



(12) 发明专利

(10) 授权公告号 CN 111445545 B

(45) 授权公告日 2023. 08. 18

(21) 申请号 202010124986.9  
 (22) 申请日 2020.02.27  
 (65) 同一申请的已公布的文献号  
 申请公布号 CN 111445545 A  
 (43) 申请公布日 2020.07.24  
 (73) 专利权人 北京大米未来科技有限公司  
 地址 100123 北京市朝阳区青年路7号院达  
 美中心T4-27层  
 (72) 发明人 谢文珍 黄恺 冯富森  
 (74) 专利代理机构 北京恒博知识产权代理有限  
 公司 11528  
 专利代理师 李宁宁  
 (51) Int. Cl.  
 G06T 11/00 (2006.01)  
 (56) 对比文件  
 CN 109522975 A, 2019.03.26  
 CN 110019675 A, 2019.07.16  
 JP 5866064 B2, 2016.02.17  
 US 10074200 B1, 2018.09.11

US 2005031219 A1, 2005.02.10  
 CN 110413788 A, 2019.11.05  
 CN 110705208 A, 2020.01.17  
 CN 109493400 A, 2019.03.19  
 CN 110489747 A, 2019.11.22  
 CN 110532381 A, 2019.12.03  
 US 2022005235 A1, 2022.01.06  
 CN 114970513 A, 2022.08.30  
 US 2010238180 A1, 2010.09.23  
 CN 109448132 A, 2019.03.08  
 US 5428754 A, 1995.06.27  
 US 5802381 A, 1998.09.01  
 US 2018341750 A1, 2018.11.29  
 Tao Xu et al..Fine-grained text to  
 image generation with attentional  
 generative diarsarial networks.《2018  
 IEEE/CVF Conference on Computer Vision  
 and Pattern Recognition》.2018,第1316-1324  
 页.

审查员 颜欣

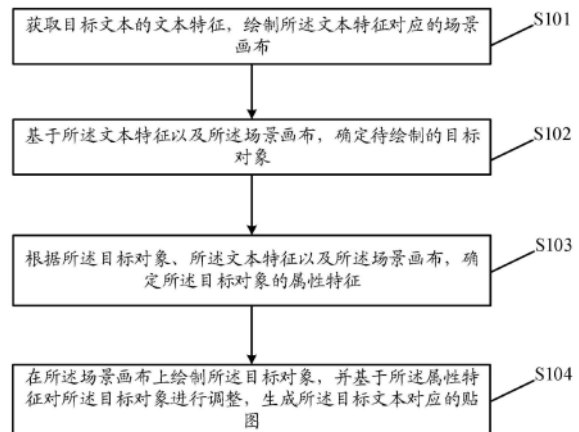
权利要求书2页 说明书19页 附图9页

(54) 发明名称

一种文本转贴图方法、装置、存储介质及电子设备

(57) 摘要

本申请实施例公开了一种文本转贴图方法、装置、存储介质及电子设备,其中,方法包括:获取目标文本的文本特征,绘制所述文本特征对应的场景画布,基于所述文本特征以及所述场景画布,确定待绘制的目标对象,根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征,在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。采用本申请实施例,可以提升贴图与文本实际描述场景的契合度,提高了生成贴图的准确度。



1. 一种文本转贴图方法,其特征在于,所述方法包括:
  - 获取目标文本的文本特征,绘制所述文本特征对应的场景画布;
  - 将所述场景画布输入至卷积网络中进行场景编码,输出场景编码后的场景特征图;
  - 基于所述文本特征以及所述场景画布,确定待绘制的目标对象;
  - 将所述文本特征以及所述目标对象输入至第二文本关注器,输出第二文本注意力向量;
  - 将所述场景特征图以及所述第二文本注意力向量输入至场景卷积网络中,输出第二场景注意力向量;
  - 将所述第二场景注意力向量、所述目标对象以及所述第二文本注意力向量输入至属性卷积网络中,输出所述目标对象的属性特征;
  - 在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。
2. 根据权利要求1所述的方法,其特征在于,所述获取目标文本的文本特征,绘制所述文本特征对应的场景画布,包括:
  - 将目标文本输入至文本编码器中,输出所述目标文本对应的文本特征;
  - 提取所述文本特征对应的场景主题,在预设的贴图索引库中索引所述场景主题对应的场景贴图,将所述场景贴图确定为场景画布。
3. 根据权利要求1所述的方法,其特征在于,所述基于所述文本特征以及所述场景画布,确定待绘制的目标对象,包括:
  - 将所述文本特征以及所述场景特征图输入至对象解码器中,输出所述目标对象。
4. 根据权利要求3所述的方法,其特征在于,所述方法还包括:
  - 获取已绘制过的历史对象;
  - 将所述文本特征以及所述场景特征图输入至对象解码器中,输出所述目标对象,包括:
    - 对所述场景特征图进行池化处理,得到池化处理后的第一场景注意力向量;
    - 将所述第一场景注意力向量、所述历史对象以及所述文本特征输入至第一文本关注器,输出第一文本注意力向量;
    - 将所述第一场景注意力向量、所述历史对象以及所述第一文本注意力向量输入至对象卷积网络中,输出所述目标对象。
5. 根据权利要求1所述的方法,其特征在于,所述在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图,包括:
  - 在预设的贴图索引库中索引所述目标对象对应的贴图对象;
  - 将所述属性特征、所述贴图对象以及所述场景画布输入至画布调整模型中,输出所述目标文本对应的贴图。
6. 一种文本转贴图装置,其特征在于,所述装置包括:
  - 场景画布绘制模块,用于获取目标文本的文本特征,绘制所述文本特征对应的场景画布;
  - 场景特征图编码模块,用于将所述场景画布输入至卷积网络中进行场景编码,输出场景编码后的场景特征图;

目标对象确定模块,用于基于所述文本特征以及所述场景画布,确定待绘制的目标对象;

属性特征确定模块,包括:文本向量输出单元,用于将所述文本特征以及所述目标对象输入至第二文本关注器,输出第二文本注意力向量;场景向量输出单元,用于将所述场景特征图以及所述第二文本注意力向量输入至场景卷积网络中,输出第二场景注意力向量;属性特征输出单元,用于将所述第二场景注意力向量、所述目标对象以及所述第二文本注意力向量输入至属性卷积网络中,输出所述目标对象的属性特征;

贴图生成模块,用于在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。

7.一种计算机存储介质,其特征在于,所述计算机存储介质存储有多条指令,所述指令适于由处理器加载并执行如权利要求1~5任意一项的方法步骤。

8.一种电子设备,其特征在于,包括:处理器和存储器;其中,所述存储器存储有计算机程序,所述计算机程序适于由所述处理器加载并执行如权利要求1~5任意一项的方法步骤。

## 一种文本转贴图方法、装置、存储介质及电子设备

### 技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种文本转贴图方法、装置、存储介质及电子设备。

### 背景技术

[0002] 文本转贴图作为文本生成图像技术的一种应用,可以将文本转化为生动形象的图片,例如,在线教育中,可以将教学文本(如学生语言写作或语言对话的文本)转化为图片,从而激发学生的语言学习的兴趣,提升在线教学效果。

[0003] 目前,文本转贴图过程中,通常是对文本进行文本编码,根据编码之后的关键词(如文本中的名词)去贴图库中索引贴图对象,将贴图对象在画布上组合生成贴图,然而采用根据关键词组合生成贴图这种方式,会忽视文本实际所表达的语义,导致贴图难以契合文本实际描述的场景,影响了生成贴图的准确度。

### 发明内容

[0004] 本申请实施例提供了一种文本转贴图方法、装置、存储介质及电子设备,可以提升贴图与文本实际描述场景的契合度,提高了生成贴图的准确度。所述技术方案如下:

[0005] 第一方面,本申请实施例提供了一种文本转贴图方法,所述方法包括:

[0006] 获取目标文本的文本特征,绘制所述文本特征对应的场景画布;

[0007] 基于所述文本特征以及所述场景画布,确定待绘制的目标对象;

[0008] 根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征;

[0009] 在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。

[0010] 第二方面,本申请实施例提供了一种文本转贴图装置,所述装置包括:

[0011] 场景画布绘制模块,用于获取目标文本的文本特征,绘制所述文本特征对应的场景画布;

[0012] 目标对象确定模块,用于基于所述文本特征以及所述场景画布,确定待绘制的目标对象;

[0013] 属性特征确定模块,用于根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征;

[0014] 贴图生成模块,用于在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。

[0015] 第三方面,本申请实施例提供一种计算机存储介质,所述计算机存储介质存储有多条指令,所述指令适于由处理器加载并执行上述的方法步骤。

[0016] 第四方面,本申请实施例提供一种电子设备,可包括:处理器和存储器;其中,所述存储器存储有计算机程序,所述计算机程序适于由所述处理器加载并执行上述的方法步

骤。

[0017] 本申请一些实施例提供的技术方案带来的有益效果至少包括：

[0018] 在本申请一个或多个实施例中，终端获取目标文本的文本特征，绘制所述文本特征对应的场景画布，基于所述文本特征以及所述场景画布，确定待绘制的目标对象，根据所述目标对象、所述文本特征以及所述场景画布，确定所述目标对象的属性特征，在所述场景画布上绘制所述目标对象，并基于所述属性特征对所述目标对象进行调整，生成所述目标文本对应的贴图。通过目标文本的文本特征分步确定目标文本所对应的场景画布、待绘制的对象以及对象的属性特征，在场景画布绘制对象时基于对象的属性特征（位置、动作、姿态等）对图像进行相应调整，可以避免根据关键词导致生成贴图难以契合文本实际描述场景的问题，可以生成一个场景准确、对象清楚以及属性清晰的贴图，贴图可以更贴近文本实际所表达的语义，可以提升了贴图与文本实际描述场景的契合度，从而提高了生成贴图的准确度。

### 附图说明

[0019] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术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。

[0020] 图1是本申请实施例提供的一种文本转贴图方法的流程示意图；

[0021] 图2是本申请实施例提供的文本转贴图方法涉及的一种属性确定模型的结构示意图；

[0022] 图3是本申请实施例提供的文本转贴图方法涉及的另一种属性确定模型的结构示意图；

[0023] 图4是本申请实施例提供的另一种文本转贴图方法的流程示意图；

[0024] 图5~图10是本申请实施例提供的文本转贴图方法涉及的转换贴图的举例示意图；

[0025] 图11是本申请实施例提供的一种文本转贴图装置的结构示意图；

[0026] 图12是本申请实施例提供的一种场景画布绘制模块的结构示意图；

[0027] 图13是本申请实施例提供的一种目标对象确定模块的结构示意图；

[0028] 图14是本申请实施例提供的一种属性特征确定模块的结构示意图；

[0029] 图15是本申请实施例提供的一种贴图生成模块的结构示意图；

[0030] 图16是本申请实施例提供的另一种文本转贴图装置的结构示意图；

[0031] 图17是本申请实施例提供的一种电子设备的结构示意图。

### 具体实施方式

[0032] 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清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0033] 在本申请的描述中,需要理解的是,术语“第一”、“第二”等仅用于描述目的,而不能理解为指示或暗示相对重要性。在本申请的描述中,需要说明的是,除非另有明确的规定和限定,“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元,而是可选地还包括没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。对于本领域的普通技术人员而言,可以具体情况理解上述术语在本申请中的具体含义。此外,在本申请的描述中,除非另有说明,“多个”是指两个或两个以上。“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[0034] 下面结合具体的实施例对本申请进行详细说明。

[0035] 在一个实施例中,如图1所示,特提出了一种文本转贴图方法,该方法可依赖于计算机程序实现,可运行于基于冯诺依曼体系的文本转贴图装置上。该计算机程序可集成在应用中,也可作为独立的工具类应用运行。

[0036] 具体的,该文本转贴图方法包括:

[0037] 步骤101:获取目标文本的文本特征,绘制所述文本特征对应的场景画布。

[0038] 所述目标文本指的是包含文本内容的文本,如,所述目标文本可以是在线教育场景中采集用户写作时的文本。通常目标文本所包含的文本内容可以理解为一种语言的书面表示形式,通常是具有完整、系统含义的一个句子或多个句子的组合。所述目标文本以英语语言为例,可以是至少一个词、至少一个句子、至少一个段落,所述目标文本可以是语言的实际运用形态,通常在具体实施中指一些语言文字。

[0039] 所述文本特征是指以文字表达的非结构化数据特有的文本属性,以一篇写作作文为例,文本特征包括作者创作意图、数据、主题说明、底层特征含义等文本要素。通过文本特征可以表达待描述对象本身的语义以及待描述对象在语言环境中语义的多种特征。以目标文本所包含的文本内容为英文短文例,所述文本特征可以是组成字母、词的顺序、词的情感信息、互信息等。

[0040] 其中,组成字母即为一个单词由哪些字母组成,这些字母的先后顺序关系。

[0041] 词顺序为表达一句话(一个意思)组成的各个单词的先后顺序。

[0042] 词的情感信息为该词语在这个句子中所表达的情感含义,所述情感含义可以理解为词在句子是褒义还是贬义、是高昂还是低沉、是欣喜还是悲伤等。

[0043] 互信息是指某个词或字和类别之间的统计独立关系,互信息常应用于度量两个对象之间的相互性。

[0044] 具体的,终端在采集到用户所输入的目标文本之后,利用文本特征获取模型获取所述目标文本的文本特征。

[0045] 可选的,所述文本特征获取模型可以是基于语境框架的文本特征信息获取方法,即先确定文本内容的特征元素(句、词、字、符号等),然后将语言语义分析融入统计算法对所述目标文本所包含的文本内容进行抽取处理,得到所述目标文本的文本特征;可以是基于本体论的文本特征获取方法,即利用本体论(On-tology)模型将所述文本内容作为输入,输出所述目标文本的文本特征;可以是基于知网的概念特征提取方法,即基于概念特征的

特征获取方法,在向量空间模型(Vector Space Model,VSM)的基础上,对所述文本内容进行语言语义分析,利用知网的数据库获取词汇的语言语义信息,将语言语义相同的词汇映射到同一主题概念,然后聚类得到聚类后的词,并作为VSM模型的文本向量的特征项,然后进行模型运算等,等等。需要说明的是所述获取所述目标文本的文本特征的方式有很多,可以是上述的一种或多种的拟合,此处不作限定。

[0046] 具体的,终端获取目标文本的文本特征之后,绘制所述文本特征对应的场景画布。其中,所述场景画布可以理解为贴图的背景图,或待添加元素的初始场景图等。

[0047] 在一种可行的实施方式中,终端获取目标文本的文本特征之后,可以基于文本语境的关键特征提取方法,即提取文本特征的关键元素(关键句、关键词、关键字、关键符号等),然后将语义分析融入统计算法对所述文本特征所包含的关键元素进行抽取处理,并确定关键元素所对应的场景主题(如沙滩、大山、河流等场景主题),然后根据场景主题在预设的贴图索引库中匹配对应的场景画布。

[0048] 步骤102:基于所述文本特征以及所述场景画布,确定待绘制的目标对象。

[0049] 所述目标对象可以理解为可绘制或可插入的任一绘图图形,可对这些图形进行更改和完善。绘图图形包含自选图形、曲线、线条等,在本申请实施例中,所述目标对象可以是待绘制的人物(如卡通人物)、动物、植物、交通工具、建筑物等等。例如,一个少年、一块海绵、一只鸭子、一头狼等都可以成为待绘制的目标对象。

[0050] 具体的,终端在获取到目标文本的文本特征以及根据所述文本特征绘制对应的场景画布之后,将所述文本特征以及场景画布输入至基于注意力机制的对象确定模型中,从而输出待绘制的目标对象。以所述文本特征以及场景画布作为神经网络模型-对象确定模型的输入,来预测绘制场景画布的每一时间步的待绘制的目标对象,例如第一时间步 $t_1$ 预测待绘制一个少年,第二时间步 $t_2$ 预测待绘制一只鸭子。

[0051] 在本申请实施例中,注意力机制可以包括至少两个方面:决定需要关注输入的哪部分以及分配有限的信息处理资源给重要的部分。对场景画布以及文本特征引入注意力机制可以凸显出场景画布中更关键的图像部分,例如当前场景画布中优先级较高的待绘制的目标对象。例如在一种具体的应用场景中,所述目标文本中会包含描述对象、对象的属性信息、场景信息等等。引入注意力机制的对象确定模型后,可以突出文本特征所包含的对象部分,而对于特征图中非对象部分(颜色、情绪、动作、位置等属性信息)进行弱化,便于在后续的处理中着重关注突出的部分。以确定接下来待绘制的目标对象。

[0052] 在本实施例中,所述对象确定模型为神经网络模型,所述神经网络模型是由许多节点中各节点的简单非线性模拟处理要素密集互连配置而成的,是一种模仿了生物神经元的系统模型。所述神经网络模型通过使至少一个节点的输入与每一个节点的输出连接形成,类似于真实神经元的突触连接。每个神经元表达了一种特定的输出函数,即激励函数,每两个神经元之间的连接都包含一个连接强度,即作用于通过该连接的信号的加权值。在本实施例中,将大量的场景画布和其对应的文本特征输入至基于注意力机制的神经网络模型进行训练,可以得到训练之后的对象确定模型,该对象确定模型具有目标对象确定过程中关键信息特征抽取、语义知识概括和学习记忆的能力,通常所述神经网络模型学习到的信息或知识储存在每个单元节点之间的连接矩阵上。

[0053] 可选的,所述神经网络模型可以是基于卷积神经网络(Convolutional Neural

Network,CNN)模型,深度神经网络(Deep Neural Network,DNN)模型、循环神经网络(Recurrent Neural Networks,RNN)、模型、嵌入(embedding)模型、梯度提升决策树(Gradient Boosting Decision Tree,GBDT)模型、逻辑回归(Logistic Regression,LR)模型等模型中的一种或多种的拟合实现的。

[0054] 具体的,终端在获取到大量的包含文本特征以及场景画布的样本数据时,对所述样本数据进行标注,所述标注可以理解为对所述样本数据对应的关键信息(待绘制的对象)进行标注,将所述文本特征以及场景画布输入至初始的对象确定模型进行训练,基于已经标注的样本数据对对象确定模型进行训练,可以得到训练好的对象确定模型。

[0055] 步骤103:根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征。

[0056] 所述属性特征是指描述目标对象的特性或属性,包括但不限于面貌特征、衣着特征、情感特征、行为特征等。例如,“凶神恶煞”、“漂亮”、“温婉贤淑”、“灰色”“风流倜傥”等文字都可以用于描述目标对象的特性。又例如,奔跑、打斗、做鬼脸、静止、玩耍、陪伴等文字都可以用于描述目标对象的特性。

[0057] 具体的,终端在确定待绘制的目标对象之后,将所述目标对象、所述文本特征以及所述场景画布输入至基于注意力机制的属性确定模型中,从而输出目标对象的属性特征。以所述目标对象、所述文本特征以及所述场景画布作为神经网络模型-属性确定模型的输入,来预测绘制场景画布的每一时间步所对应目标对象的属性特征,例如第一时间步 $t_1$ 预测待绘制一个少年的表情、行为等特性,第二时间步 $t_2$ 预测待绘制一只鸭子的面貌、行为等特性。

[0058] 在本申请实施例中,注意力机制可以包括至少两个方面:决定需要关注输入的哪部分以及分配有限的信息处理资源给重要的部分。对所述目标对象、所述文本特征以及所述场景画布引入注意力机制可以凸显出场景画布中更关键的图像部分,例如当前场景画布中优先级较高的目标对象的属性特征。例如在一种具体的应用场景中,所述目标文本中会包含描述对象、对象的属性信息、场景信息等等。引入注意力机制的对象确定模型后,可以突出文本特征所包含的属性部分(位置、颜色、情绪、动作、位置等属性信息),而对于特征图中对象部分(人物、动物、植物等对象)进行弱化,便于在后续的处理中着重关注突出的部分。以确定接下来待绘制的目标对象对应的属性特征。

[0059] 其中,基于注意力机制的属性确定模型可以基于输入的文本特征关注目标文本的上下文语义,对待绘制目标对象的相关内容(即属性信息)进行编码,以及在场景画布中确定目标对象的属性信息-位置信息等,最终由属性确定模型编码后输出所述目标对象的属性特征。

[0060] 可选的,所述基于注意力机制的属性确定模型可以是基于seq2seq框架中的decoder模型。如图2所示,图2是一种属性确定模型的结构示意图,在图2中,所述目标对象 $x_1$ 、所述文本特征 $x_2$ 以及所述场景画布 $x_3$ 构成属性确定模型的输入 $x$ , $h_1$ 、 $h_2$ ... $h_n$ 构成decoder模型中的神经网络计算单元,每两个神经网络计算单元之间的连接都包含一个连接强度,即作用于通过该连接的信号的加权值。该对象确定模型具有目标对象确定过程中关键信息特征抽取、语义知识概括和学习记忆的能力,通常所述神经网络模型学习到的信息或知识储存在每个神经网络计算单元之间的连接矩阵上。需要说明的是,在如图2所示的



模型中,可以看到上一时刻的输出属性 $y$ 会作为当前时刻的输入,而且其中“输入 $x$ ”只作为初始状态参与运算,后面的运算都与“输入 $x$ ”无关。如,上一时刻的神经网络计算单元 $h_1$ 对应的输出属性 $y_1$ 会作为当前时刻的神经网络计算单元 $h_2$ 的输入参与计算。从而输出包含属性 $y_1$ 、属性 $y_2 \dots$ 属性 $y_n$ 的输出 $y$ (即属性特征)。

[0061] 在一种可行的实施方式中,所述属性确定模型的结构可以是如图3所示的一种decoder模型结构。在如图3所示的模型中,可以看到上一时刻的输出属性 $y$ 会作为当前时刻的输入,而且其中“输入 $x$ ”作为稳态参与每个神经网络计算单元的运算,即后面的运算都与“输入 $x$ ”有关。如,上一时刻的神经网络计算单元 $h_1$ 对应的输出属性 $y_1$ 会作为当前时刻的神经网络计算单元 $h_2$ 的输入参与计算。从而输出包含属性 $y_1$ 、属性 $y_2 \dots$ 属性 $y_n$ 的输出 $y$ (即属性特征)。其中“输入 $x$ ”作为稳态参与神经网络计算单元 $h_2$ 运算,且后面的运算都与“输入 $x$ ”有关。

[0062] 在本申请实施例中,在创建初始属性确定模型之后,获取大量的包含场景画布、目标对象以及所述文本特征的样本数据输入至基于注意力机制的神经网络模型进行训练,可以得到训练之后的属性确定模型。其中,在对所述属性确定模型进行训练时可以是采用基于动态时间规整的训练方法(DTW),可以是基于矢量量化的训练方法(VQ),还可以是基于图像信号的时间序列的训练方法(HMM),等等。

[0063] 步骤104:在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。

[0064] 所述贴图可以理解为终端在执行本申请实施例的文本转贴图方法最终将目标文本转换为相应文本语义的图像。

[0065] 在一种可行的实施方式中,终端可以调用对象生成程序在场景画布上初步生成目标对象的特征贴图。特征贴图是基于对象特征向量初步生成的图像。在本实施例中,特征贴图通常可以是低分辨率的图像,例如分辨率为 $32 \times 32$ 或 $64 \times 64$ 的图像。一个对象在终端生对应应有相应的对象特征向量,终端可以通过获取所述目标对象对应的对象特征向量,然后根据对象特征向量以生成特征贴图。并可以同步或异步基于所述属性特征对场景画布上目标对象的特征贴图进行图像调整,需要说明的是,所述图像调整的对象还包括对所述目标对象相关联的图像部分进行调整,例如对所述目标对象相关联的图像背景(植物、动物等)进行调整,在一个具体的应用场景中,例如采用幼鸟数据集(CUB),输入待处理文本为“一只胸前为白色的灰色小鸟,灰色小鸟的心情不好”,终端根据小鸟特征向量以在场景画布上生成小鸟的特征贴图,并可以同步或异步根据小鸟对象所对应的的属性特征,对场景画布上目标对象的特征贴图进行图像调整:将小鸟的外貌特征调整为皮毛为灰色、小鸟胸部区域为白色,将小鸟情绪特征调整为面部表征为心情低沉、与小鸟关联的图像部分-场景环境表征为阴雨天(如添加阴云、雨滴,将场景亮度、对比度对应调整)等。待根据属性特征对所述目标对象进行调整之后,即可得到调整之后的场景画布,即所述目标文本对应的贴图。

[0066] 在一种可行的实施方式中,终端储存有贴图索引库,所述贴图索引库至少存储有多个贴图元素(对象所对应的贴图),终端可以在贴图索引库中获取所述目标对象所对应的贴图对象,在场景画布上添加该贴图对象,并可以同步或异步基于所述属性特征对场景画布上目标对象的特征贴图进行图像调整,所述图像调整的对象还包括对所述目标对象相关联的图像部分进行调整,例如对所述目标对象相关联的图像背景(植物、动物等)进行调整。

[0067] 在本申请实施例中,终端获取目标文本的文本特征,绘制所述文本特征对应的场景画布,基于所述文本特征以及所述场景画布,确定待绘制的目标对象,根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征,在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。通过目标文本的文本特征分步确定目标文本所对应的场景画布、待绘制的对象以及对象的属性特征,在场景画布绘制对象时基于对象的属性特征(位置、动作、姿态等)对图像进行相应调整,可以避免根据关键词导致生成贴图难以契合文本实际描述场景的问题,可以生成一个场景准确、对象清楚以及属性清晰的贴图,贴图可以更贴近文本实际所表达的语义,可以提升了贴图与文本实际描述场景的契合度,从而提高了生成贴图的准确度。

[0068] 请参见图4,图4是本申请提出的一种文本转贴图方法的另一种实施例的流程示意图。具体的:

[0069] 步骤201:将目标文本输入至文本编码器中,输出所述目标文本对应的文本特征。

[0070] 所述文本编码器用于对输入文本编码器的目标文本输出编码好的特征表达,即输出目标文本的所对应的文本特征,在实际应用中所述文本特征通常以编码向量的形式进行表征。

[0071] 具体的,文本编码器可以利用深度神经网络将目标文本进行压缩,得到每个时刻对应的编码向量;具体方式为使用深度神经网络中的长短时记忆网络模型,将目标文本的每个文本元素(词、句子、符号等)依次输入网络,得到每个时刻(即时间步)对应的隐层表示 $h_i$ 。

[0072] 具体地,使用深度神经网络中的长短时记忆网络(LSTM),在开始时刻输入目标文本中的一个文本元素(如一个字词、一个字符等)到长短时记忆网络(LSTM),将这个字词压缩成一个向量,然后将压缩后得到的向量传入下一时刻;下一时刻循环神经网络输入上一时刻的压缩向量和原文本的下一个文本元素,将二者压缩成新的向量传入下一时刻;在压缩完所有文本后得到的每一时刻的编码向量,即是在对编码向量进行在解码(包含属性解码、对象解码、场景解码等)时需要的特征信息;本实施例的时刻数为句子的字词个数;每个时刻对应的隐层向量为长短时记忆网络(LSTM)将字词压缩成的向量,经上述编码过程即可得到所述目标文本对应的文本特征。

[0073] 在一种具体的实施场景中,文本编码器可以是基于LSTM的演化网络GRU网络构成,即采用带有门控循环单元(GRU)的单层双向循环网络(BiGRU)组成。它将文本元素-每个字词的线性嵌入作为输入,并且每个方向的隐藏维数可以是固定的维数,如所述维数可均为256维。这里采用来自GloVe方法的预训练参数来初始化由单词嵌入网络构成的文本编码器。将词嵌入向量作为文本特征被固定为抽象场景和语义布局生成,用于后续步骤的合成图像-贴图的生成。

[0074] 例如,文本编码器的编码过程可由下式表征,对于给定的目标文本,我们为每个文本元素-字词计算:

$$[0075] \quad h_i^E = BiGRU(X_i, h_{i-1}^E, h_{i+1}^E)$$

[0076] 这里BiGRU表示带有双向GRU单元的神经网络, $X_i$ 是与第*i*个单词对应的单词嵌入矢量,而 $h_i^E$ 是编码文本元素所对应的上下文隐藏矢量(以特征向量形式表示),这里由上下

文隐藏矢量  $h_i^E$  和单词嵌入矢量  $x_i$  组成的矢量对  $[h_i^E, x_i]$  即为文本编码器的输出-文本特征。

[0077] 步骤202:提取所述文本特征对应的场景主题,在预设的贴图索引库中索引所述场景主题对应的场景贴图,将所述场景贴图确定为场景画布。

[0078] 所述贴图索引库为预先建立的包含大量贴图素材的图像库,所述图像库包含多个场景主题对应的场景贴图,在本申请实施例中,终端可以在贴图索引库查找场景主题对应的场景贴图。

[0079] 具体的,终端获取目标文本的文本特征之后,提取所述文本特征对应的场景主题,在预设的贴图索引库中索引所述场景主题对应的场景贴图,将所述场景贴图确定为场景画布。其中,所述场景画布可以理解为贴图的背景图,或待添加元素的初始场景图等。

[0080] 在一种可行的实施方式中,终端获取目标文本的文本特征之后,可以基于文本语境的关键特征提取方法,即提取文本特征的关键元素(关键句、关键词、关键字、关键符号等),然后将语义分析融入统计算法对所述文本特征所包含的关键元素进行抽取处理,并确定关键元素所对应的场景主题(如沙滩、大山、河流等场景主题),然后根据场景主题在预设的贴图索引库中匹配对应的场景贴图,待查找到所述场景贴图之后,可以将所述场景贴图确定为场景画布。

[0081] 步骤203:将所述场景画布输入至卷积网络中进行场景编码,输出场景编码后的场景特征图。

[0082] 所述场景特征图可以理解为通过卷积网络对待识别图像进行识别,得到的场景特征图。在本申请实施例中,所述场景特征图为通过卷积网络对场景画布进行识别得到的,通常场景特征图中包括至少一个场景特征值。

[0083] 例如,场景特征图可以通过卷积网络(CNN网络)提取待识别图像的场景特征值而得到的,具体过程可以为:

[0084] 卷积网络中包括一个或多个用于从场景画布的像素矩阵中提取特征信息的卷积核用卷积核按照一定的步长遍历待识别图像的像素矩阵,将所述场景画布输入至卷积网络中进行场景编码,即可以得到至少一个场景特征值,由至少一个场景特征值组成场景特征图。

[0085] 在一种具体的实施方式中,所述卷积网络可以由至少一个门控循环单元(ConvGRU单元)组成的卷积GRU网络。该卷积GRU网络中的每个卷积层都有一个 $3 \times 3$ 的卷积内核,步幅(即步长)为1,隐藏维数为512维。并填充每个卷积的输入,以便场景特征图具有与输入的场景画布相同的空间分辨率。通过在空间上复制文本编码器所输出的文本特征的最后一个隐藏状态来初始化卷积网络隐层的隐藏状态。

[0086] 示意性的,所述卷积网络可以由下式表征,采用场景画布  $B_t$  作为卷积网络的输入:

$$[0087] \quad h_t^D = \text{ConvGRU}(\Omega(B_t), h_{t-1}^D)$$

[0088] 其中,  $h_t^D$  为当前时间步的场景状态,也可理解为场景特征值,  $\text{ConvGRU}(\Omega)$  为带有门控循环单元的卷积网络,  $h_{t-1}^D$  为历史时间步的场景状态,这里各  $h_t^D$  表征场景中每个空间(网格)位置的时间动态的信息性,则可以以集合  $\{h_t^D\}$  表示卷积网络的输出-当前场景状态的  $C \times H \times W$  场景特征图,其中  $C$  是通道数(由输入通道数和输出通道数构成),  $H$  和  $W$  是卷积高度

和卷积宽度。

[0089] 步骤204:将所述文本特征以及所述场景特征图输入至对象解码器中,输出所述目标对象。

[0090] 所述对象解码器是一种基于注意力的神经网络模型,可根据所有可能对象的似然分数输出目标对象。它以场景特征图中的各循环场景状态、文本特征作为输入,输出当前时间步中待绘制的目标对象。在实际应用中,所述对象解码器为一种对象确定模型,所述对象解码器包括但不限于场景池化部分、文本关注部分以及对象卷积部分构成。

[0091] 在本实施例中,所述对象解码器是由许多节点中各节点的简单非线性模拟处理要素密集互连配置而成的,是一种模仿了生物神经元的系统模型。所述对象解码器通过使至少一个节点的输入与每一个节点的输出连接形成,类似于真实神经元的突触连接。每个神经元表达了一种特定的输出函数,即激励函数,每两个神经元之间的连接都包含一个连接强度,即作用于通过该连接的信号的加权值。在本实施例中,将大量的场景画布和其对应的文本特征输入至基于注意力机制的神经网络模型进行训练,可以得到训练之后的对象解码器,该对象解码器具有目标对象确定过程中关键信息特征抽取、语义知识概括和学习记忆的能力,通常所述对象解码器学习到的信息或知识储存在每个单元节点之间的连接矩阵上。其中,在将所述文本特征以及所述场景特征图输入至对象解码器中之后,对象解码器的场景池化部分会对所述场景特征图进行池化处理,以收集对象预测所需的场景空间上下文,例如当前时间步中场景特征图所对应的画布是否添加目标对象以及添加的历史对象,然后通过池化处理以控制池化神经单元将参与的对象空间特征融合到场景注意力向量中,对象解码器的文本关注部分对场景注意力向量以及文本特征进行处理,具体为通过注意力机制关注目标文本的语义上下文并突出所对应的的文本特征所包含的对象部分,而对于特征图中非对象部分(颜色、情绪、动作、位置等属性信息)进行弱化,以确定待绘制对象在文本特征中所包含的文本信息,该文本信息以文本注意力向量表征,最后由对象解码器的对象卷积部分对文本注意力向量以及场景注意力向量进行卷积计算,从而由卷积网络的输出层输出所述目标对象。

[0092] 在一种可行的实施方式中,为了使对象解码器预测的待绘制对象更精确,可以将已绘制历史对象纳入参考,以提高对象解码器预测结果的准确性。所述历史对象可以理解为当前时刻(即当前时间步)之前的时刻所绘制的对象,在实际应用中,通常历史对象为上一时刻(即上一时间步)所绘制的对象,假设当前时间步T所对应的对象为目标对象1,则历史对象为T-1时刻(即上一时间步)所绘制的对象。具体为:

[0093] 1、终端可以获取已绘制历史对象,具体为获取上一时间步对象编码器所预测的对象 $0_{t-1}$ ,通常该对象 $0_{t-1}$ 以高维度的对象特征向量进行编码表示。

[0094] 2、对所输入所述场景特征图进行池化处理,得到池化处理后的第一场景注意力向量 $u_t^0$ ,即

$$[0095] \quad u_t^0 = \text{AvgPooling}(\Psi^0(h_t^D))$$

[0096] 其中,AvgPooling()表示对象编码器的场景池化部分,通常为对象编码器中的池化层, $\Psi^0$ 为对象编码器的卷积网络, $h_t^D$ 为当前时间步的场景状态,通过卷积网络 $\Psi^0$ 可以关注 $h_t^D$ 的空间状态。

[0097] 终端将所述文本特征、所述场景特征图、历史对象输入至对象编码器,对象解码器的场景池化部分会对所述场景特征图进行池化处理,以收集对象预测所需的场景空间上下文,即  $h_t^D$  的空间状态,例如当前时间步中场景特征图所对应的画布是否添加目标对象以及添加的历史对象,然后通过池化处理以控制池化层的池化神经单元将参与的对象空间特征融合到场景注意力向量中,从而得到池化处理后的第一场景注意力向量  $u_t^0$ 。

[0098] 3.将所述第一场景注意力向量、所述历史对象以及所述文本特征输入至对象编码器的文本关注部分-第一文本关注器,输出第一文本注意力向量  $c_t^0$ ,即

$$[0099] \quad c_t^0 = \phi^1([u_t^0; O_{t-1}], [h_t^E; x_i])$$

[0100] 其中,  $\phi^1$  为对象编码器的文本关注部分-第一文本关注器,  $u_t^0$  为第一场景注意力向量,  $O_{t-1}$  为历史对象,  $[h_t^E, x_i]$  为文本特征。

[0101] 对象解码器的文本关注部分-第一文本关注器对场景注意力向量、历史对象以及文本特征进行处理,具体为通过注意力机制对输入的  $[u_t^0; O_{t-1}]$  来关注目标文本的语义上下文即  $[h_t^E, x_i]$ , 并突出所对应的文本特征所包含的对象部分,而对于特征图中非对象部分(颜色、情绪、动作、位置等属性信息)进行弱化,以确定待绘制对象在文本特征中所包含的文本信息,该文本信息以第一文本注意力向量表征。

[0102] 4.将所述第一场景注意力向量、所述历史对象以及所述第一文本注意力向量输入至对象卷积网络中,输出待绘制的目标对象。所述对象卷积网络通常为一定层数的卷积感知器,例如预定层数为2的卷积感知器;以第一场景注意力向量、所述历史对象以及所述第一文本注意力向量作为输入,根据第一场景注意力向量、所述历史对象以及所述第一文本注意力向量在卷积感知器的全连接层将各输入的特征综合起来至输出层,使用激励函数(sigmoid函数)来预测下一个对象的似然性,从而输出似然性高的目标对象,即所述目标对象。

[0103] 步骤205:将所述文本特征以及所述目标对象输入至第二文本关注器,输出第二文本注意力向量。

[0104] 其中,所述第二文本关注器、场景卷积网络、属性卷积网络共同构成了属性解码器,所述属性解码器是一种基于注意力的神经网络模型,对所述目标对象、所述文本特征以及所述场景画布引入注意力机制的属性解码器可以凸显出场景画布中更关键的图像部分,例如当前场景画布中优先级较高的目标对象的属性特征。在实际应用中,由第二文本关注器、场景卷积网络、属性卷积网络构成的属性解码器为一种属性确定模型。需要说明的是,所述属性解码器根据具体实施环境包括但不限于第二文本关注器、场景卷积网络、属性卷积网络。

[0105] 具体的,终端将所述文本特征以及所述目标对象输入至属性解码器的文本关注部分-第二文本关注器,输出第一文本注意力向量  $c_t^a$ ,即

$$[0106] \quad c_t^a = \phi^2(O_t, [h_t^E; x_i])$$

[0107] 其中,  $\phi^2$  为属性解码器的文本属性部分-第二文本关注器,  $O_t$  为目标对象,  $[h_t^E, x_i]$

为文本特征。

[0108] 属性解码器的文本关注部分-第二文本关注器对目标对象以及文本特征进行处理,具体为通过注意力机制对输入的 $O_t$ 来关注目标文本的语义上下文即 $[h_i^E, x_i]$ ,并突出所对应的文本特征所包含的属性部分(颜色、情绪、动作、位置等属性信息),而对于对象部分进行弱化,在文本关注器中通过大量样本数据预先训练所包含的计算矩阵,学习计算注意力分数以确定待绘制对象在文本特征中所包含的文本信息,该文本信息以第二文本注意力向量表征。

[0109] 步骤206:将所述场景特征图以及所述第二文本注意力向量输入至场景卷积网络中,输出第二场景注意力向量。

[0110] 具体的,所述场景卷积网络为一个基于图像(场景特征图)的注意力模块,通过该场景卷积网络可以收集待添加对象在场景画布中的有关场景信息,如待添加对象在场景画布中的添加位置等场景信息。通常该场景卷积网络可以是由两个卷积层组成的场景空间注意力模块。即

$$[0111] \quad u_t^a = \Psi^a(h_t^D; C_t^a)$$

[0112] 其中, $u_t^a$ 为第二场景注意力向量, $\Psi^a$ 为场景卷积网络, $h_t^D$ 为场景特征图, $C_t^a$ 为第二文本注意力向量。

[0113] 具体的,场景卷积网络,例如是包括输入层、卷积层、池化层、全连接层和输出层。在一些实施方式中,以场景特征图以及第二文本注意力向量作为输入,输入至场景卷积网络,例如是通过场景卷积网络的输入层接收场景特征图以及第二文本注意力向量,输入层可以将输入的数据(场景特征图以及第二文本注意力向量)进行标准化处理,有利于提升场景卷积网络的学习效率和表现。而后场景特征图经卷积层进行特征提取与计算,传递至池化层进行文本及场景特征选择和信息过滤,再经过全连接层将特征综合起来至输出层从而输出第二场景注意力向量。通过将场景特征图以及第二文本注意力向量输入至场景卷积神经网络,可以在场景特征图的基础上基于文本注意力向量累计更多的特征信息,从而获取特征图的场景内容表征,以得到场景特征图以及第二文本注意力向量共同对应的第二场景注意力向量。

[0114] 步骤207:将所述第二场景注意力向量、所述目标对象以及所述第二文本注意力向量输入至属性卷积网络中,输出所述目标对象的属性特征。

[0115] 所述属性卷积网络通常为一预定层数的卷积感知器,例如预定层数为4的卷积网络(CNN网络);以第二场景注意力向量、所述目标对象以及所述第二文本注意力向量作为输入,根据第二场景注意力向量、所述目标对象以及所述第二文本注意力向量在卷积感知器的全连接层将各输入的特征综合起来至输出层,使用激励函数(softmax函数)来预测下一个目标对象的至少一个属性,从而输出目标对象的属性特征。属性特征 $P()$ 表示如下:

$$[0116] \quad P(l_t, [R_t^k]) = \Theta(u_t^a, O_t, C_t^a)$$

[0117] 其中 $\Theta$ 为属性卷积网络, $\Theta$ 的输出层具有“ $1 + \sum_k R^k$ ”个输出通道,其中 $R_t^k$ 表示当前时间步第k个属性的离散范围, $l_t$ 为属性特征-位置属性,在实际应用中, $\Theta$ 的输出层的第一个通道预测使用softmax函数在空间域上对象最小化位置的可能性。其余通道预测每个关

注网格位置(由场景注意力向量确定)的属性。在训练过程中,来自真实位置的可能性用于计算损失。在确定属性的每个时间步中,从属性卷积网络中采样前一位置,然后从该采样位置收集目标对象对应的属性信息,直至对第二场景注意力向量所预测的所有采样位置(即每个关注网格位置)收集完目标对象对应的属性信息,以在输出层使用激励函数(softmax函数)输出目标对象的属性特征。

[0118] 步骤208:在预设的贴图索引库中索引所述目标对象对应的贴图对象。

[0119] 具体的,终端储存有贴图索引库,所述贴图索引库至少存储有多个贴图元素(对象所对应的贴图),终端可以采用图像检索技术在贴图索引库中获取目标贴图元素,将目标贴图元素作为所述目标对象所对应的贴图对象。

[0120] 其中,所述贴图索引库的贴图元素可以是已有的图像数据库(如WIDER FACE数据集、IJB-C测试集、AVA数据集、coco数据集等)中获取全部或部分样本图像作为贴图元素,和/或获取采用具有拍照功能的设备对实际环境下拍摄的样本图像。通过获取大量样本图像,对所述样本图像进行预处理,所述预处理包含数字化、几何变换、归一化、平滑、复原增强等处理过程,以得到处理之后的贴图元素。

[0121] 可选的,所述图像检索技术可以是一种图像检索算法,所述图像检索算法包括但不限于局部敏感哈希(Locality Sensitive Hashing,LSH)算法、SH谱哈希算法、SDH监督离散检索算法、局部聚合向量(vector of locally aggregated descriptors,VLAD)检索算法、K-D树检索算法等。

[0122] 在一种可行的实施方式中,在图像数据库所包含的各贴图元素中,在将获取的样本图像作为贴图元素存储时,可以为所述贴图元素分配一个图像标识,所述图像标识可以是图像id、编号、特定的字符串等,以图像标识来识别贴图元素。然后在检索时,可以直接基于目标对象,获取所述目标对象的图像标识,在图像数据库中查询所述图像标识对应的所述贴图元素。

[0123] 步骤209:将所述属性特征、所述贴图对象以及所述场景画布输入至画布调整模型中,输出所述目标文本对应的贴图。

[0124] 所述画布调整模型可以是基于神经网络模型中的一种,用于根据模型的输入(属性特征、贴图对象以及场景画布)在场景画布上绘制并调整贴图对象,并将调整之后的场景画布作为输出,输出目标文本对应的贴图。

[0125] 在本申请实施例中,所述画布调整模型以属性特征为基准,将当前确定的贴图对象绘制到场景画布上,其不直接生成高分辨率图像,而是将图像调整任务简化,画布调整模型分层从低到高分辨率根据属性特征一步步模拟图像分布,在场景画布上对贴图对象进行图像调整。需要说明的是,所述图像调整的对象还包括对所述目标对象相关联的图像部分进行调整,例如对所述目标对象相关联的图像背景(植物、动物等)进行调整。在一个具体的应用场景中,例如输入待处理文本为“一只胸前为白色的灰色小鸟,灰色小鸟的心情不好”,终端根据小鸟特征向量以在场景画布上绘制小鸟的贴图对象,并可以同步或异步根据小鸟对象所对应的属性特征,对场景画布上目标对象的特征贴图进行图像调整:将小鸟的外貌特征调整为皮毛为灰色、小鸟胸部区域为白色,将小鸟情绪特征调整为面部表征为心情低沉、与小鸟关联的图像部分-场景环境表征为阴雨天(如添加阴云、雨滴,将场景亮度、对比度对应调整)等。待根据属性特征对所述目标对象进行调整之后,即可得到调整之后的场景

画布,即所述目标文本对应的贴图。

[0126] 在实际应用中,目标文本在所述文本转贴图的过程中,通常会对应多个待绘制的对象,各对象的添加及绘制过程可以分别对应一个时间步,可以理解的是,在一个时间步中,场景画布可以只绘制一个对象,也可以同步或异步绘制多个对象,在本实施例中不作具体限定,待所有对象均添加完之后,画布调整模型即完成对场景画布的图像调整,此时输出目标文本对应的贴图。

[0127] 示意性的,例如所述目标文本可以是:

[0128] Tim is holding a hotdog.Amy is sitting in the sandbox.Amy is holding the shovel.

[0129] 终端在处理上述目标文本时,可以采用所述文本转贴图方法分步生成一个贴图,具体为:

[0130] 在时间步T1:对上述目标文本进行文本编码,得到目标文本对应的文本特征。提取文本特征对应的场景主题1,在预设的贴图索引库中索引场景主题1对应的场景贴图,如所述场景贴图如图5所示,图5是一种场景贴图对应的示意图,将如图5所示场景贴图确定为场景画布。

[0131] 在时间步T2:关注的待绘制的目标对象可以是“sandbox”,然后确定“sandbox”的属性(位置、尺寸、颜色、空间关系等),在场景画布上绘制目标对象-“sandbox”...,并基于所述属性特征对所述目标对象进行调整。调整之后的场景贴图如图6所示,可以注意到图6中绘制了一个“sandbox”。

[0132] 在时间步T3:关注的待绘制的目标对象可以是“Tim”,历史对象为“sandbox”,然后确定“Tim”的属性(位置、尺寸、颜色、空间关系等),如“holding”...,在场景画布上绘制目标对象-“Tim”,并基于所述属性特征对所述目标对象进行调整。调整之后的场景贴图如图7所示,可以注意到图7中绘制了一个人物“Tim”,图中“Tim”站在“sandbox”旁,动作为“左手托着”.....。

[0133] 在时间步T4:关注的待绘制的目标对象可以是“Amy”历史对象为“Tim”,然后确定“Jenny”的属性(位置、尺寸、颜色、空间关系等),如“holding”“sitting”,在场景画布上绘制目标对象-“Amy”,并基于所述属性特征对所述目标对象进行调整。调整之后的场景贴图如图8所示,可以注意到图8中绘制了一个人物“Amy”,图中“Amy”站在“sandbox”上,动作为“左手托着”.....。

[0134] 在时间步T5:关注的待绘制的目标对象可以是“hotdog”历史对象为“Tim”、,然后确定“hotdog”的属性(位置、尺寸、颜色、空间关系等),如“holding”“Timholding”,在场景画布上绘制目标对象-“hotdog”,并基于所述属性特征对所述目标对象进行调整。调整之后的场景贴图如图9所示,可以注意到图9中绘制了一物品“hotdog”,图中“hotdog”握持在“Tim”手上,.....。

[0135] 在时间步T6:关注的待绘制的目标对象可以是“shovel”历史对象为“Amy”...,然后确定“shovel”的属性(位置、尺寸、颜色、空间关系等),如“holding”“shovel holding”,在场景画布上绘制目标对象-“shovel”,并基于所述属性特征对所述目标对象进行调整。调整之后的场景贴图如图,10所示,可以注意到图10中绘制了一物品“shovel”,图中“shovel”握持在“Amy”手上,.....。



[0136] 此时,所有对象均添加完之后,画布调整模型即完成对场景画布1的图像调整,此时输出目标文本对应的贴图,贴图如图10所示。需要说明的上述目标文本仅仅是为了便于理解以更好的释义本申请实施例,释义中所涉及到的添加细节不作具体限定。

[0137] 在一种可行的实施方式中,所述画布调整模型可以是一种端到端的生成对抗模型,来模拟一系列多尺度图像分布。画布调整模型可由的许多个生成器和判别器。不同分辨率的贴图元素在模型的各个分支上绘制生成及修正。在每个分支上,生成器捕捉相应分辨率上的场景画布的图像分布,由判别器判别生成图像和相应尺寸的真实图像,以联合训练生成器,从而来逼近多层分布,从而有利于保证文本生成的贴图在语义上逼近目标文本的真实语义,是生成的贴图的准确性更高。

[0138] 在一种可行的实施方式中,所述画布调整模型可以是一种基于贴图检索的图像合成模型,具体为画布调整模型可以基于属性特征中的外观向量,在预设的贴图索引库中检索相匹配的至少一个贴图元素,从贴图元素获取补丁(即贴图元素的全部或部分图像),由各补丁对场景画布的目标对象以及与目标对象相关联的部分进行画布调整处理。具体为针对场景画布所对应的特征图中的每个位置预测外观向量,根据外观向量在贴图索引库中检索相似的补丁,同时在画布调整模型创建时,基于CNN网络训练一个补丁嵌入器,画布调整模型的补丁嵌入器可以将检索到的前景补丁(即检索到的补丁)在“场景画布绘制对象贴图的过程”中将前景补丁嵌入到场景画布中,具体通过将前景补丁经网络中的神经计算单元缩减为指定维数的矢量,如一维矢量 $F_t$ ,使用三重态嵌入方法即在模型处理过程中计算三重态损失,来缩小 $l_t$ 与 $F_t$ 之间的欧式距离,待根据属性特征对所述目标对象进行调整之后,即可得到调整之后的场景画布,即所述目标文本对应的贴图。

[0139] 在本申请实施例中,终端获取目标文本的文本特征,绘制所述文本特征对应的场景画布,基于所述文本特征以及所述场景画布,确定待绘制的目标对象,根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征,在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。通过目标文本的文本特征分步确定目标文本所对应的场景画布、待绘制的对象以及对象的属性特征,在场景画布绘制对象时基于对象的属性特征(位置、动作、姿态等)对图像进行相应调整,可以避免根据关键词导致生成贴图难以契合文本实际描述场景的问题,可以生成一个场景准确、对象清楚以及属性清晰的贴图,贴图可以更贴近文本实际所表达的语义,可以提升了贴图与文本实际描述场景的契合度,从而提高了生成贴图的准确度。

[0140] 下述为本申请装置实施例,可以用于执行本申请方法实施例。对于本申请装置实施例中未披露的细节,请参照本申请方法实施例。

[0141] 请参见图11,其示出了本申请一个示例性实施例提供的文本转贴图装置的结构示意图。该文本转贴图装置可以通过软件、硬件或者两者的结合实现成为装置的全部或一部分。该装置1包括场景画布绘制模块11、目标对象确定模块12、属性特征确定模块13和贴图生成模块14。

[0142] 场景画布绘制模块11,用于获取目标文本的文本特征,绘制所述文本特征对应的场景画布;

[0143] 目标对象确定模块12,用于基于所述文本特征以及所述场景画布,确定待绘制的目标对象;

- [0144] 属性特征确定模块13,用于根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征;
- [0145] 贴图生成模块14,用于在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。
- [0146] 可选的,如图12所示,所述场景画布绘制模块11,包括:
- [0147] 文本特征输出单元111,用于将目标文本输入至文本编码器中,输出所述目标文本对应的文本特征;
- [0148] 场景画布确定单元112,用于提取所述文本特征对应的场景主题,在预设的贴图索引库中索引所述场景主题对应的场景贴图,将所述场景贴图确定为场景画布。
- [0149] 可选的,如图16所示,所述装置1,还包括:
- [0150] 场景特征图编码模块15,用于将所述场景画布输入至卷积网络中进行场景编码,输出场景编码后的场景特征图。
- [0151] 可选的,所述目标对象确定模块12,具体用于:
- [0152] 将所述文本特征以及所述场景特征图输入至对象解码器中,输出所述目标对象。
- [0153] 可选的,如图16所示,所述装置1,还包括:
- [0154] 历史对象获取模块16,用于获取已绘制的历史对象;
- [0155] 如图13所示,所述目标对象确定模块12,包括:
- [0156] 向量池化单元121,用于对所述场景特征图进行池化处理,得到池化处理后的第一场景注意力向量;
- [0157] 文本向量输出单元122,用于将所述第一场景注意力向量、所述历史对象以及所述文本特征输入至第一文本关注器,输出第一文本注意力向量;
- [0158] 目标对象输出单元123,用于将所述第一场景注意力向量、所述历史对象以及所述第一文本注意力向量输入至对象卷积网络中,输出所述目标对象。
- [0159] 可选的,如图14所示,所述属性特征确定模块13,包括:
- [0160] 文本向量输出单元131,用于将所述文本特征以及所述目标对象输入至第二文本关注器,输出第二文本注意力向量;
- [0161] 场景向量输出单元132,用于将所述场景特征图以及所述第二文本注意力向量输入至场景卷积网络中,输出第二场景注意力向量;
- [0162] 属性特征输出单元133,用于将所述第二场景注意力向量、所述目标对象以及所述第二文本注意力向量输入至属性卷积网络中,输出所述目标对象的属性特征。
- [0163] 可选的,如图15所示,所述贴图生成模块14,包括:
- [0164] 贴图对象绘制单元141,用于在预设的贴图索引库中索引所述目标对象对应的贴图对象;
- [0165] 场景画布调整单元142,用于将所述属性特征、所述贴图对象以及所述场景画布输入至画布调整模型中,输出所述目标文本对应的贴图。
- [0166] 需要说明的是,上述实施例提供的文本转贴图装置在执行文本转贴图方法时,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将设备的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。另外,上述实施例提供的文本转贴图装置与文本转贴图方法实施例属于

同一构思,其体现实现过程详见方法实施例,这里不再赘述。

[0167] 上述本申请实施例序号仅仅为了描述,不代表实施例的优劣。

[0168] 在本申请实施例中,终端获取目标文本的文本特征,绘制所述文本特征对应的场景画布,基于所述文本特征以及所述场景画布,确定待绘制的目标对象,根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征,在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。通过目标文本的文本特征分步确定目标文本所对应的场景画布、待绘制的对象以及对象的属性特征,在场景画布绘制对象时基于对象的属性特征(位置、动作、姿态等)对图像进行相应调整,可以避免根据关键词导致生成贴图难以契合文本实际描述场景的问题,可以生成一个场景准确、对象清楚以及属性清晰的贴图,贴图可以更贴近文本实际所表达的语义,可以提升了贴图与文本实际描述场景的契合度,从而提高了生成贴图的准确度。

[0169] 本申请实施例还提供了一种计算机存储介质,所述计算机存储介质可以存储有多条指令,所述指令适于由处理器加载并执行如上述图1-图10所示实施例的所述文本转贴图方法,具体执行过程可以参见图1-图10所示实施例的具体说明,在此不进行赘述。

[0170] 本申请还提供了一种计算机程序产品,该计算机程序产品存储有至少一条指令,所述至少一条指令由所述处理器加载并执行如上述图1-图10所示实施例的所述文本转贴图方法,具体执行过程可以参见图1-图10所示实施例的具体说明,在此不进行赘述。

[0171] 请参见图17,为本申请实施例提供了一种电子设备的结构示意图。如图17所示,所述电子设备1000可以包括:至少一个处理器1001,至少一个网络接口1004,用户接口1003,存储器1005,至少一个通信总线1002。

[0172] 其中,通信总线1002用于实现这些组件之间的连接通信。

[0173] 其中,用户接口1003可以包括显示屏(Display)、摄像头(Camera),可选用户接口1003还可以包括标准的有线接口、无线接口。

[0174] 其中,网络接口1004可选的可以包括标准的有线接口、无线接口(如WI-FI接口)。

[0175] 其中,处理器1001可以包括一个或者多个处理核心。处理器1001利用各种借口和线路连接整个服务器1000内的各个部分,通过运行或执行存储在存储器1005内的指令、程序、代码集或指令集,以及调用存储在存储器1005内的数据,执行服务器1000的各种功能和处理数据。可选的,处理器1001可以采用数字信号处理(Digital Signal Processing, DSP)、现场可编程门阵列(Field-Programmable Gate Array, FPGA)、可编程逻辑阵列(Programmable Logic Array, PLA)中的至少一种硬件形式来实现。处理器1001可集成中央处理器(Central Processing Unit, CPU)、图像处理(Graphics Processing Unit, GPU)和调制解调器等中的一种或几种的组合。其中,CPU主要处理操作系统、用户界面和应用程序等;GPU用于负责显示屏所需要显示的内容的渲染和绘制;调制解调器用于处理无线通信。可以理解的是,上述调制解调器也可以不集成到处理器1001中,单独通过一块芯片进行实现。

[0176] 其中,存储器1005可以包括随机存储器(Random Access Memory, RAM),也可以包括只读存储器(Read-Only Memory)。可选的,该存储器1005包括非瞬时性计算机可读介质(non-transitory computer-readable storage medium)。存储器1005可用于存储指令、程序、代码、代码集或指令集。存储器1005可包括存储程序区和存储数据区,其中,存储程序区

可存储用于实现操作系统的指令、用于至少一个功能的指令(比如触控功能、声音播放功能、图像播放功能等)、用于实现上述各个方法实施例的指令等;存储数据区可存储上面各个方法实施例中涉及到的数据等。存储器1005可选的还可以是至少一个位于远离前述处理器1001的存储装置。如图17所示,作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及文本转贴图应用程序。

[0177] 在图17所示的电子设备1000中,用户接口1003主要用于为用户提供输入的接口,获取用户输入的数据;而处理器1001可以用于调用存储器1005中存储的文本转贴图应用程序,并具体执行以下操作:

[0178] 获取目标文本的文本特征,绘制所述文本特征对应的场景画布;

[0179] 基于所述文本特征以及所述场景画布,确定待绘制的目标对象;

[0180] 根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征;

[0181] 在所述场景画布上绘制所述目标对象,并基于所述属性特征对所述目标对象进行调整,生成所述目标文本对应的贴图。

[0182] 在一个实施例中,所述处理器1001在执行所述获取目标文本的文本特征,绘制所述文本特征对应的场景画布时,具体执行以下操作:

[0183] 将目标文本输入至文本编码器中,输出所述目标文本对应的文本特征;

[0184] 提取所述文本特征对应的场景主题,在预设的贴图索引库中索引所述场景主题对应的场景贴图,将所述场景贴图确定为场景画布。

[0185] 在一个实施例中,所述处理器1001在执行所述获取目标文本的文本特征,绘制所述文本特征对应的场景画布之后,还执行以下操作:

[0186] 将所述场景画布输入至卷积网络中进行场景编码,输出场景编码后的场景特征图。

[0187] 在一个实施例中,所述处理器1001在执行所述基于所述文本特征以及所述场景画布,确定待绘制的目标对象时,具体执行以下操作:

[0188] 将所述文本特征以及所述场景特征图输入至对象解码器中,输出所述目标对象。

[0189] 在一个实施例中,所述处理器1001在执行所述文本转贴图的方法时,还执行以下操作:

[0190] 获取已绘制过的历史对象;

[0191] 所述将所述文本特征以及所述场景特征图输入至对象解码器中,输出所述目标对象,包括:

[0192] 对所述场景特征图进行池化处理,得到池化处理后的第一场景注意力向量;

[0193] 将所述第一场景注意力向量、所述历史对象以及所述文本特征输入至第一文本关注器,输出第一文本注意力向量;

[0194] 将所述第一场景注意力向量、所述历史对象以及所述第一文本注意力向量输入至对象卷积网络中,输出所述目标对象。

[0195] 在一个实施例中,所述处理器1001在执行所述根据所述目标对象、所述文本特征以及所述场景画布,确定所述目标对象的属性特征时,具体执行以下操作:

[0196] 将所述文本特征以及所述目标对象输入至第二文本关注器,输出第二文本注意力

向量；

[0197] 将所述场景特征图以及所述第二文本注意力向量输入至场景卷积网络中，输出第二场景注意力向量；

[0198] 将所述第二场景注意力向量、所述目标对象以及所述第二文本注意力向量输入至属性卷积网络中，输出所述目标对象的属性特征。

[0199] 在一个实施例中，所述处理器1001在执行所述场景画布上绘制所述目标对象，并基于所述属性特征对所述目标对象进行调整，生成所述目标文本对应的贴图时，具体执行以下操作：

[0200] 在预设的贴图索引库中索引所述目标对象对应的贴图对象；

[0201] 将所述属性特征、所述贴图对象以及所述场景画布输入至画布调整模型中，输出所述目标文本对应的贴图。

[0202] 在本申请实施例中，终端获取目标文本的文本特征，绘制所述文本特征对应的场景画布，基于所述文本特征以及所述场景画布，确定待绘制的目标对象，根据所述目标对象、所述文本特征以及所述场景画布，确定所述目标对象的属性特征，在所述场景画布上绘制所述目标对象，并基于所述属性特征对所述目标对象进行调整，生成所述目标文本对应的贴图。通过目标文本的文本特征分步确定目标文本所对应的场景画布、待绘制的对象以及对象的属性特征，在场景画布绘制对象时基于对象的属性特征(位置、动作、姿态等)对图像进行相应调整，可以避免根据关键词导致生成贴图难以契合文本实际描述场景的问题，可以生成一个场景准确、对象清楚以及属性清晰的贴图，贴图可以更贴近文本实际所表达的语义，可以提升了贴图与文本实际描述场景的契合度，从而提高了生成贴图的准确度。

[0203] 本领域的技术人员可以清楚地了解到本申请的技术方案可借助软件和/或硬件来实现。本说明书中的“单元”和“模块”是指能够独立完成或与其他部件配合完成特定功能的软件和/或硬件，其中硬件例如可以是现场可编程门阵列(Field-Programmable Gate Array, FPGA)、集成电路(Integrated Circuit, IC)等。

[0204] 需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请并不受所描述的动作顺序的限制，因为依据本申请，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本申请所必须的。

[0205] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中沒有详述的部分，可以参见其他实施例的相关描述。

[0206] 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，可通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些服务接口，装置或单元的间接耦合或通信连接，可以是电性或其它的形式。

[0207] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个

网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0208] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0209] 所述集成的单元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器中。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全部或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器中,包括若干指令用以使得一台计算机设备(可为个人计算机、服务器或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储器包括:U盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0210] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通进程序来指令相关的硬件来完成,该程序可以存储于一计算机可读取存储器中,存储器可以包括:闪存盘、只读存储器(Read-Only Memory,ROM)、随机存取器(Random Access Memory,RAM)、磁盘或光盘等。

[0211] 以上所述者,仅为本公开的示例性实施例,不能以此限定本公开的范围。即但凡依本公开教导所作的等效变化与修饰,皆仍属本公开涵盖的范围内。本领域技术人员在考虑说明书及实践这里的公开后,将容易想到本公开的其它实施方案。本申请旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未记载的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的范围和精神由权利要求限定。

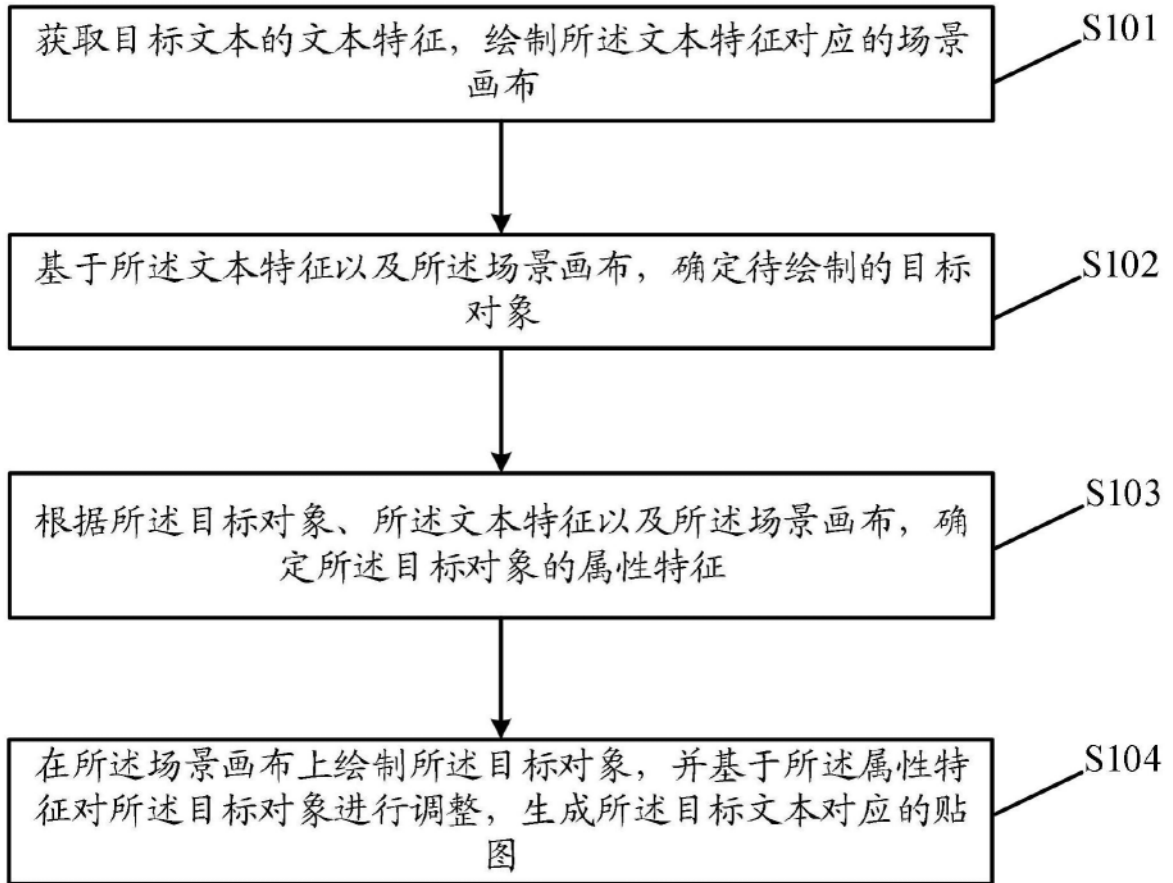


图1

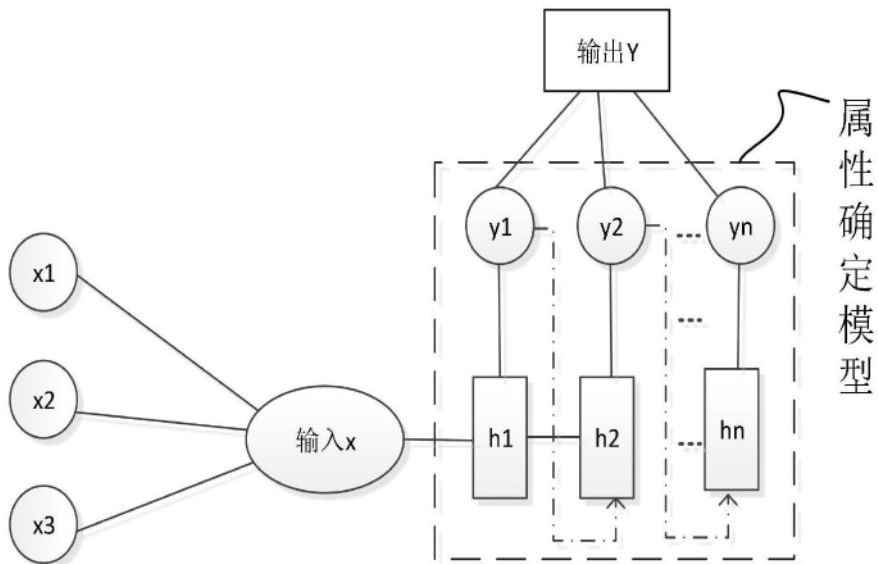


图2

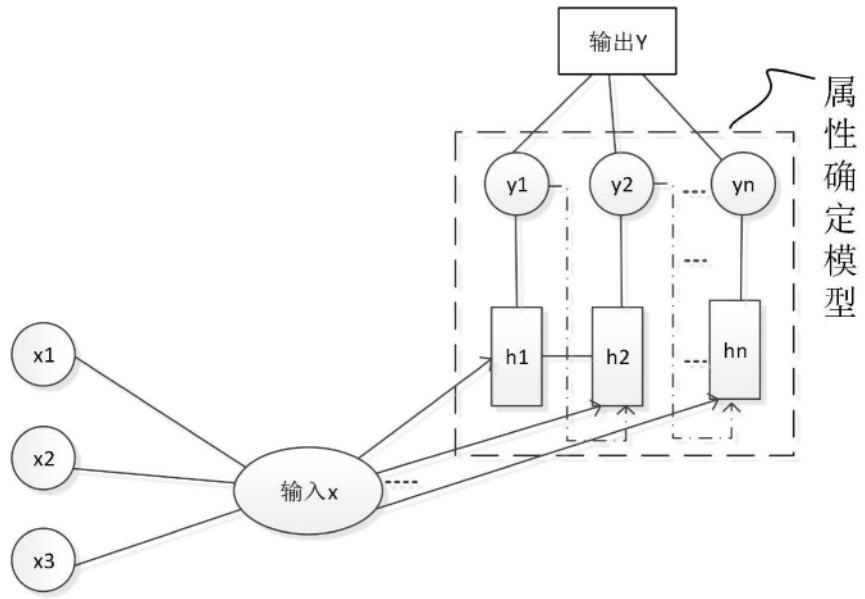


图3



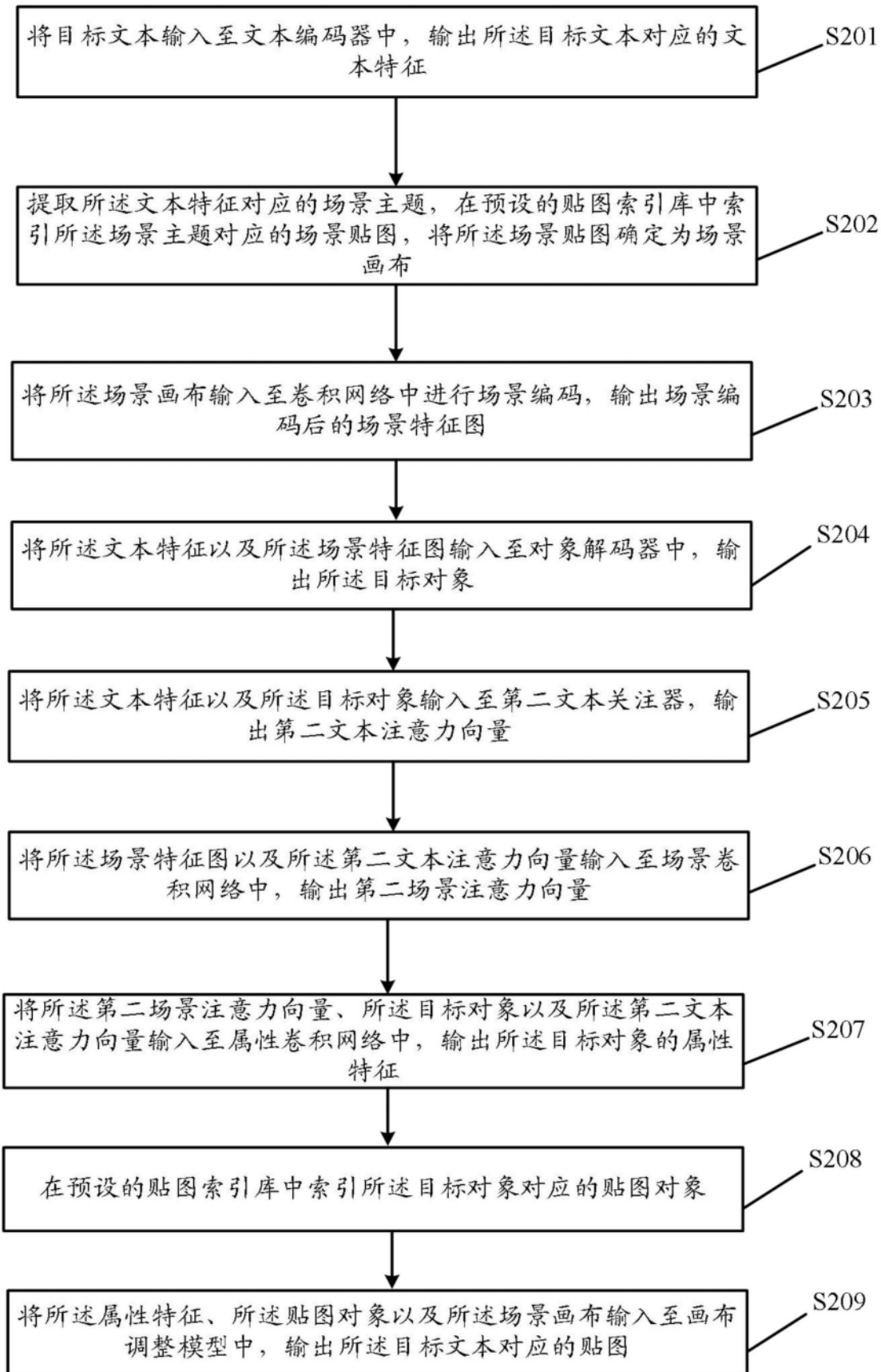


图4

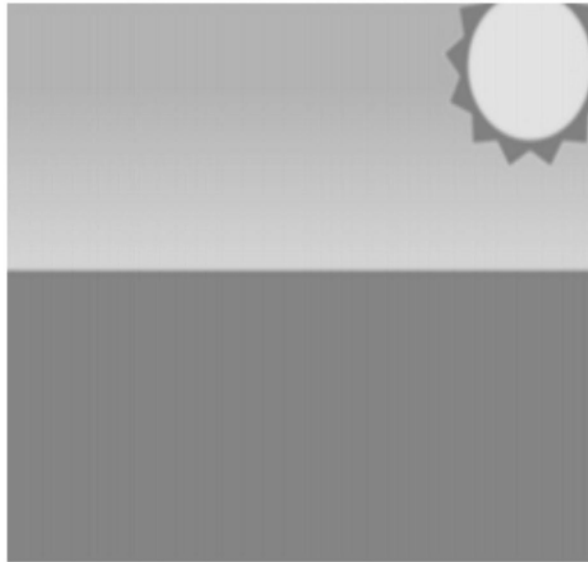


图5

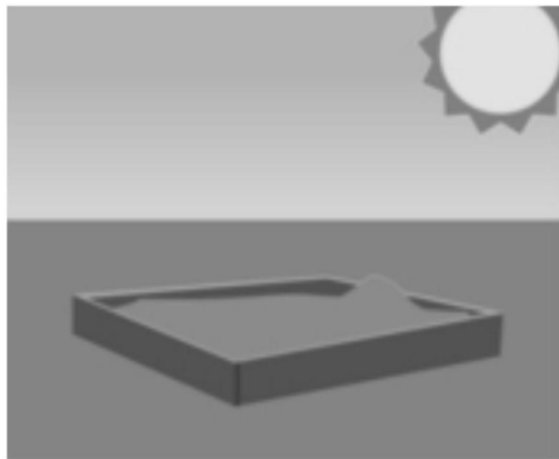


图6

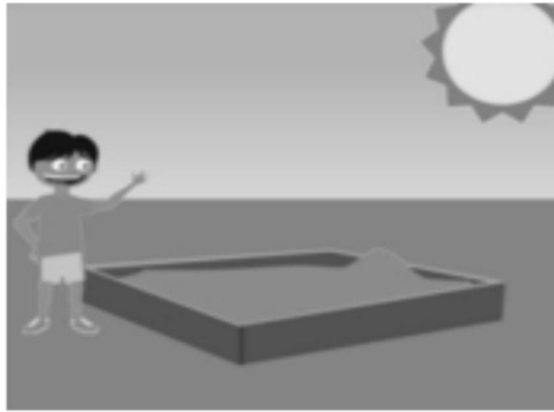


图7

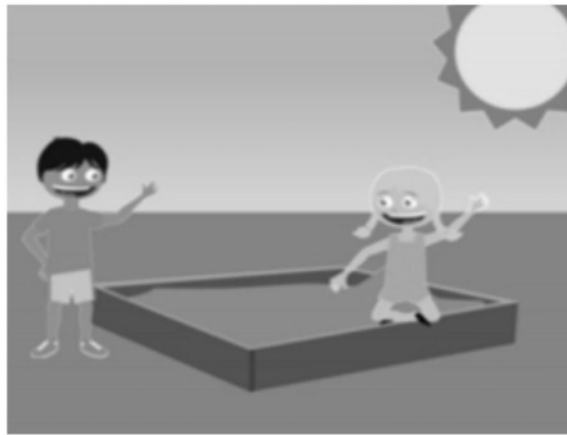


图8



图9

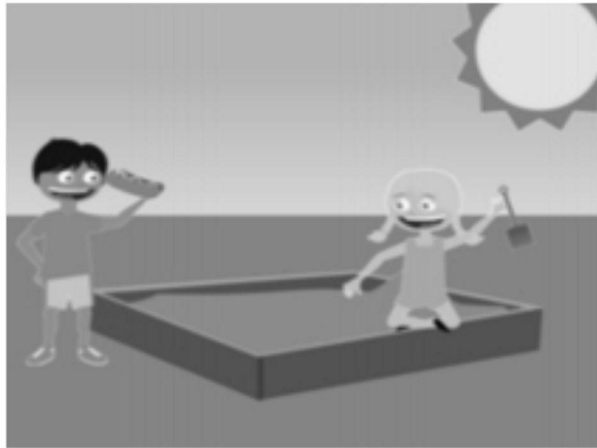


图10

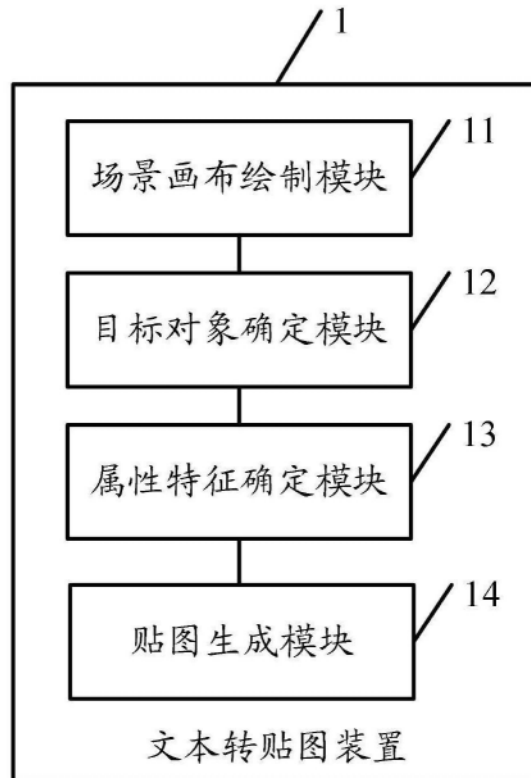


图11

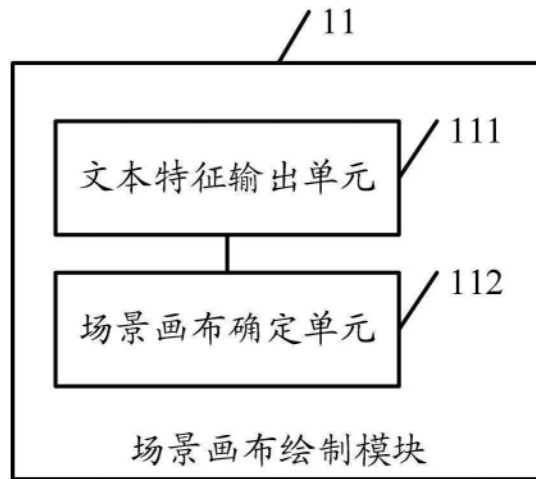


图12

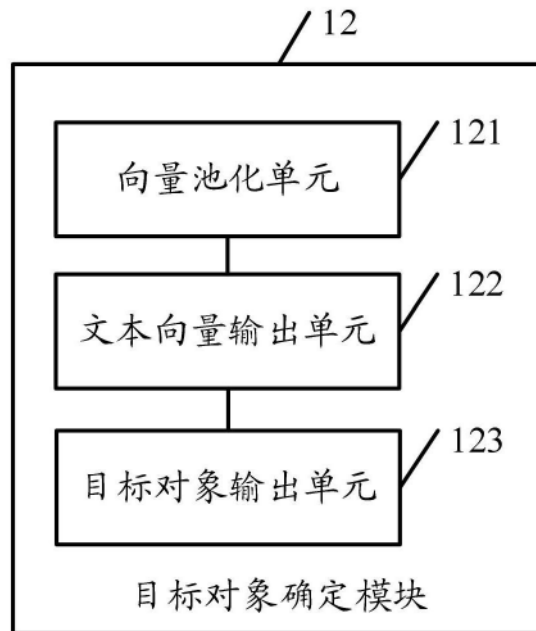


图13

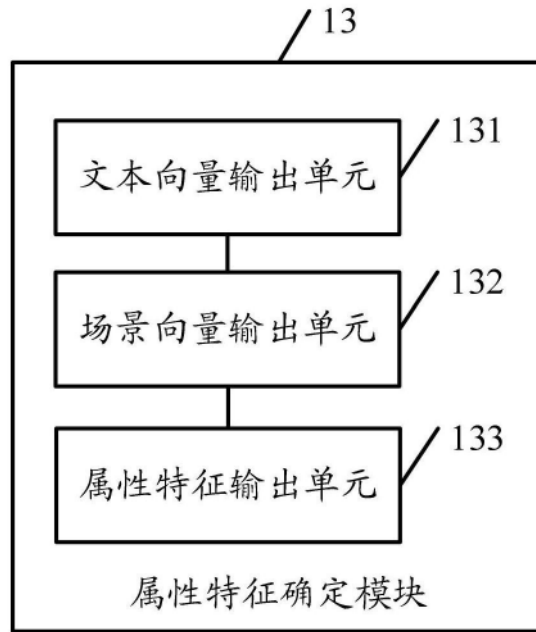


图14

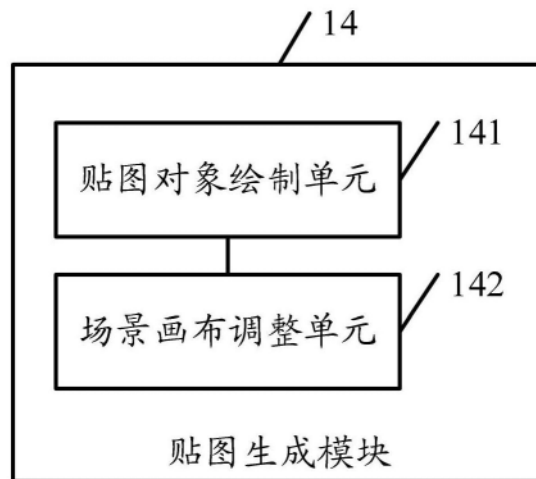


图15

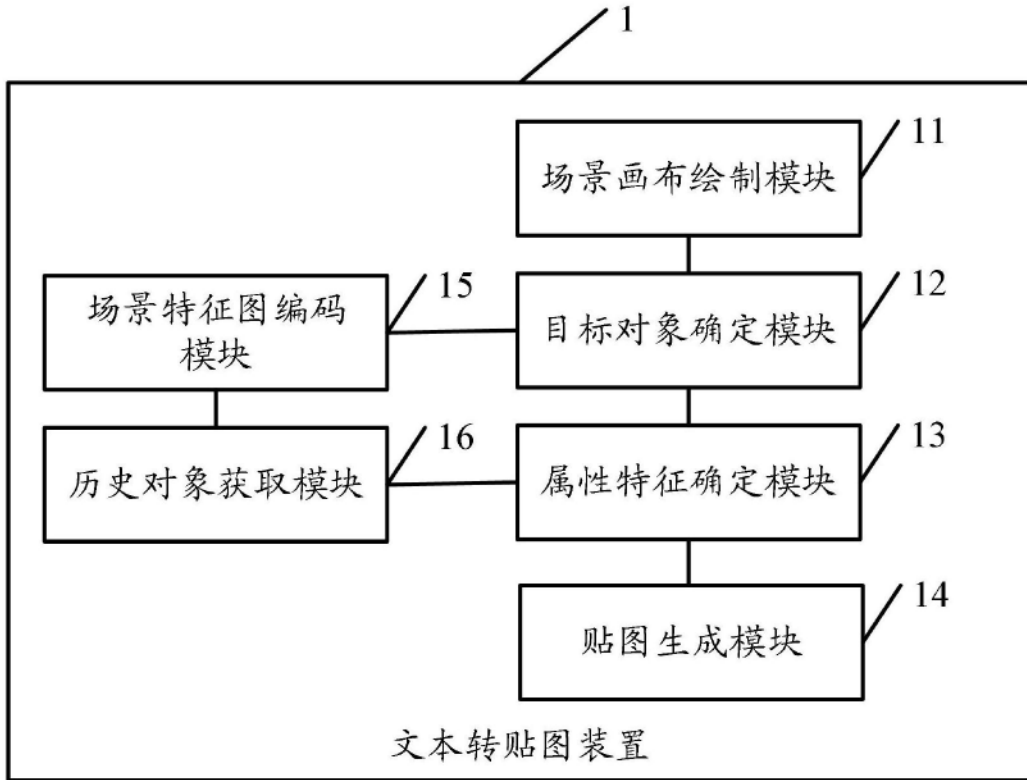


图16

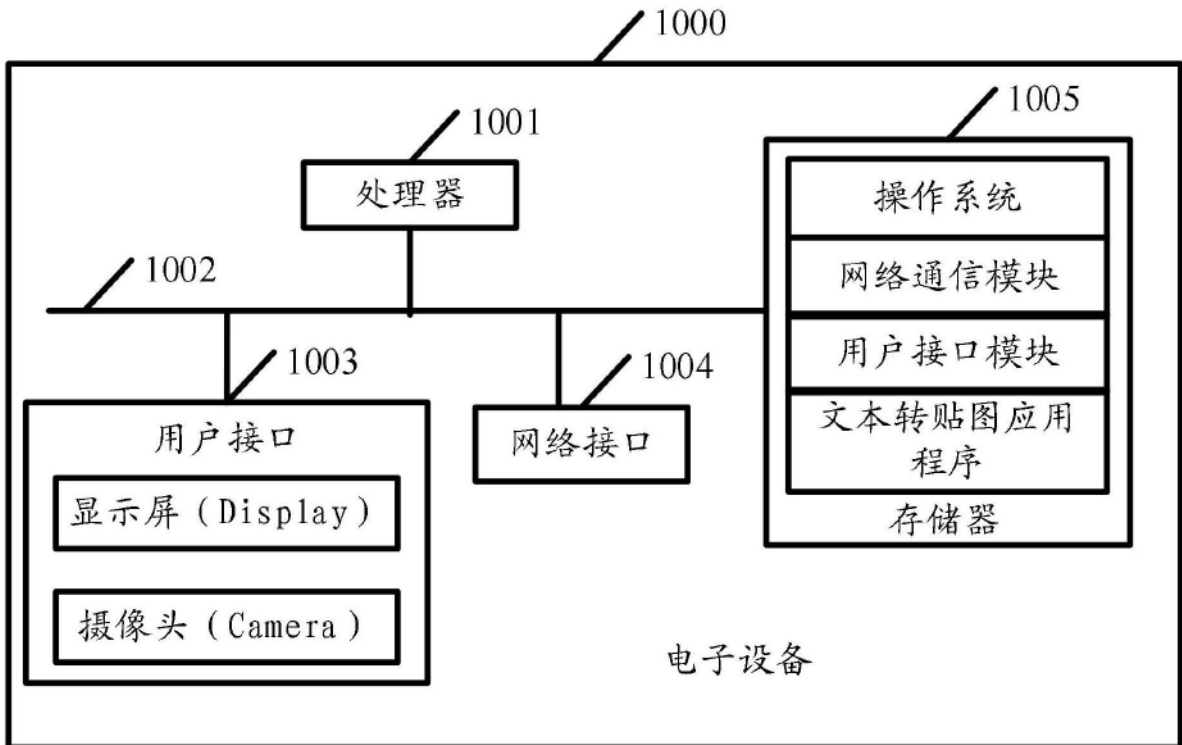


图17