储器用于存储计算机程序,所述处理器运行所述计算机程序以使所述电子设备执行根据权利要求1所述的文本图像生成方法。

## 一种文本图像生成方法、装置及电子设备

### 技术领域

[0001] 本申请涉及图像生成技术领域,具体而言,涉及一种文本图像生成方法、装置及电子设备。

### 背景技术

[0002] 近年来,随着移动互联网的飞速发展,海量数据不断产生,文本、语音、图像、视频等数据形式日益多样化。随着数据的爆炸式增长,计算机硬件水平不断提高,越来越多的应用场景需要完成语义图像分割、图像分类、目标获取等任务。

[0003] 然而,需要将文字转换成图片的应用场景也越来越多,例如,设计师设计出的图片常常无法满足客户的描述,即使反复修改,依然无法达到客户的要求;在犯罪现场,目击案件发生及犯罪嫌疑人的目击者往往只能通过口头表达的形式将犯罪嫌疑人的外貌特征描述出来,这些需要将描述转化成图片的应用场景,往往需要专业人士、社会人士的参与才可以实现,费时费力,并且不一定能获得较好的效果。

### 发明内容

[0004] 本申请实施例的目的在于提供一种文本图像生成方法、装置及电子设备,解决了一些应用场景下无法快速、准确得到文字对应的图片的问题,可以轻松、有效的生成图片,不需要浪费大量人力资源。

[0005] 第一方面,本申请实施例提供了一种文本图像生成方法,所述方法包括:

[0006] 获取文本和所述文本对应的原始图像;

[0007] 根据所述文本获得文本特征向量;

[0008] 将所述文本特征向量输入到多级生成对抗模型中,获得初始文本图像;

[0009] 将所述初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图;

[0010] 根据所述初始文本图像空间信息二值图及所述原始图像空间信息二值图获得文本图像。

[0011] 在上述实现过程中,通过语义分割模型、多级生成对抗模型等方式,将获取到的文本和文本对应的原始图像进行训练,输出文本图像,通过语义分割模型得出初始文本图像的空间信息,通过与原始图像的真实的空间信息比较,从而提高生成的文本图像的空间正确性,将文本描述的内容以文本图像的形式呈现,无需消耗大量人力制作图片。

[0012] 进一步地,所述根据所述文本获得文本特征向量的步骤,包括:

[0013] 根据所述原始图像和所述文本获得所述文本与所述原始图像的关联度;

[0014] 根据所述关联度获得文本特征向量。

[0015] 在上述实现过程中,文本特征向量作为输入可以增强多级生成对抗模型的有效性 & 多样性,使得模型可以更好地学习到文本的特征,得到的文本图像也更加贴合文本。

[0016] 进一步地,所述根据所述原始图像和所述文本获得所述文本与所述原始图像的关

联度的步骤,包括:

[0017] 获取所述原始图像的目标区域;

[0018] 获取所述文本中的词汇数据;

[0019] 根据所述目标区域和所述词汇数据获得所述文本与所述原始图像的关联度。

[0020] 在上述实现过程中,获取原始图像的目标区域后,计算其与输入的文本之间的关联度,可以使得文本与原始图像之间的联系更加紧密,从而提升输入的文本特征向量的空间代表性。

[0021] 进一步地,所述根据所述关联度获得文本特征向量的步骤,包括:

[0022] 获得所述词汇数据的嵌入向量;

[0023] 将所述嵌入向量和所述关联度输入多级生成对抗模型,获得所述文本特征向量。

[0024] 在上述实现过程中,词汇数据的嵌入向量用于在生成文本图像过程中提取出文本中的重要词汇,并且找出与之对应的原始图像的目标区域,提高该区域的注意力,以此提高生成的文本图像与文本语义的一致性。

[0025] 进一步地,所述根据所述初始文本图像空间信息二值图及所述原始图像空间信息二值图获得文本图像的步骤,包括:

[0026] 将所述初始文本图像空间信息二值图及所述原始图像空间信息二值图输入到空间约束判别模型中,分别获得初始文本图像二值图向量和原始生成图像二值图向量;

[0027] 根据所述初始文本图像二值图向量和所述原始生成图像二值图向量获得文本图像。

[0028] 在上述实现过程中,通过初始文本图像二值图向量和原始生成图像二值图向量来得到文本图像的空间正确性,使得模型学习到的文本图像更加符合真实图像。

[0029] 进一步地,所述空间约束判别模型对应的空间约束公式为:

$$[0030] \quad \sum_i || \varphi (f_{t,i}) - \varphi (f_{g,i}) ||^2;$$

[0031] 其中, $f_{t,i}$ 为初始文本图像二值图向量, $f_{g,i}$ 为原始生成图像二值图向量。

[0032] 在上述实现过程中,初始文本图像二值图向量和原始生成图像二值图向量可以帮助模型获得更多原始图像的真实图像信息,使得模型更新的参数更加符合空间正确性。

[0033] 进一步地,所述将所述文本特征向量输入到多级生成对抗模型中,获得初始文本图像的步骤,包括:

[0034] 所述多级生成对抗模型的鉴别器损失函数为:

$$[0035] \quad L_D = E_{(I,t) \sim P_{data}} [\log D(I, \varphi_t)] + E_{s \sim P_G, t \sim P_{data}} [\log (1 - D(G(s, c), \varphi_t))] + \sum_i || \phi(I) - \phi(G(s, c)) ||^2 ;$$

[0036] 其中, $E_{(I,t) \sim P_{data}}$ 为原始图像的均方误差, $\varphi_t$ 为文本的特征向量, $E_{s \sim P_G, t \sim P_{data}}$ 为初始文本图像的均方误差, $G()$ 表示生成器的输出结果, $D()$ 表示鉴别器的输出结果, $I$ 为原始图像, $s$ 为文本, $c$ 为噪声, $L_D$ 为鉴别器损失函数。

[0037] 在上述实现过程中,鉴别器损失函数可以表现原始图像与文本图像的差距程度。

[0038] 进一步地,所述将所述文本特征向量输入到多级生成对抗模型中,获得初始文本图像的步骤,包括:

[0039] 所述多级生成对抗模型的生成器损失函数为:

$$[0040] \quad L_G = E_{s \sim P_G, t \sim P_{data}} [\log (1 - D(G(s, c), \varphi_t))] ;$$

[0041] 其中,  $E_{s \sim p_G, I \sim p_{data}}$  为初始文本图像的均方误差,  $G()$  表示生成器的输出结果,  $D()$  表示鉴别器的输出结果,  $I$  为原始图像,  $s$  为文本,  $c$  为噪声,  $\phi_t$  为文本的特征向量,  $L_G$  为生成器损失函数。

[0042] 在上述实现过程中,生成器损失函数可以提高模型的鲁棒性。

[0043] 第二方面,本申请实施例还提供了一种文本图像生成装置,所述装置包括:

[0044] 获取模块,用于获取文本和所述文本对应的原始图像;

[0045] 转换模块,用于根据所述文本获得文本特征向量;

[0046] 生成模块,用于将所述文本特征向量输入到多级生成对抗模型中,获得初始文本图像;

[0047] 语义分割模块,用于将所述初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图;

[0048] 图像生成模块,用于根据所述初始文本图像空间信息二值图及所述原始图像空间信息二值图获得文本图像。

[0049] 在上述实施过程中,通过语义分割模型、多级生成对抗模型等方式,将获取到的文本和文本对应的原始图像进行训练,输出文本图像,通过语义分割模型得出初始文本图像的空间信息,通过与原始图像的真实的空间信息比较,更新模型的参数,从而提高生成的文本图像的空间正确性,将文本描述的内容以文本图像的形式呈现,无需消耗大量人力制作图片。

[0050] 第三方面,本申请实施例提供的一种电子设备,包括:存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如第一方面任一项所述的方法的步骤。

[0051] 第四方面,本申请实施例提供的一种计算机程序产品,所述计算机程序产品在计算机上运行时,使得计算机执行如第一方面任一项所述的方法。

[0052] 本公开的其他特征和优点将在随后的说明书中阐述,或者,部分特征和优点可以从说明书推知或毫无疑义地确定,或者通过实施本公开的上述技术即可得知。

[0053] 并可依照说明书的内容予以实施,以下以本申请的较佳实施例并配合附图详细说明如后。

## 附图说明

[0054] 为了更清楚地说明本申请实施例的技术方案,下面将对本申请实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本申请的某些实施例,因此不应被看作是对范围的限定,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相关的附图。

[0055] 图1为本申请实施例提供的文本图像生成方法的流程示意图;

[0056] 图2为本申请实施例提供的文本图像生成装置的结构组成示意图;

[0057] 图3为本申请实施例提供的电子设备的结构组成示意图。

## 具体实施方式

[0058] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行描述。

[0059] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步定义和解释。同时,在本申请的描述中,术语“第一”、“第二”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0060] 下面结合附图和实施例,对本申请的具体实施方式作进一步详细描述。以下实施例用于说明本申请,但不用来限制本申请的范围。

[0061] 实施例一

[0062] 图1是本申请实施例提供的文本图像生成方法的流程示意图,如图1所示,该方法包括:

[0063] S1,获取文本和文本对应的原始图像;

[0064] S2,根据文本获得文本特征向量;

[0065] S3,将文本特征向量输入到多级生成对抗模型中,获得初始文本图像;

[0066] S4,将初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图;

[0067] S5,根据初始文本图像空间信息二值图及原始图像空间信息二值图获得文本图像。

[0068] 在上述实现过程中,通过语义分割模型、多级生成对抗模型等方式,将获取到的文本和文本对应的原始图像进行训练,输出文本图像,通过语义分割模型得出初始文本图像的空间信息,通过与原始图像的真实的空间信息比较,更新模型的参数,从而提高生成的文本图像的空间正确性,将文本描述的内容以文本图像的形式呈现,无需消耗大量人力制作图片。

[0069] 在S1中,获取文本和文本对应的原始图像。

[0070] 以本实施例为例,文本和原始图像为相互对应的关系,文本为对原始图像的描述,包括原始图像的场景、风格、细节等,文本和文本对应的原始图像可以从数据库网站中获得。

[0071] 在S2中,根据文本获得文本特征向量,包括:

[0072] 根据原始图像和文本获得文本与原始图像的关联度;

[0073] 根据关联度获得文本特征向量。

[0074] 以本实施例为例,文本特征向量作为输入可以增强生成的图片的有效性及多样性,使得模型可以更好地学习到文本的特征,获得的文本图像也更加贴合文本。

[0075] 进一步地,根据原始图像和文本获得文本与原始图像的关联度,包括:

[0076] 获取原始图像的目标区域;

[0077] 获取文本中的词汇数据;

[0078] 根据目标区域和词汇数据获得文本与原始图像的关联度。

[0079] 示例性地,原始图像的目标区域利用yolo-v3进行提取,可以获取到图原始像中与文本对应的目标区域,同时选择自底向上和自顶向下的注意力模型,用于提取每个词汇相对于原始图像的视觉特征。获取到yolo-v3预训练模型后,再训练多级生成对抗模型的过程中,获取模型输入的方法为:将文本及与文本对应的原始图像输入到记忆力机制文本编码器中,记忆力机制文本编码器中包含预训练好的yolo-v3,通过yolo-v3获取到输入原始图像的目标区域,可选低,目标区域可以是10个子区域,然后利用自底向上和自顶向下的注意

力模型计算出文本中每个单词与这10个区域间的关联度,计算方法如下:

$$[0080] \quad q_n = \arg \max_{i=1}^{10} a_{i,n}, m_r = \frac{\sum_{n=1}^N a_{q_n,n} v_{q_n,n}}{\sum_{n=1}^N a_{q_n,n}} ;$$

[0081] 其中,N为训练数据中文本中到第r个词汇的相关原始图像的数量, $a_{i,n}$ 指相关的第n个原始图像上第i个区域的权重, $q_n$ 是权重最大的区域的索引, $v_{q_n,n}$ 是第n个原始图像的区域视觉特征向量。

[0082] 以本实施例为例,获取图像目标区域后,计算其与输入的文本之间的关联度,可以使得文本与原始图像之间的联系更加紧密,从而提升输入的文本特征向量的空间代表性。

[0083] 进一步地,根据关联度获得文本特征向量,包括:

[0084] 获得词汇数据的嵌入向量;

[0085] 将嵌入向量和关联度输入多级生成对抗模型,获得文本特征向量。

[0086] 示例性地,多级生成对抗模型主要由3个生成器主要。第一阶段生成器由4个反卷积块组成,每个反卷积块由上采样层以及谱归一化层组成。上采样层每次将三维特征向量的通道数减少为原来一半,同时将特征宽高扩大为原来2倍。第一阶段生成特征向量的维度为 $3 \times 64 \times 64$ 。谱归一化层用于提高多级生成对抗模型在训练过程中的稳定性,避免模式崩溃等问题。第二阶段及第三阶段生成器主要由卷积层、残差层以及上采样层组成。卷积层及残差层处理上一阶段生成的特征图像,上采样层提升图像像素。

[0087] 可选地,文本特征向量输入到下一阶段生成器前,通过一个文本-图像注意力机制,该注意力机制有两部分输入,分别为文本的词嵌入特征矩阵以及生成器的输出特征。通过该文本-图像注意力机制,可以计算出子图中与关键词最相关的部分,提升子图质量。

[0088] 进一步的,每个阶段的生成器都有与之对应的鉴别器,鉴别器有3部分输入,分别为文本特征向量、初始文本图像、原始图像。初始文本图像与原始图像作为一个样本对,构成非条件约束,文本特征向量与初始文本图像以及原始图像构成条件约束。非条件约束用于提升文本图像的真实性,条件约束用于提升文本图像与文本的语义的一致性。

[0089] 以本实施例为例,词汇数据的嵌入向量用于在生成文本图像过程中提取出文本中的重要词汇,并且找出与之对应的原始图像的目标区域,提高该区域的注意力,以此提高生成的文本图像与文本语义的一致性。

[0090] 进一步地,将文本特征向量输入到多级生成对抗模型中,获得初始文本图像的步骤,包括:

[0091] 多级生成对抗模型的鉴别器损失函数为:

$$[0092] \quad L_D = E_{(I,t) \sim p_{data}} [\log D(I, \varphi_t)] + E_{s \sim p_G, t \sim p_{data}} [\log(1 - D(G(s, c), \varphi_t))] + \sum_i \|\phi(I) - \phi(G(s, c))\|^2 ;$$

[0093] 其中, $E_{(I,t) \sim p_{data}}$ 为原始图像的均方误差, $\varphi_t$ 为文本的特征向量, $E_{s \sim p_G, t \sim p_{data}}$ 为初始文本图像的均方误差,G()表示生成器的输出结果,D()表示鉴别器的输出结果,I为原始图像,s为文本,c为噪声, $L_D$ 为鉴别器损失函数。

[0094] 以本实施例为例,鉴别器损失函数可以表现原始图像与文本图像的差距程度, $D(I, \varphi_t)$ 表示将原始图像和文本的特征向量输入鉴别器中得到的结果,G(s, c)表示将文本和噪声输入生成器中得到的结果, $D(G(s, c), \varphi_t)$ 表示将文本的特征向量和上述生成器中得

到的结果,输入到鉴别器中,得到的结果。

[0095] 进一步地,将文本特征向量输入到多级生成对抗模型中,获得初始文本图像的步骤,包括:

[0096] 多级生成对抗模型的生成器损失函数为:

$$[0097] \quad L_G = E_{s \sim p_G, t \sim p_{data}} [\log(1 - D(G(s, c), \varphi_t))];$$

[0098] 其中,  $E_{s \sim p_G, t \sim p_{data}}$  为初始文本图像的均方误差,  $G()$  表示生成器的输出结果,  $D()$  表示鉴别器的输出结果,  $I$  为原始图像,  $s$  为文本,  $c$  为噪声,  $\varphi_t$  为文本的特征向量,  $L_G$  为生成器损失函数。

[0099] 以本实施例为例,生成器损失函数可以提高模型的鲁棒性,  $G(s, c)$  表示将文本和噪声输入生成器中得到的结果,  $D(G(s, c), \varphi_t)$  表示将文本的特征向量和上述生成器中得到的结果,输入到鉴别器中,得到的结果。

[0100] 在S3中,将文本特征向量输入到多级生成对抗模型中,获得初始文本图像。

[0101] 在S4中,将初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图。

[0102] 语义分割模型主要包含两部分:编码器以及解码器。编码器使用卷积层及池化层组成,卷积层用于计算相邻像素之间的关系及降低特征图大小,池化层用于减小参数数量。解码器用于将编码器中池化层产生的高维特征向量进行反卷积,提升特征图大小,最终产生与原始出入尺寸相等的语义分割图。

[0103] 将初始文本图像及原始图像分别输入到语义分割模型,经过语义分割模型中卷积层和池化层的作用,获得初始文本图像空间信息二值图以及原始图像空间信息二值图。

[0104] 在S5中,根据初始文本图像空间信息二值图及原始图像空间信息二值图获得文本图像,包括:

[0105] 将初始文本图像空间信息二值图及原始图像空间信息二值图输入到空间约束判别模型中,分别获得初始文本图像二值图向量和原始生成图像二值图向量;

[0106] 根据初始文本图像二值图向量和原始生成图像二值图向量获得文本图像。

[0107] 以本实施例为例,通过初始文本图像二值图向量和原始生成图像二值图向量来获得文本图像的空间正确性,并通过模型的迭代,使得模型学习到的文本图像更加符合真实图像。

[0108] 进一步地,空间约束判别模型对应的空间约束公式为:

$$[0109] \quad \sum_i || \varphi(f_{t,i}) - \varphi(f_{g,i}) ||^2;$$

[0110] 其中,  $f_{t,i}$  为初始文本图像二值图向量,  $f_{g,i}$  为原始生成图像二值图向量。

[0111] 可选地,条件约束中包含空间约束。具体而言,将初始文本图像空间信息二值图及原始图像空间信息二值图,作为初始文本图像的空间信息,通过vgg网络获得初始文本图像二值图向量和原始生成图像二值图向量,再计算出空间约束信息,作为条件约束一部分,用于更新鉴别器及生成器。

[0112] 进一步的,该模型训练过程中,一轮训练分为2个过程:首先固定生成器参数,使用鉴别器损失函数更新鉴别器参数,然后固定鉴别器参数,使用生成器损失函数更新生成器参数,以此类推,完成600次的训练,学习率设置为0.0002。

[0113] 通过上述方式可以获得基于记忆机制的文本及空间约束的多级生成对抗模型,后续使用时只需要将文本输入到训练好的模型即可产生对应的文本图像,无需再使用鉴别器、图像语义分割模型等模型。

[0114] 以本实施例为例,初始文本图像二值图向量和原始生成图像二值图向量可以帮助模型获得更多原始图像的真实图像信息,使得模型更新的参数更加符合空间正确性。可选地,本申请的实施方式并不受上述实施例的限制。

[0115] 实施例二

[0116] 为了执行上述实施例一对应的方法,以实现相应的功能和技术效果,下面提供一种文本图像生成装置,如图2所示,该装置包括:

[0117] 获取模块1,用于获取文本和文本对应的原始图像;

[0118] 转换模块2,用于根据文本获得文本特征向量;

[0119] 生成模块3,用于将文本特征向量输入到多级生成对抗模型中,获得初始文本图像;

[0120] 语义分割模块4,用于将初始文本图像及原始图像分别输入到语义分割模型中,获得初始文本图像空间信息二值图以及原始图像空间信息二值图;

[0121] 图像生成模块5,用于根据初始文本图像空间信息二值图及原始图像空间信息二值图获得文本图像。

[0122] 进一步地,转换模块2用于:

[0123] 根据原始图像和文本获得所述文本与原始图像的关联度;

[0124] 根据关联度获得文本特征向量。

[0125] 进一步地,转换模块2还用于:

[0126] 获取原始图像的目标区域;

[0127] 获取文本中的词汇数据;

[0128] 根据目标区域和词汇数据获得文本与原始图像的关联度。

[0129] 进一步地,转换模块2还用于:

[0130] 获得词汇数据的嵌入向量;

[0131] 将嵌入向量和关联度输入多级生成对抗模型,获得文本特征向量。

[0132] 进一步地,图像生成模块5还用于:

[0133] 将初始文本图像空间信息二值图及原始图像空间信息二值图输入到空间约束判别模型中,分别获得初始文本图像二值图向量和原始生成图像二值图向量;

[0134] 根据初始文本图像二值图向量和原始生成图像二值图向量获得文本图像。

[0135] 空间约束判别模型对应的空间约束公式为:

$$[0136] \quad \sum_i || \varphi(f_{t,i}) - \varphi(f_{g,i}) ||^2;$$

[0137] 其中, $f_{t,i}$ 为初始文本图像二值图向量, $f_{g,i}$ 为原始生成图像二值图向量。

[0138] 进一步地,生成模块3还用于:

[0139] 多级生成对抗模型的鉴别器损失函数为:

$$[0140] \quad L_D = E_{(I,t) \sim p_{data}} [\log D(I, \varphi_t)] + E_{s \sim p_G, t \sim p_{data}} [\log(1 - D(G(s, c), \varphi_t))] + \sum_i || \phi(I) - \phi(G(s, c)) ||^2 ;$$

[0141] 其中, $E_{(I,t) \sim p_{data}}$ 为原始图像的均方误差, $\varphi_t$ 为文本的特征向量, $E_{s \sim p_G, t \sim p_{data}}$ 为初始文

本图像的均方误差,  $G()$  表示生成器的输出结果,  $D()$  表示鉴别器的输出结果,  $I$  为原始图像,  $s$  为文本,  $c$  为噪声,  $L_D$  为鉴别器损失函数。

[0142] 多级生成对抗模型的生成器损失函数为:

$$[0143] \quad L_G = E_{s \sim p_G, t \sim p_{data}} [\log(1 - D(G(s, c), \varphi_t))];$$

[0144] 其中,  $E_{s \sim p_G, t \sim p_{data}}$  为初始文本图像的均方误差,  $G()$  表示生成器的输出结果,  $D()$  表示鉴别器的输出结果,  $I$  为原始图像,  $s$  为文本,  $c$  为噪声,  $\varphi_t$  为文本的特征向量,  $L_G$  为生成器损失函数。

[0145] 上述的文本图像生成装置可实施上述实施例一的方法。上述实施例一中的可选项也适用于本实施例, 这里不再详述。

[0146] 本申请实施例的其余内容可参照上述实施例一的内容, 在本实施例中, 不再进行赘述。

[0147] 实施例三

[0148] 本申请实施例提供一种电子设备, 包括存储器及处理器, 该存储器用于存储计算机程序, 该处理器运行计算机程序以使电子设备执行实施例一的文本图像生成方法。

[0149] 可选地, 上述电子设备可以是服务器。

[0150] 请参见图3, 图3为本申请实施例提供的电子设备的结构组成示意图。该电子设备可以包括处理器31、通信接口32、存储器33和至少一个通信总线34。其中, 通信总线34用于实现这些组件直接的连接通信。其中, 本申请实施例中设备的通信接口32用于与其他节点设备进行信令或数据的通信。处理器31可以是一种集成电路芯片, 具有信号的处理能力。

[0151] 上述的处理器31可以是通用处理器, 包括中央处理器 (Central Processing Unit, CPU)、网络处理器 (Network Processor, NP) 等; 还可以是数字信号处理器 (DSP)、专用集成电路 (ASIC)、现成可编程门阵列 (FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器31也可以是任何常规的处理器等。

[0152] 存储器33可以是, 但不限于, 随机存取存储器 (Random Access Memory, RAM), 只读存储器 (Read Only Memory, ROM), 可编程只读存储器 (Programmable Read-Only Memory, PROM), 可擦除只读存储器 (Erasable Programmable Read-Only Memory, EPROM), 电可擦除只读存储器 (Electric Erasable Programmable Read-Only Memory, EEPROM) 等。存储器33中存储有计算机可读取指令, 当计算机可读取指令由所述处理器31执行时, 设备可以执行上述图1方法实施例涉及各个步骤。

[0153] 可选地, 电子设备还可以包括存储控制器、输入输出单元。存储器33、存储控制器、处理器31、外设接口、输入输出单元各元件相互之间直接或间接地电性连接, 以实现数据的传输或交互。例如, 这些元件相互之间可通过一条或多条通信总线34实现电性连接。处理器31用于执行存储器33中存储的可执行模块, 例如设备包括的软件功能模块或计算机程序。

[0154] 输入输出单元用于提供给用户创建任务以及为该任务创建启动可选时段或预设执行时间以实现用户与服务器的交互。输入输出单元可以是, 但不限于, 鼠标和键盘等。

[0155] 可以理解, 图3所示的结构仅为示意, 电子设备还可包括比图3中所示更多或者更少的组件, 或者具有与图3所示不同的配置。图3中所示的各组件可以采用硬件、软件或其组

合实现。

[0156] 本申请实施例还提供一种计算机程序产品,该计算机程序产品在计算机上运行时,使得计算机执行方法实施例所述的方法。

[0157] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,也可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,附图中的流程图和框图显示了根据本申请的多个实施例的装置、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现方式中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0158] 另外,在本申请各个实施例中的各功能模块可以集成在一起形成一个独立的部分,也可以是各个模块单独存在,也可以两个或两个以上模块集成形成一个独立的部分。

[0159] 所述功能如果以软件功能模块的形式实现并作为独立的产品销售或使用,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,本申请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)执行本申请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、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0160] 以上所述仅为本申请的实施例而已,并不用于限制本申请的保护范围,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步定义和解释。

[0161] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应所述以权利要求的保护范围为准。

[0162] 需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

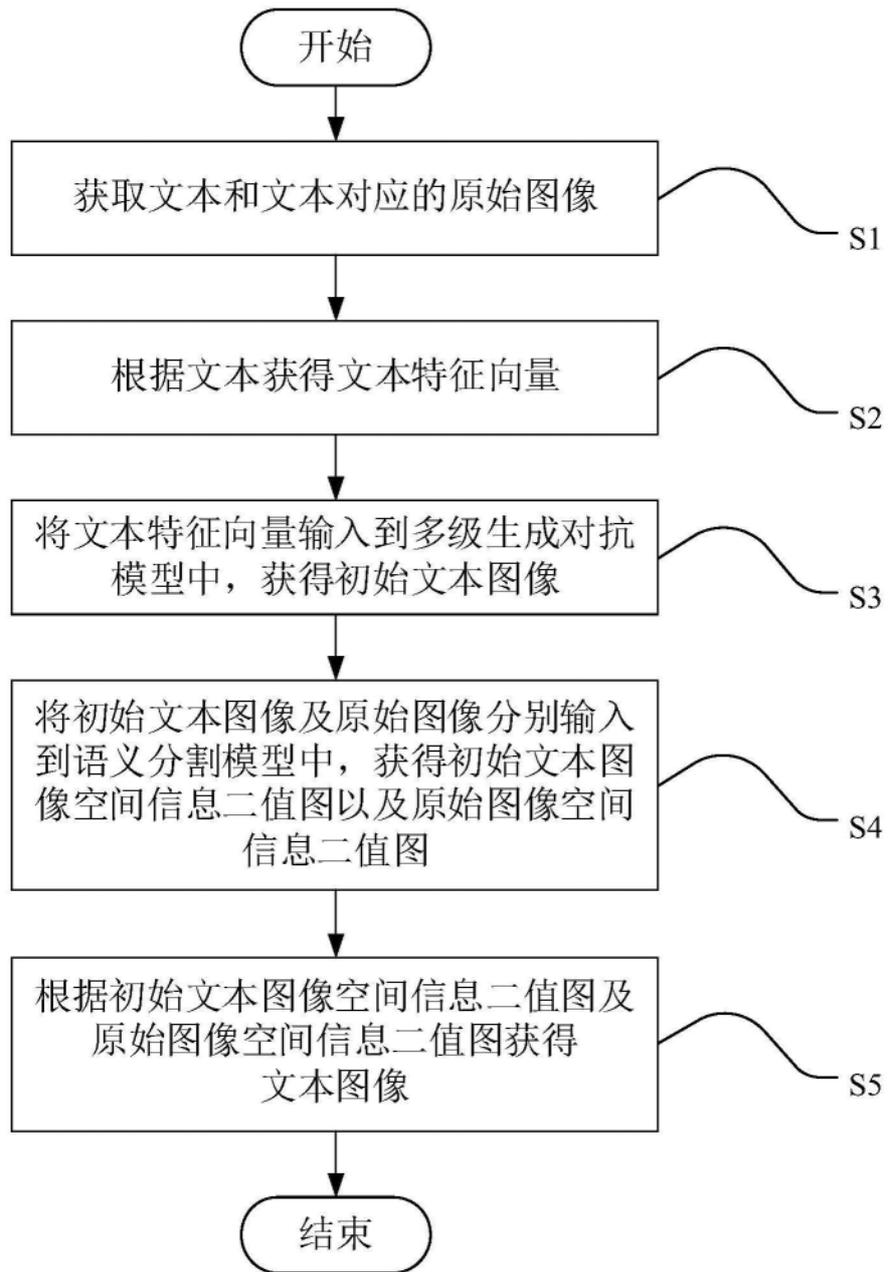


图1

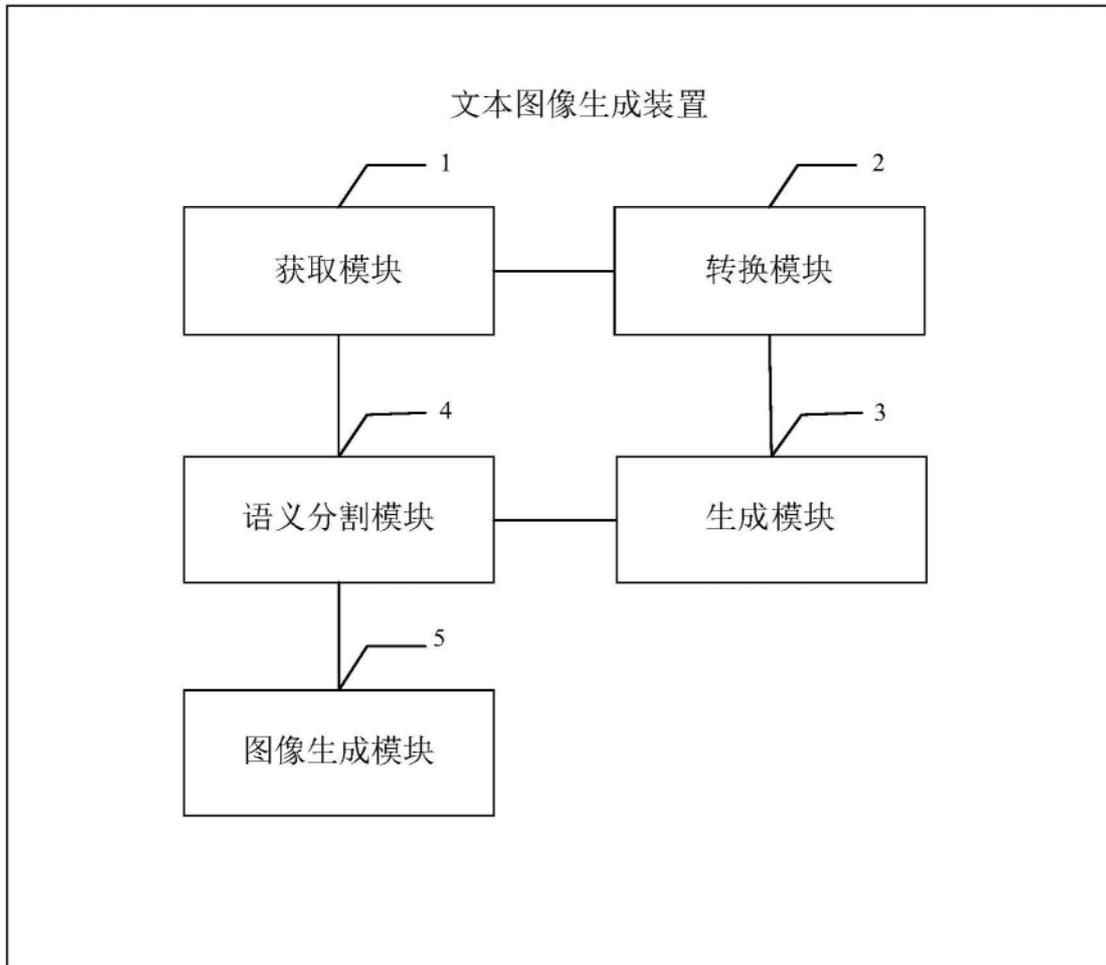


图2

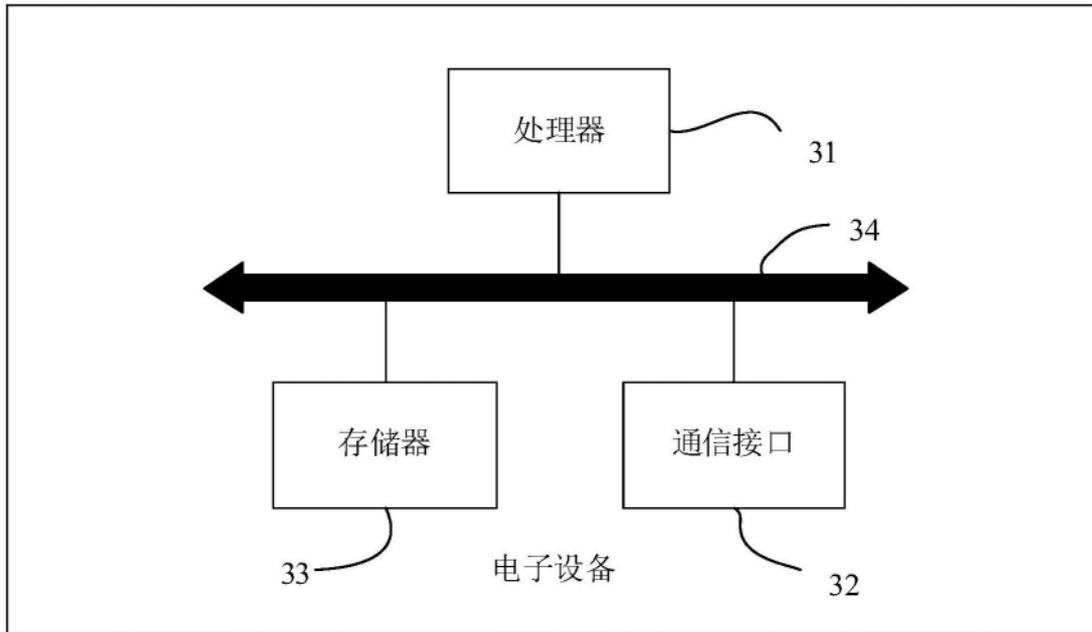


图3