(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 119128200 B
(45)授权公告日 2025.03.14

(21)申请号 202411605346.4
(22)申请日 2024.11.12
(65)同一申请的已公布的文献号
申请公布号 CN 119128200 A
(43)申请公布日 2024.12.13
(73)专利权人 杭州喔影网络科技有限公司
地址 311100 浙江省杭州市余杭区仓前街道龙园路88号3幢A座801室
(72)发明人 袁峰 邓豪 王豪 李浬
(74)专利代理机构 杭州华进联浙知识产权代理有限公司 33250
专利代理师 盛影影
(51)Int.Cl.
G06F 16/583(2019.01)
G06F 16/64(2019.01)
G06F 16/334(2025.01)
G06F 16/34(2025.01)
G06F 40/30(2020.01)
(56)对比文件
CN 113158630 A,2021.07.23
CN 113448477 A,2021.09.28
审查员 赵会玲

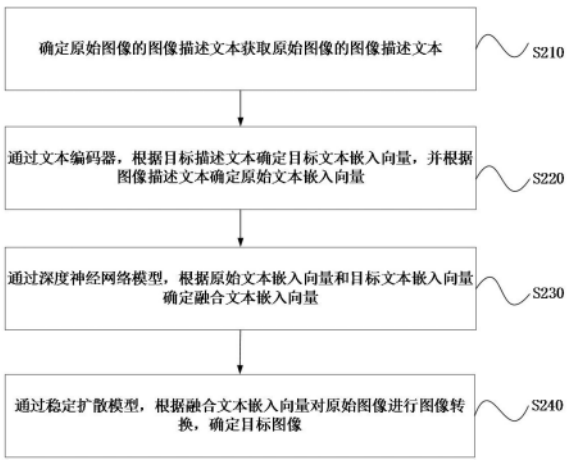
权利要求书2页 说明书16页 附图5页

(54)发明名称

图像转换方法、系统、计算机设备以及存储介质

(57)摘要

本申请涉及一种图像转换方法、系统、计算机设备以及存储介质。包括:获取原始图像的图像描述文本,并通过文本编码器,根据目标描述文本和图像描述文本确定原始文本嵌入向量和目标文本嵌入向量;通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。上述方案,能够在原始图像进行图像转换的过程中,尽量保持原始图像的特征结构,提高了图像转换结果的准确性,使得获取的目标图像具有期望的视觉观感。



CN 119128200 B

1.一种图像转换方法,其特征在于,包括:

获取原始图像的图像描述文本;

通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据所述图像描述文本确定原始文本嵌入向量;

通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量;通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量,包括:将目标文本嵌入向量输入深度神经网络模型,通过所述深度神经网络模型的多头自注意力层对目标文本嵌入向量进行特征提取,生成目标文本的语义特征;将原始文本嵌入向量和所述目标文本的语义特征输入所述深度神经网络模型的交叉注意力层,生成所述图像描述文本的融合语义信息,将所述目标文本嵌入向量和所述目标文本的语义特征输入所述深度神经网络模型的交叉注意力层,生成所述目标文本的融合语义信息;通过所述深度学习网络的前馈网络层,根据所述目标文本的融合语义信息和所述图像描述文本的融合语义信息确定所述原始文本嵌入向量和所述目标文本嵌入向量的融合文本嵌入向量;

通过稳定扩散模型,根据所述融合文本嵌入向量对所述原始图像进行图像转换,确定目标图像。

2.根据权利要求1所述的方法,其特征在于,获取原始图像的图像描述文本,包括:

将原始图像输入图像描述生成器,确定所述原始图像的子图像;

通过神经网络模型的线性嵌入向量层将所述子图像转换为图像嵌入向量;

通过图像描述生成器中的视觉编码器,根据所述图像嵌入向量确定所述原始图像的图像特征;

通过所述图像描述生成器中的视觉解码器,根据所述图像特征确定所述原始图像的图像描述文本。

3.根据权利要求1所述的方法,其特征在于,通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据所述图像描述文本确定原始文本嵌入向量,包括:

确定所述图像描述文本的原始词元序列,以及所述目标描述文本的目标词元序列;

根据所述原始词元序列确定原始输入嵌入向量,并根据所述目标词元序列确定目标输入嵌入向量;

根据所述原始输入嵌入向量确定原始嵌入向量矩阵,并根据所述目标输入嵌入向量确定目标嵌入向量矩阵;

通过文本编码器,根据所述原始嵌入向量矩阵确定原始文本嵌入向量,并根据所述目标嵌入向量矩阵确定目标文本嵌入向量。

4.根据权利要求3所述的方法,其特征在于,根据所述原始词元序列确定原始输入嵌入向量,并根据所述目标词元序列确定目标输入嵌入向量,包括:

确定所述原始词元序列对应的原始词嵌入向量和原始位置嵌入向量,以及所述目标次元序列对应的目标词嵌入向量和目标位置嵌入向量;

根据所述原始词嵌入向量和原始位置嵌入向量确定原始词元序列的原始输入嵌入向量;

根据所述目标词嵌入向量和所述目标位置嵌入向量确定目标词元序列的目标输入嵌

入向量。

5.根据权利要求1所述的方法,其特征在于,通过稳定扩散模型,根据所述融合文本嵌入向量对所述原始图像进行图像转换,确定目标图像,包括:

通过多模态预训练模型和图像分类模型,根据所述图像描述文本对所述原始图像进行语义分割,确定分割图像;

根据所述分割图像的像素信息确定所述分割图像的掩码图;

将掩码图输入到稳定扩散模型,基于所述融合文本嵌入向量生成特征空间向量,通过去噪扩散概率模型对特征空间向量进行降噪处理,确定目标空间向量;

根据所述目标空间向量对所述原始图像进行图像转换,确定目标图像。

6.根据权利要求5所述的方法,其特征在于,根据所述分割图像的像素信息确定所述分割图像的掩码图,包括:

根据所述分割图像的像素信息确定所述分割图像的图像置信度,并根据所述图像置信度确定置信度矩阵;

对置信度矩阵中的矩阵元素进行反处理,确定目标矩阵;

根据所述目标矩阵和分割图像确定掩码图。

7.一种图像转换系统,其特征在于,所述图像转换系统包括:

原始图像输入界面,用于输入原始图像;

描述文本输入界面,用于输入目标描述文本;

目标图像展示界面,用于展示目标图像;所述目标图像的生成方式为:通过文本编码器,根据原始图像的目标描述文本和所述图像描述文本确定原始文本嵌入向量和目标文本嵌入向量;通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量;通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量,包括:将目标文本嵌入向量输入深度神经网络模型,通过所述深度神经网络模型的多头自注意力层对目标文本嵌入向量进行特征提取,生成目标文本的语义特征;将原始文本嵌入向量和所述目标文本的语义特征输入所述深度神经网络模型的交叉注意力层,生成所述图像描述文本的融合语义信息,将所述目标文本嵌入向量和所述目标文本的语义特征输入所述深度神经网络模型的交叉注意力层,生成所述目标文本的融合语义信息;通过所述深度学习网络的前馈网络层,根据所述目标文本的融合语义信息和所述图像描述文本的融合语义信息确定所述原始文本嵌入向量和所述目标文本嵌入向量的融合文本嵌入向量;通过稳定扩散模型,根据所述融合文本嵌入向量对所述原始图像进行图像转换,确定目标图像。

8.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至权利要求6中任一项所述的方法的步骤。

9.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至权利要求6中任一项所述的方法的步骤。

图像转换方法、系统、计算机设备以及存储介质

技术领域

[0001] 本申请涉及图像处理技术领域，特别是涉及一种图像转换方法、系统、计算机设备以及存储介质。

背景技术

[0002] 近年来，深度学习和人工智能技术的迅猛发展已经对图像翻译领域产生了影响。图像翻译作为计算机视觉领域的一个重要分支，旨在将一张图像的内容转化为另一张具有不同风格、内容或领域的图像，同时尽可能地保留源图像的内容和结构。图像翻译技术的社会应用越来越广泛，其重要性日益显著。

[0003] 域迁移图像翻译是图像翻译的一个重要子任务，旨在将一个图像从源域转换到目标域，同时保持原始图像的内容结构信息不变。当前在将原始图像进行域迁移时，往往通过深度神经网络，根据原始图像的图像特征和目标图像的描述信息将原始图像从源域转换到目标域，存在生成的目标图像与原始图像视觉观感不一致的缺陷，因此，如何在对原始图像进行域迁移时，保证图像转换结果的准确性，使得获取的目标图像具有期望的视觉观感，是需要解决的问题。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种能够在对原始图像进行域迁移时，保证图像转换结果的准确性，使得获取的目标图像具有期望的视觉观感，是需要解决的问题的面部表情的图像转换方法、系统、计算机设备以及存储介质。

[0005] 第一方面，本申请提供了一种图像转换方法，所述方法包括：

[0006] 获取原始图像的图像描述文本；

[0007] 通过文本编码器，根据目标描述文本确定目标文本嵌入向量，并根据所述图像描述文本确定原始文本嵌入向量；

[0008] 通过深度神经网络模型，根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量；

[0009] 通过稳定扩散模型，根据所述融合文本嵌入向量对所述原始图像进行图像转换，确定目标图像。

[0010] 在其中一个实施例中，获取原始图像的图像描述文本，包括：

[0011] 将原始图像输入图像描述生成器，确定所述原始图像的子图像；

[0012] 通过神经网络模型的线性嵌入向量层将所述子图像转换为图像嵌入向量；

[0013] 通过图像描述生成器中的视觉编码器，根据所述图像嵌入向量确定所述原始图像的图像特征；

[0014] 通过所述图像描述生成器中的视觉解码器，根据所述图像特征确定所述原始图像的图像描述文本。

[0015] 在其中一个实施例中，通过文本编码器，根据目标描述文本确定目标文本嵌入向

量,并根据所述图像描述文本确定原始文本嵌入向量,包括:

[0016] 确定所述图像描述文本的原始词元序列,以及所述目标描述文本的目标词元序列;

[0017] 根据所述原始词元序列确定原始输入嵌入向量,并根据所述目标词元序列确定目标输入嵌入向量;

[0018] 根据所述原始输入嵌入向量确定原始嵌入向量矩阵,并根据所述目标输入嵌入向量确定目标嵌入向量矩阵;

[0019] 通过文本编码器,根据所述原始嵌入向量矩阵确定原始文本嵌入向量,并根据所述目标嵌入向量矩阵确定目标文本嵌入向量。

[0020] 在其中一个实施例中,根据所述原始词元序列确定原始输入嵌入向量,并根据所述目标词元序列确定目标输入嵌入向量,包括:

[0021] 确定所述原始词元序列对应的原始词嵌入向量和原始位置嵌入向量,以及所述目标次元序列对应的目标词嵌入向量和目标位置嵌入向量;

[0022] 根据所述原始词嵌入向量和原始位置嵌入向量确定原始词元序列的原始输入嵌入向量;

[0023] 根据所述目标词嵌入向量和所述目标位置嵌入向量确定目标词元序列的目标输入嵌入向量。

[0024] 在其中一个实施例中,通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量,包括:

[0025] 将目标文本嵌入向量输入深度神经网络模型,通过所述深度神经网络模型的多头自注意力层对目标文本嵌入向量进行特征提取,生成目标文本的语义特征;

[0026] 将原始文本嵌入向量和目标文本嵌入向量输入所述深度神经网络模型的交叉注意力层,生成所述目标文本的融合语义信息,以及所述图像描述文本的融合语义信息;

[0027] 通过所述深度学习网络的前馈网络层,根据所述目标文本的融合语义信息和所述图像描述文本的融合语义信息确定所述原始文本嵌入向量和所述目标文本嵌入向量的融合文本嵌入向量。

[0028] 在其中一个实施例中,通过稳定扩散模型,根据所述融合文本嵌入向量对所述原始图像进行图像转换,确定目标图像,包括:

[0029] 通过多模态预训练模型和图像分类模型,根据所述图像描述文本对所述原始图像进行语义分割,确定分割图像;

[0030] 根据所述分割图像的像素信息确定所述分割图像的掩码图;

[0031] 将掩码图输入到稳定扩散模型,基于所述融合文本嵌入向量生成特征空间向量,通过去噪扩散概率模型对特征空间向量进行降噪处理,确定目标空间向量;

[0032] 根据所述目标空间向量对所述原始图像进行图像转换,确定目标图像。

[0033] 在其中一个实施例中,根据所述分割图像的像素信息确定所述分割图像的掩码图,包括:

[0034] 根据所述分割图像的像素信息确定所述分割图像的图像置信度,并根据所述图像置信度确定置信度矩阵;

[0035] 对置信度矩阵中的矩阵元素进行反处理,确定目标矩阵;

[0036] 根据所述目标矩阵和分割图像确定掩码图。

[0037] 第二方面,本申请还提供了一种图像转换系统,所述图像转换系统包括:

[0038] 原始图像输入界面,用于输入原始图像;

[0039] 描述文本输入界面,用于输入目标描述文本;

[0040] 目标图像展示界面,用于展示目标图像;所述目标图像的生成方式为:通过文本编码器,根据原始图像的目标描述文本和所述图像描述文本确定原始文本嵌入向量和目标文本嵌入向量;通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量;通过稳定扩散模型,根据所述融合文本嵌入向量对所述原始图像进行图像转换,确定目标图像。

[0041] 第三方面,本申请还提供了一种计算机设备,所述计算机设备包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现以下步骤:

[0042] 获取原始图像的图像描述文本;

[0043] 通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据所述图像描述文本确定原始文本嵌入向量;

[0044] 通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量;

[0045] 通过稳定扩散模型,根据所述融合文本嵌入向量对所述原始图像进行图像转换,确定目标图像。

[0046] 第四方面,本申请还提供了一种计算机可读存储介质,所述计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现以下步骤:

[0047] 获取原始图像的图像描述文本;

[0048] 通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据所述图像描述文本确定原始文本嵌入向量;

[0049] 通过深度神经网络模型,根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量;

[0050] 通过稳定扩散模型,根据所述融合文本嵌入向量对所述原始图像进行图像转换,确定目标图像。

[0051] 上述图像转换方法、系统、计算机设备以及存储介质,获取原始图像的图像描述文本,通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据图像描述文本确定原始文本嵌入向量;通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。解决了直接由文本条件主导图像过程而忽略了原始图像的内容和结构语义,存在生成的目标图像与原始图像视觉观感不一致的问题。上述方案,在对原始图像进行图像转换时,基于原始图像的图像描述文本,和期望生成的目标图像的目标描述文本确定融合文本嵌入向量,使得融合文本嵌入向量能够表征图像描述文本的文本特征和目标描述文本的文本特征,采用融合文本嵌入向量指导稳定扩散模型对原始图像进行图像转换,能够在原始图像进行图像转换的过程中,尽量保持原始图像的特征结构,提高了图像转换结果的准确性,使得获取的目标图像具有期望的视觉观感。

附图说明

[0052] 图1为一个实施例中图像转换方法的应用环境图；

[0053] 图2为一个实施例中图像转换方法的流程示意图；

[0054] 图3为另一个实施例中图像转换方法的流程示意图；

[0055] 图4为另一个实施例中图像转换方法的流程示意图；

[0056] 图5为另一个实施例中图像转换方法的流程示意图；

[0057] 图6为一个实施例中图像转换装置的结构框图；

[0058] 图7为一个实施例中计算机设备的内部结构图。

具体实施方式

[0059] 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。

[0060] 本申请实施例提供的图像转换方法，可以应用于如图1所示的应用环境中。其中，终端102通过网络与服务器104进行通信。数据存储系统可以存储服务器104需要处理的数据。数据存储系统可以集成在服务器104上，也可以放在云上或其他网络服务器上。服务器104获取原始图像的图像描述文本，并通过文本编码器，根据目标描述文本和所述图像描述文本确定原始文本嵌入向量和目标文本嵌入向量；通过深度神经网络模型，根据所述原始文本嵌入向量和所述目标文本嵌入向量确定融合文本嵌入向量；通过稳定扩散模型，根据所述融合文本嵌入向量对所述原始图像进行图像转换，确定目标图像，通过通信网络将目标图像发送至终端102。其中，终端102可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑、物联网设备和便携式可穿戴设备，物联网设备可为智能音箱、智能电视、智能空调、智能车载设备等。便携式可穿戴设备可为智能手表、智能手环、头戴设备等。服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

[0061] 在一个实施例中，如图2所示，提供了一种图像转换方法，本实施例以该方法应用于终端进行举例说明，可以理解的是，该方法也可以应用于服务器，还可以应用于包括终端和服务器的系统，并通过终端和服务器的交互实现。本实施例中，该方法包括以下步骤：

[0062] S210、获取原始图像的图像描述文本。

[0063] 其中，原始图像是指需要进行域迁移的图像，图像描述文本是指描述原始图像中的图像内容的文本信息，图像描述文本可以是一个短语或句子，用于描述了原始图像中的主要对象。

[0064] S220、通过文本编码器，根据目标描述文本确定目标文本嵌入向量，并根据图像描述文本确定原始文本嵌入向量。

[0065] 其中，目标描述文本是指对原始图像进行图像转换时，所需要原始图像转换成的目标图像的描述文本。原始文本嵌入向量是指原始图像对应的文本嵌入向量，目标文本嵌入向量是指目标文本对应的文本嵌入向量。文本嵌入向量是对文本进行转换获取的固定大小的实数向量，这些向量能够捕获文本中的语义信息，使得语义上相似的文本在嵌入向量空间中具有相似的向量表示 。

[0066] 具体的，可以通过文本编码器，采用自然语言处理技术，根据目标描述文本和图像

描述文本确定原始文本嵌入向量和目标文本嵌入向量。

[0067] S230、通过深度神经网络模型，根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量。

[0068] 具体的，可以将原始文本嵌入向量和目标文本嵌入向量输入深度神经网络模型，通过深度神经网络模型中的语义融合网络层对原始文本嵌入向量和目标文本嵌入向量进行融合处理，确定原始文本嵌入向量和目标文本嵌入向量的融合文本嵌入向量。

[0069] S240、通过稳定扩散模型，根据融合文本嵌入向量对原始图像进行图像转换，确定目标图像。

[0070] 其中，稳定扩散模型是一种基于深度学习的图像生成技术，它通过模拟图像数据在噪声中的逐步扩散与恢复过程，能够生成高质量、多样化的图像。这种模型类似于一个智能画家，能够从随机噪声中“绘制”出逼真的图像，为图像创作、编辑及多领域应用提供了强大工具。例如，如果用户向稳定扩散模型输入“夏日海滩”的文本描述或关键词，模型能够捕捉到这个概念的核心特征，如蓝天、白云、金色的沙滩和波光粼粼的海面，然后生成一系列与之相符的夏日海滩景象图像，每一张都独具特色且细节丰富。

[0071] 具体的，采用融合文本嵌入向量作为引导条件，引导预训练的稳定扩散模型对原始图像进行图像转换，生成目标图像。

[0072] 上述图像转换方法中，获取原始图像的图像描述文本，通过文本编码器，根据目标描述文本确定目标文本嵌入向量，并根据图像描述文本确定原始文本嵌入向量；通过深度神经网络模型，根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量；通过稳定扩散模型，根据融合文本嵌入向量对原始图像进行图像转换，确定目标图像。解决了直接由文本条件主导图像过程而忽略了原始图像的内容和结构语义，存在生成的目标图像与原始图像视觉观感不一致的问题。上述方案，在对原始图像进行图像转换时，基于原始图像的图像描述文本，和期望生成的目标图像的目标描述文本确定融合文本嵌入向量，使得融合文本嵌入向量能够表征图像描述文本的文本特征和目标描述文本的文本特征，采用融合文本嵌入向量指导稳定扩散模型对原始图像进行图像转换，能够在原始图像进行图像转换的过程中，尽量保持原始图像的特征结构，提高了图像转换结果的准确性，使得获取的目标图像具有期望的视觉观感。

[0073] 在一个实施例中，获取原始图像的图像描述文本，包括：

[0074] 将原始图像输入图像描述生成器，确定所述原始图像的子图像；通过神经网络模型的线性嵌入向量层将所述子图像转换为图像嵌入向量；通过图像描述生成器中的视觉编码器，根据所述图像嵌入向量确定所述原始图像的图像特征；通过所述图像描述生成器中的视觉解码器，根据所述图像特征确定所述原始图像的图像描述文本。

[0075] 其中，图像描述生成器由视觉编码器和视觉文本解码器两部分组成，视觉编码器共有12层，每一层由层归一化、多头自注意力层、层归一化和多层感知机块组成。多头自注意力层被用于对图像的分割块进行自注意力计算，以捕捉不同分割块之间的依赖关系，使得整个图像的全局信息可以被编码到每个分割块的表示中。另外凭借其处理长距离像素之间关系的能力，更好的捕捉图像的语义信息。多层感知机块则通过对每个图像分割块进行特征提取进一步学习图像的复杂特征以及输出维度的调整满足后续的处理。同时通过残差连接更好地传递信息，减少信息丢失，并且减轻梯度消失带来的影响，可以更有效地学习到

输入数据的变化和重要特征。

[0076] 上述方案,通过神经网络模型和图像描述生成器获取原始图像的图像描述文本,能够直接根据原始图像获取到准确的图像描述文本,节约了人力,提高了图像描述文本的获取效率。

[0077] 在一个实施例中,如图3所示,根据目标描述文本和所述图像描述文本确定原始文本嵌入向量和目标文本嵌入向量,包括:

[0078] S310、确定图像描述文本的原始词元序列,以及目标描述文本的目标词元序列。

[0079] 其中,词元序列即token序列,指自然语言处理过程中用来表示单词或短语的符号。

[0080] 具体的,对图像描述文本进行分词处理,确定图像描述文本的token序列,即原始词元序列;对目标描述文本进行分词处理,确定目标描述文本的token序列,即目标词元序列。

[0081] S320、根据原始词元序列确定原始输入嵌入向量,并根据目标词元序列确定目标输入嵌入向量。

[0082] S330、根据原始输入嵌入向量确定原始嵌入向量矩阵,并根据目标输入嵌入向量确定目标嵌入向量矩阵。

[0083] 具体的,将原始输入嵌入向量进行整合,确定原始嵌入向量矩阵,将目标输入嵌入向量进行整合,确定目标嵌入向量矩阵。

[0084] 示例性的,原始嵌入向量矩阵如公式(1)所示:

[0085] $$X^0=\left[x_1^0,x_2^0,...,x_n^0\right] \quad (1)$$

[0086] 其中,X^0为原始嵌入向量矩阵。

[0087] S340、通过文本编码器,根据原始嵌入向量矩阵确定原始文本嵌入向量,并根据目标嵌入向量矩阵确定目标文本嵌入向量。

[0088] 具体的,将原始嵌入向量矩阵输入文本编码器,通过文本编码器,基于注意力机制的多层感知机确定原始嵌入向量矩阵中各原始输入嵌入向量之间的向量关系,根据各原始输入嵌入向量之间的向量关系对原始嵌入向量进行层归一化,确定原始文本嵌入向量。将目标嵌入向量矩阵输入文本编码器,通过文本编码器,基于注意力机制的多层感知机确定原始嵌入向量矩阵中各目标输入嵌入向量之间的向量关系,根据各目标输入嵌入向量之间的向量关系对目标嵌入向量进行层归一化,确定目标文本嵌入向量。

[0089] 示例性的,原始文本嵌入向量的计算公式如公式(2)所示:

[0090] $$emb_{text}=LN\left(MLP\left(Attention\left(X^{l-1}\right)\right)\right) \quad (2)$$

[0091] 其中,$\mathrm{emb}_{\mathrm{text}}$为原始文本嵌入向量,LN表示层归一化处理,MLP表示多层感知机,Attention表示注意力机制,X^{l-1}表示前一个文本编码器的输出数据。

[0092] 上述方案,根据原始词元序列确定原始输入嵌入向量,并根据目标词元序列确定目标输入嵌入向量,根据原始输入嵌入向量确定原始嵌入向量矩阵,并根据所述目标输入嵌入向量确定目标嵌入向量矩阵,根据原始嵌入向量矩阵确定原始文本嵌入向量,并根据目标嵌入向量矩阵确定目标文本嵌入向量,能够提高原始文本嵌入向量和目标文本嵌入向

量的可靠性。

[0093] 在一个实施例中,根据原始词元序列确定原始输入嵌入向量,并根据目标词元序列确定目标输入嵌入向量,包括:

[0094] 确定原始词元序列对应的原始词嵌入向量和原始位置嵌入向量,以及目标次元序列对应的目标词嵌入向量和目标位置嵌入向量;根据原始词嵌入向量和原始位置嵌入向量确定原始词元序列的原始输入嵌入向量;根据目标词嵌入向量和目标位置嵌入向量确定目标词元序列的目标输入嵌入向量。

[0095] 具体的,根据预定义的词汇表将获取到的原始词元序列和目标词元序列分别映射到一个特定的索引,以根据映射到的索引确定原始词元序列的token嵌入向量和位置嵌入向量,即原始词元序列的原始词嵌入向量和原始位置嵌入向量。将原始词元序列的原始词嵌入向量和原始词元序列的位置嵌入向量相加确定原始词元序列的原始输入嵌入向量。根据映射到的索引确定目标词元序列的目标词嵌入向量和目标位置嵌入向量。将目标词元序列的目标词嵌入向量和目标位置嵌入向量相加确定目标词元序列的目标输入嵌入向量。

[0096] 示例性的,原始输入嵌入向量的计算公式如公式(3)所示:

[0097] $$x_i^0 = e_i + p_i \quad (3)$$

[0098] 其中, x_i^0 为第i个原始输入嵌入向量, e_i 为原始词元序列的token嵌入向量, p_i 为原始词元序列的位置嵌入向量。

[0099] 上述方案,将原始词嵌入向量和原始位置嵌入向量相加,确定原始词元序列的原始输入嵌入向量,将目标词嵌入向量和目标位置嵌入向量相加,确定目标词元序列的目标输入嵌入向量,可以使得原始输入嵌入向量能够表征原始词元序列的token嵌入向量和位置嵌入向量,目标输入嵌入向量能够表征目标词元序列的token嵌入向量和位置嵌入向量,从而提高原始输入嵌入向量和目标输入嵌入向量的可靠性。

[0100] 在一个实施例中,如图4所示,通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量,包括:

[0101] S410、将目标文本嵌入向量输入深度神经网络模型,通过深度神经网络模型的多头自注意力层,确定目标文本的语义特征。

[0102] 需要说明的是,一般情况下,目标文本的描述信息更为详尽,因此将目标文本嵌入向量输入深度神经网络模型,能更好的通过深度神经网络模型的多头自注意力层捕捉到目标文本复杂的语义特征,便于后续更好的理解目标文本的语义。

[0103] S420、将原始文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层,生成图像描述文本的融合语义信息,将目标文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层,生成目标文本的融合语义信息。

[0104] 具体的,将原始文本嵌入向量、目标文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层,将原始文本嵌入向量作为查询语句query,将目标文本的语义特征作为键值key-value,通过目标文本的语义特征指导深度神经网络模型的交叉注意力层,生成原始文本嵌入向量对应的图像描述文本的融合语义信息,以及目标文本嵌入向量对应的目标文本的融合语义信息。

[0105] S430、通过深度学习网络的前馈网络层,根据目标文本的融合语义信息和图像描

述文本的融合语义信息确定原始文本嵌入向量和目标文本嵌入向量的融合文本嵌入向量。

[0106] 上述方案,通过深度神经网络模型的交叉注意力层,根据原始文本嵌入向量和目标文本的语义特征,确定目标文本和原始图像文本的融合语义信息,再深度学习网络的前馈网络层,根据目标文本的融合语义信息和图像描述文本的融合语义信息确定原始文本嵌入向量和目标文本嵌入向量的融合文本嵌入向量,能够提高原始文本嵌入向量和目标文本嵌入向量的准确性。

[0107] 在一个实施例中,如图5所示,通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像,包括:

[0108] S501、通过多模态预训练模型和图像分类模型,根据图像描述文本对原始图像进行语义分割,确定分割图像。

[0109] 其中,多模态预训练模型即CLIP(Contrastive Language-Image Pre-Training,多模态预训练模型)模型,CLIP模型是一种多模态预训练神经网络,其核心思想是使用大量图像和文本的配对数据进行预训练,以学习图像和文本之间的对齐关系。图像分类模型即ViT(VisionTransformer,图像分类)模型,ViT模型的核心思想是利用Transformer的注意力机制来对图像块之间的关系进行建模。注意力机制允许模型根据输入序列中的不同元素之间的关联性来分配不同的权重。通过多层的自注意力机制,ViT模型能够对图像块之间的关系进行编码和捕捉,从而实现对图像的全局理解。

[0110] 需要说明的是,图像是由许多像素组成,而图像语义分割就是将像素按照图像中表达语义含义的不同进行分组。

[0111] 具体的,通过多模态预训练模型中的Transformer解码器,根据原始图像的图像描述文本确定多模态嵌入向量,多模态嵌入向量即CLIP嵌入向量。将原始图像输入图像分类模型,提取图像分类模型中间层的激活函数,并将中间层的激活函数映射到解码器内部,并通过解码器根据原始图像的文本向量、中间层的激活函数和多模态嵌入向量输出原始图像的分割图像。在此基础上,原始图像根据图像描述文本的语义信息被分割为前景区域和背景区域,前景区域是指图像转换任务中需要进行图像翻译的主体区域,背景区域是指可以不进行图像转换的可保留区域。

[0112] 示例性的,多模态嵌入向量的确定公式如公式(4)所示:

[0113] $$emb_{clip} = Transformer_{Text}\left(Embed\left(Tokenize\left(Text_{desc}\right) + PosEnc\left(i\right)\right)\right) \quad (4)$$

[0114] 其中,emb_{clip}为多模态嵌入向量,$Tokenize(\cdot)$表示多模态预训练模型的标记函数,$Embed(\cdot)$为多模态嵌入向量的映射函数,即将图像描述文本的词元序列映射到多模态嵌入向量的函数,$PosEnc(\cdot)$为多模态预训练模型的位置编码函数,i为图像描述文本的词元序列映射到多模态嵌入向量时,图像描述文本的词元序列的位置索引。

[0115] S502、根据分割图像的像素信息确定分割图像的掩码图。

[0116] 具体的,根据分割图像的像素信息确定分割图像的置信度,将带有置信度的分割图像作为分割图像的掩码图。

[0117] S503、将掩码图输入到稳定扩散模型,基于融合文本嵌入向量生成特征空间向量,通过去噪扩散概率模型对特征空间向量进行降噪处理,确定目标空间向量。

[0118] 需要说明的是,去噪扩散概率模型是一种生成模型,它通过模拟物理过程中的扩

散和逆扩散来生成数据。这个过程包括两个主要部分:前向过程和逆向过程,其中前向过程即加噪过程,逆向过程即去噪过程。在加噪过程中,去噪扩散概率模型逐步地向数据中引入噪声,直到数据完全变成噪声。这个过程可以被看作是一个马尔可夫链,其中每一步都是确定的,并且最终会达到一个平衡状态,即数据的分布变为标准的高斯分布。例如,用户能够逐渐给一张清晰的图片添加随机的噪声,随着时间推移,用户不断地增加噪声,原本清晰的图片会变得越来越模糊,最终,原本清晰的图片会完全变成一片白噪声。去噪过程是加噪过程的逆操作,目标是从噪声中恢复出原始数据。这个过程通过逐步预测并去除噪声来实现。在实际应用中,通常使用一个神经网络来预测每一步中的噪声,然后从当前的噪声图像中减去这个预测的噪声,以此来还原前一时间步的图像。例如,若用户有一张完全由噪声组成的图片,它实际上是从清晰图片逐渐加噪得到的,去噪过程就是要逆转这个过程,就像是逐渐把雾吹散,让图片重新变得清晰,用户能够从噪声开始,尝试猜测和去除掉一些噪声,恢复出下一步的图片,通过不断重复这个过程,一步步地去除噪声,能够恢复出原始的清晰图片。

[0119] 具体的,将掩码图作为结构条件输入到稳定扩散模型,通过稳定扩散模型基于融合文本嵌入向量将掩码图从掩码图对应的像素空间转换为与原始图像的潜在表示相匹配的大小,再通过卷积神经网络对转换后的掩码图进行特征提取,从而确定掩码图对应的特征空间向量。

[0120] 需要说明的是,卷积神经网络由具有3×3卷积核以及2×2步长的卷积层组成。图像的潜在表示是通过自编码器技术,将图像数据压缩成尺寸更小的表示形式。通过去噪扩散概率模型对特征空间向量进行降噪处理,确定目标空间向量。

[0121] S504、根据目标空间向量对原始图像进行图像转换,确定目标图像。

[0122] 上述方案,采用多模态预训练模型和图像分类模型对原始图像进行语义分割确定分割图像,采用稳定扩散模型,将融合文本嵌入向量作为条件,引导稳定扩散模型对掩码图进行翻译,确定特征空间向量,通过降噪后的特征空间向量指导原始图像进行图像翻译,确定目标图像,能够提高目标图像的精确度。

[0123] 在一个实施例中,根据分割图像的像素信息确定分割图像的掩码图,包括:

[0124] 根据分割图像的像素信息确定分割图像的图像置信度,并根据图像置信度确定置信度矩阵;对置信度矩阵中的矩阵元素进行反处理,确定目标矩阵;根据目标矩阵和分割图像确定掩码图。

[0125] 需要说明的是,由于是以原始图像的描述作为条件对图像进行语义分割,分割的结果显示是前景部分更显著。为了将分割图的背景区域引入到图像翻译的过程中,需要让背景信息保留模块关注到分割图中背景部分而不是前景部分。分割图像的图像置信度可以根据分割图像中所包含的像素点的平均像素值确定,对原始图像的分割结果是一个带置信度的掩码图,置信度矩阵的大小与掩码图一致。需要对掩码图的置信度矩阵进行转换处理得到带有新置信度的掩码图以辅助后续图像翻译过程更好保留背景结构,即通过对置信度矩阵的每个矩阵元素进行反处理,根据反处理后的矩阵元素确定目标矩阵,可以使得掩码图符合后续处理的要求,此时经过语义分割处理后的掩码图会在背景区域有着更高的置信度,而在前景需要进行翻译的区域像素则为低置信度的掩码图。

[0126] 示例性的,对置信度矩阵中的矩阵元素进行反处理,以确定目标矩阵的公式如公

式(5)所示：

[0127] $$C'_{i,j} = 1 - C_{i,j} \quad (5)$$

[0128] 其中，$C'_{i,j}$为目标矩阵中中第i行第j列的矩阵元素，$C_{i,j}$表示置信度矩阵中第i行第j列的矩阵元素。

[0129] 需要说明的是，在上述实施例的基础上，稳定扩散模型的主干网络U-Net由8×8、16×16、32×32和64×64大小的成对编码器和解码器组成。

[0130] 去噪扩散概率模型主要是通过用主干网络预测在反向过程中特定时间步进行采样去噪图像的噪声与从高斯分布中随机选择的噪声进行损失函数的计算，该损失函数的计算公式如公式(6)所示：

[0131] $$\mathcal{L}_{DM} = \mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,1),t}\left[\left|\left|\epsilon - \epsilon_\theta(x_t,t)\right|\right|_2^2\right] \quad (6)$$

[0132] 其中，x_t是指时间步t的去噪样本，L_{DM}为噪扩散概率模型计算出的损失函数。

[0133] 稳定扩散模型的模型损失函数对文生图任务具有比较好的效果，但是并不符合图像转换任务的要求，因此在对稳定扩散模型进行训练时，将稳定扩散模型的模型损失函数作为模型训练过程中的总损失函数的一部分，以保证稳定扩散模型的模型输出图像与于融合文本嵌入向量对应的融合语义相匹配，强化了融合文本嵌入向量在图像转换过程中的指导作用。示例性的，稳定扩散模型的模型损失函数如公式(7)所示：

[0134] $$\mathcal{L}_{LDM} = \mathbb{E}_{\varepsilon(x),y,\epsilon\sim\mathcal{N}(0,1),t}\left[\left|\left|\epsilon - \epsilon_\theta(z_t,t,\tau_\theta(y))\right|\right|_2^2\right] \quad (7)$$

[0135] 其中，L_{LDM}为稳定扩散模型的模型损失函数，y是指文本条件；$\tau_\theta(y)$表示文本转换为条件向量的函数。

[0136] 进一步的，掩码图包括前景区域对应的图像和背景区域对应的图像，将背景区域对应的掩码图作为结构条件，能够在图像转换的过程中保留背景区域。基于噪扩散概率模型确定的损失函数对稳定扩散模型的模型损失函数进行调整，确定噪声预测损失函数，噪声预测损失函数的公式如公式(8)所示：

[0137] $$\mathcal{L}_{noise} = \mathbb{E}_{\varepsilon(x),c_{text},c_{bg},\epsilon\sim\mathcal{N}(0,1),t}\left[\left|\left|\epsilon - \epsilon_\theta(z_t,t,c_{text},c_{bg})\right|\right|_2^2\right] \quad (8)$$

[0138] 其中，z_t为时间步t时刻的潜在表示；c_{text}是指目标文本条件；c_{bg}是指背景结构条件。

[0139] 此外，为了确保目标图像和目标描述文本对其，使用多模态预训练模型的图像编

码器和文本编码器提取目标描述文本的文本特征和原始图像转换后获取到的目标图像的图像特征,根据目标表述文本的文本特征和目标图像的图像特征之间的特征差异性确定目标描述文本和目标图像之间的对其程度,即多模态预训练模型的模型损失函数,多模态预训练模型的模型损失函数的计算公式如公式(9)所示:

[0140] $$L_{CLIP} = -sim\left(E_{image}\left(I_{target}\right), E_{text}\left(T_{target}\right)\right) \quad (9)$$

[0141] 其中,L_{CLIP}为多模态预训练模型的模型损失函数,$E_{image}(\cdot)$为多模态预训练模型的图像编码器,$E_{text}(\cdot)$为多模态预训练模型的文本编码器,I_{target}为目标图像,T_{target}为目标描述文本。

[0142] 为了保证目标图像的背景区域和原始图像的背景区域尽可能一致,引入背景损失函数作为模型总损失函数的一部分,背景损失函数的计算公式如公式(10)所示:

[0143] $$L_{bg} = \sum_{i \in W, j \in H} C'_{i,j} \| I_{target,i,j} - I_{src,i,j} \|^2 \quad (10)$$

[0144] 其中,L_{bg}表示背景损失函数,$I_{target,i,j}$表示目标图像在(i,j)处像素值,$I_{src,i,j}$表示原始图像在(i,j)处像素值,W表示图像像素的宽度取值范围,H表示图像像素的高度取值范围。

[0145] 根据噪声预测损失函数、背景损失函数和多模态预训练模型的模型损失函数确定对稳定扩散模型进行训练时的总损失函数为如公式(11)所示:

[0146] $$L_{total} = \lambda_{noise} L_{noise} + \lambda_{CLIP} L_{CLIP} + \lambda_{bg} L_{bg} \quad (11)$$

[0147] 其中,L_{total}为总损失函数,λ_{noise}为噪声预测损失函数的权重,λ_{CLIP}为多模态预训练模型的模型损失函数的权重,λ_{bg}为背景损失函数的权重。

[0148] 示例性的,在上述实施例的基础上,图像转换方法包括:

[0149] 将原始图像输入图像描述生成器,确定所述原始图像的子图像;通过神经网络模型的线性嵌入向量层将所述子图像转换为图像嵌入向量;通过图像描述生成器中的视觉编码器,根据所述图像嵌入向量确定所述原始图像的图像特征;通过所述图像描述生成器中的视觉解码器,根据所述图像特征确定所述原始图像的图像描述文本。

[0150] 对图像描述文本进行分词处理,确定图像描述文本的token序列,即原始词元序列;对目标描述文本进行分词处理,确定目标描述文本的token序列,即目标词元序列。根据预定义的词汇表将获取到的原始词元序列和目标词元序列分别映射到一个特定的索引,以根据映射到的索引确定原始词元序列的token嵌入向量和位置嵌入向量,即原始词元序列的原始词嵌入向量和原始位置嵌入向量。将原始词元序列的原始词嵌入向量和原始词元序列的位置嵌入向量相加确定原始词元序列的原始输入嵌入向量。根据映射到的索引确定目标词元序列的目标词嵌入向量和目标位置嵌入向量。将目标词元序列的目标词嵌入向量和目标位置嵌入向量相加确定目标词元序列的目标输入嵌入向量。将原始输入嵌入向量进行整合,确定原始嵌入向量矩阵,将目标输入嵌入向量进行整合,确定目标嵌入向量矩阵。将原始嵌入向量矩阵输入文本编码器,通过文本编码器,基于注意力机制的多层感知机确定原始嵌入向量矩阵中各原始输入嵌入向量之间的向量关系,根据各原始输入嵌入向量之间

的向量关系对原始嵌入向量进行层归一化,确定原始文本嵌入向量。将目标嵌入向量矩阵输入文本编码器,通过文本编码器,基于注意力机制的多层感知机确定原始嵌入向量矩阵中各目标输入嵌入向量之间的向量关系,根据各目标输入嵌入向量之间的向量关系对目标嵌入向量进行层归一化,确定目标文本嵌入向量。

[0151] 将目标文本嵌入向量输入深度神经网络模型,通过深度神经网络模型的多头自注意力层,确定目标文本的语义特征。将原始文本嵌入向量、目标文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层,将原始文本嵌入向量作为查询语句query,将目标文本的语义特征作为键值key-value,通过目标文本的语义特征指导深度神经网络模型的交叉注意力层,生成原始文本嵌入向量对应的图像描述文本的融合语义信息,以及目标文本嵌入向量对应的目标文本的融合语义信息。通过深度学习网络的前馈网络层,根据目标文本的融合语义信息和图像描述文本的融合语义信息确定原始文本嵌入向量和目标文本嵌入向量的融合文本嵌入向量。

[0152] 通过多模态预训练模型中的Transformer解码器,根据原始图像的图像描述文本确定多模态嵌入向量,多模态嵌入向量即CLIP嵌入向量。将原始图像输入图像分类模型,提取图像分类模型中间层的激活函数,并将中间层的激活函数映射到解码器内部,并通过解码器根据原始图像的文本向量、中间层的激活函数和多模态嵌入向量输出原始图像的分割图像。在此基础上,原始图像根据图像描述文本的语义信息被分割为前景区域和背景区域,前景区域是指图像转换任务中需要进行图像翻译的主体区域,背景区域是指可以不进行图像转换的可保留区域。

[0153] 根据分割图像的像素信息确定分割图像的图像置信度,并根据图像置信度确定置信度矩阵;对置信度矩阵中的矩阵元素进行反处理,确定目标矩阵;根据目标矩阵和分割图像确定掩码图。将掩码图作为结构条件输入到稳定扩散模型,通过稳定扩散模型基于融合文本嵌入向量将掩码图从掩码图对应的像素空间转换为与原始图像的潜在表示相匹配的大小,再通过卷积神经网络对转换后的掩码图进行特征提取,从而确定掩码图对应的特征空间向量。通过降噪后的特征空间向量指导原始图像进行图像翻译,确定目标图像。

[0154] 上述图像转换方法中,获取原始图像的图像描述文本,并通过文本编码器,根据目标描述文本和图像描述文本确定原始文本嵌入向量和目标文本嵌入向量;通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。解决了直接由文本条件主导图像过程而忽略了原始图像的内容和结构语义,存在生成的目标图像与原始图像视觉观感不一致的问题。上述方案,在对原始图像进行图像转换时,基于原始图像的图像描述文本,和期望生成的目标图像的目标描述文本确定融合文本嵌入向量,使得融合文本嵌入向量能够表征图像描述文本的文本特征和目标描述文本的文本特征,采用融合文本嵌入向量指导稳定扩散模型对原始图像进行图像转换,能够在原始图像进行图像转换的过程中,尽量保持原始图像的特征结构,提高了图像转换结果的准确性,使得获取的目标图像具有期望的视觉观感。

[0155] 应该理解的是,虽然如上所述的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而

且,如上所述的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0156] 基于同样的发明构思,本申请实施例还提供了一种用于实现上述所涉及的图像转换方法的图像转换装置。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似,故下面所提供的一个或多个图像转换装置实施例中的具体限定可以参见上文中对于图像转换方法的限定,在此不再赘述。

[0157] 在一个实施例中,提供了一种图像转换系统,包括:

[0158] 原始图像输入界面,用于输入原始图像;

[0159] 描述文本输入界面,用于输入目标描述文本;

[0160] 目标图像展示界面,用于展示目标图像;目标图像的生成方式为:通过文本编码器,根据原始图像的目标描述文本和图像描述文本确定原始文本嵌入向量和目标文本嵌入向量;通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。

[0161] 在一个实施例中,如图6所示,提供了一种图像转换装置,包括:图像描述文本确定模块601、目标文本嵌入确定模块602、融合文本嵌入确定模块603和目标图像确定模块604,其中:

[0162] 图像描述文本确定模块601,用于获取原始图像的图像描述文本;

[0163] 目标文本嵌入确定模块602,用于通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据图像描述文本确定原始文本嵌入向量;

[0164] 融合文本嵌入确定模块603,用于通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;

[0165] 目标图像确定模块604,用于通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。

[0166] 示例性的,图像描述文本确定模块601具体用于:

[0167] 将原始图像输入图像描述生成器,确定原始图像的子图像;

[0168] 通过神经网络模型的线性嵌入向量层将子图像转换为图像嵌入向量;

[0169] 通过图像描述生成器中的视觉编码器,根据图像嵌入向量确定原始图像的图像特征;

[0170] 通过图像描述生成器中的视觉解码器,根据图像特征确定原始图像的图像描述文本。

[0171] 示例性的,目标文本嵌入确定模块602具体用于:

[0172] 确定图像描述文本的原始词元序列,以及目标描述文本的目标词元序列;

[0173] 根据原始词元序列确定原始输入嵌入向量,并根据目标词元序列确定目标输入嵌入向量;

[0174] 根据原始输入嵌入向量确定原始嵌入向量矩阵,并根据目标输入嵌入向量确定目标嵌入向量矩阵;

[0175] 通过文本编码器,根据原始嵌入向量矩阵确定原始文本嵌入向量,并根据目标嵌入向量矩阵确定目标文本嵌入向量。

[0176] 示例性的,目标文本嵌入确定模块602还具体用于:

[0177] 确定原始词元序列对应的原始词嵌入向量和原始位置嵌入向量,以及目标次元序列对应的目标词嵌入向量和目标位置嵌入向量;

[0178] 根据原始词嵌入向量和原始位置嵌入向量确定原始词元序列的原始输入嵌入向量;

[0179] 根据目标词嵌入向量和目标位置嵌入向量确定目标词元序列的目标输入嵌入向量。

[0180] 示例性的,融合文本嵌入确定模块603具体用于:

[0181] 将目标文本嵌入向量输入深度神经网络模型,通过深度神经网络模型的多头自注意力层对目标文本嵌入向量进行特征提取,生成目标文本的语义特征;

[0182] 将原始文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层,生成图像描述文本的融合语义信息,将目标文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层,生成目标文本的融合语义信息;

[0183] 通过深度学习网络的前馈网络层,根据目标文本的融合语义信息和图像描述文本的融合语义信息确定原始文本嵌入向量和目标文本嵌入向量的融合文本嵌入向量。

[0184] 示例性的,目标图像确定模块604具体用于:

[0185] 通过多模态预训练模型和图像分类模型,根据图像描述文本对原始图像进行语义分割,确定分割图像;

[0186] 根据分割图像的像素信息确定分割图像的掩码图;

[0187] 将掩码图输入到稳定扩散模型,基于融合文本嵌入向量生成特征空间向量,通过去噪扩散概率模型对特征空间向量进行降噪处理,确定目标空间向量;

[0188] 根据目标空间向量对原始图像进行图像转换,确定目标图像。

[0189] 示例性的,目标图像确定模块604还具体用于:

[0190] 根据分割图像的像素信息确定分割图像的图像置信度,并根据图像置信度确定置信度矩阵;

[0191] 对置信度矩阵中的矩阵元素进行反处理,确定目标矩阵;

[0192] 根据目标矩阵和分割图像确定掩码图。

[0193] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是终端,其内部结构图可以如图7所示。该计算机设备包括处理器、存储器、输入/输出接口、通信接口、显示单元和输入装置。其中,处理器、存储器和输入/输出接口通过系统总线连接,通信接口、显示单元和输入装置通过输入/输出接口连接到系统总线。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端进行有线或无线方式的通信,无线方式可通过WIFI、移动蜂窝网络、NFC(近场通信)或其他技术实现。该计算机程序被处理器执行时以实现一种图像转换方法。该计算机设备的显示单元用于形成视觉可见的画面,可

以是显示屏、投影装置或虚拟现实成像装置。显示屏可以是液晶显示屏或者电子墨水显示屏,该计算机设备的输入装置可以是显示屏上覆盖的触摸层,也可以是计算机设备外壳上设置的按键、轨迹球或触控板,还可以是外接的键盘、触控板或鼠标等。

[0194] 本领域技术人员可以理解,图7中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0195] 在一个实施例中,提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现以下步骤:

[0196] 步骤一、获取原始图像的图像描述文本;

[0197] 步骤二、通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据图像描述文本确定原始文本嵌入向量;

[0198] 步骤三、通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;

[0199] 步骤四、通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。

[0200] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现以下步骤:

[0201] 步骤一、获取原始图像的图像描述文本;

[0202] 步骤二、通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据图像描述文本确定原始文本嵌入向量;

[0203] 步骤三、通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;

[0204] 步骤四、通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。

[0205] 在一个实施例中,提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现以下步骤:

[0206] 步骤一、获取原始图像的图像描述文本;

[0207] 步骤二、通过文本编码器,根据目标描述文本确定目标文本嵌入向量,并根据图像描述文本确定原始文本嵌入向量;

[0208] 步骤三、通过深度神经网络模型,根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量;

[0209] 步骤四、通过稳定扩散模型,根据融合文本嵌入向量对原始图像进行图像转换,确定目标图像。

[0210] 需要说明的是,本申请所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据,且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

[0211] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机

可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存、光存储器、高密度嵌入向量式非易失性存储器、阻变存储器(ReRAM)、磁变存储器(MagnetoresistiveRandom Access Memory,MRAM)、铁电存储器(Ferroelectric Random Access Memory,FRAM)、相变存储器(PhaseChange Memory,PCM)、石墨烯存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器等。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(StaticRandom Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等,不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等,不限于此。

[0212] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0213] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对本申请专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请的保护范围应以所附权利要求为准。

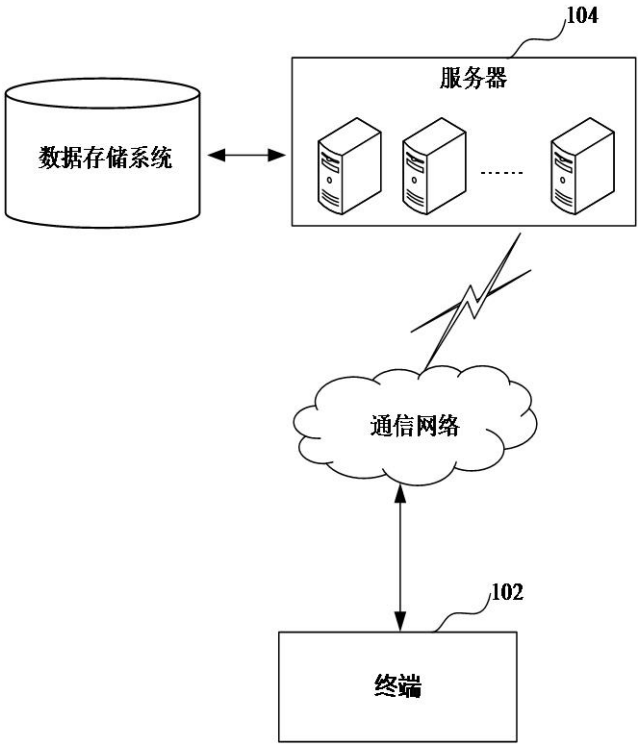


图 1

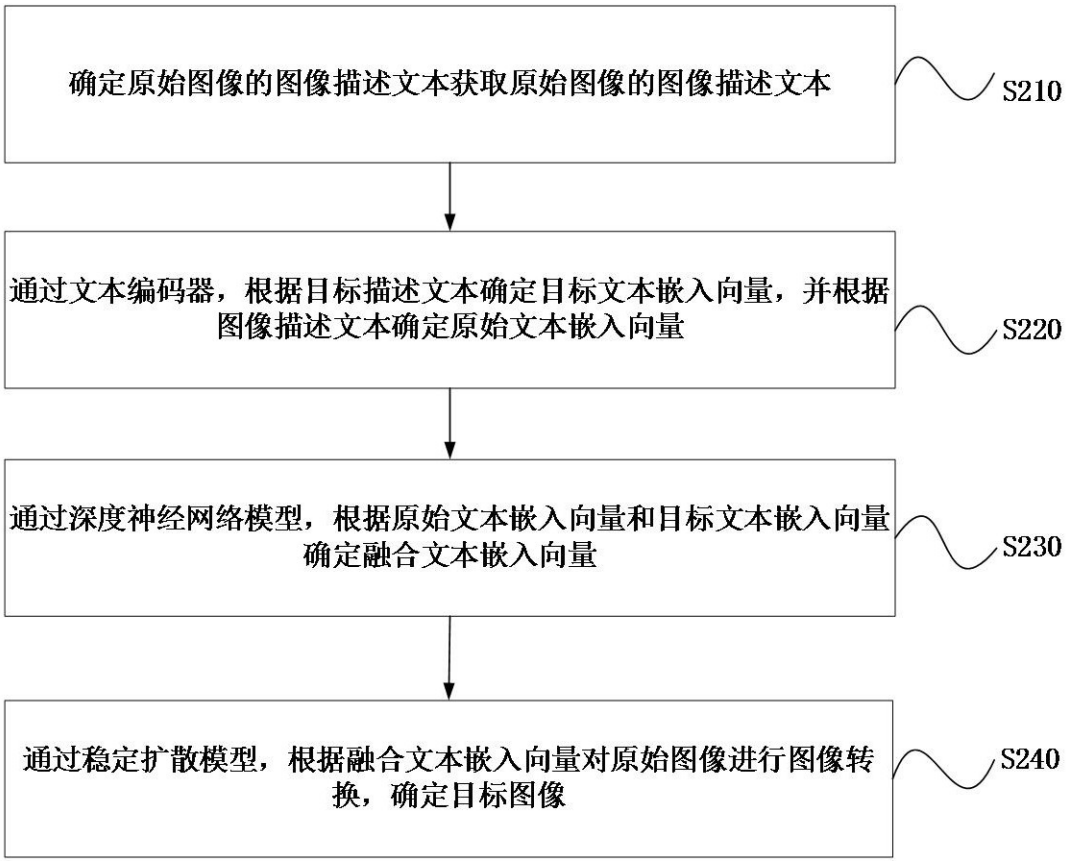

确定原始图像的图像描述文本获取原始图像的图像描述文本
S210
通过文本编码器，根据目标描述文本确定目标文本嵌入向量，并根据图像描述文本确定原始文本嵌入向量
S220
通过深度神经网络模型，根据原始文本嵌入向量和目标文本嵌入向量确定融合文本嵌入向量
S230
通过稳定扩散模型，根据融合文本嵌入向量对原始图像进行图像转换，确定目标图像
S240

图 2

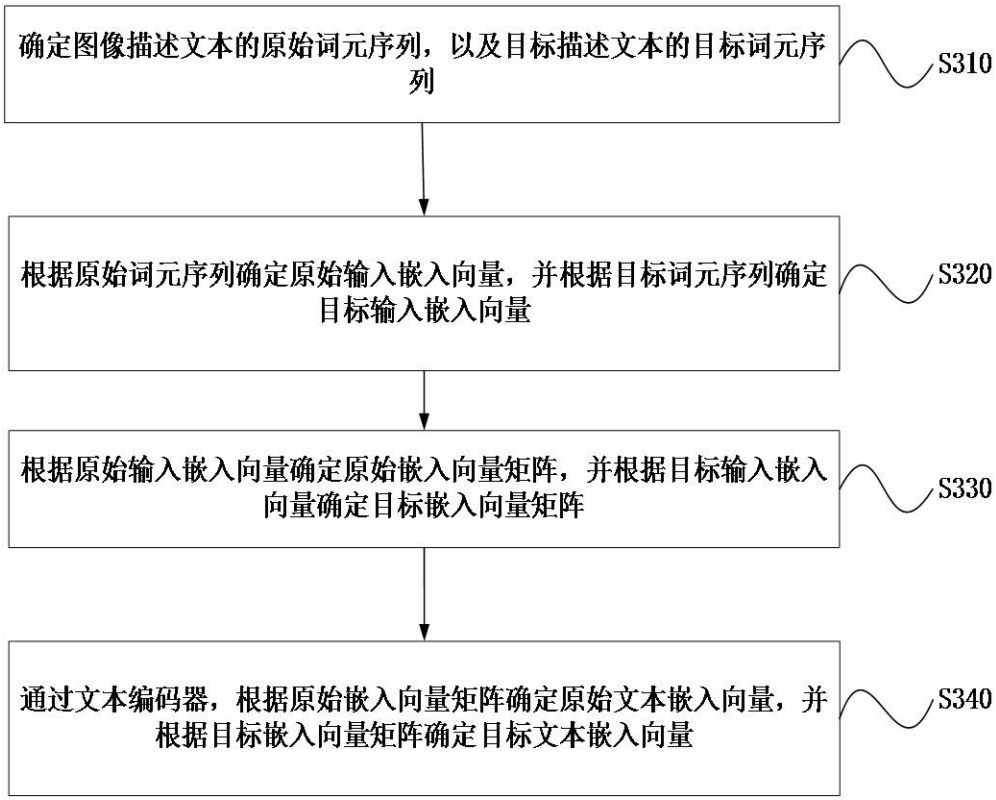

确定图像描述文本的原始词元序列，以及目标描述文本的目标词元序列
S310
根据原始词元序列确定原始输入嵌入向量，并根据目标词元序列确定目标输入嵌入向量
S320
根据原始输入嵌入向量确定原始嵌入向量矩阵，并根据目标输入嵌入向量确定目标嵌入向量矩阵
S330
通过文本编码器，根据原始嵌入向量矩阵确定原始文本嵌入向量，并根据目标嵌入向量矩阵确定目标文本嵌入向量
S340

图3

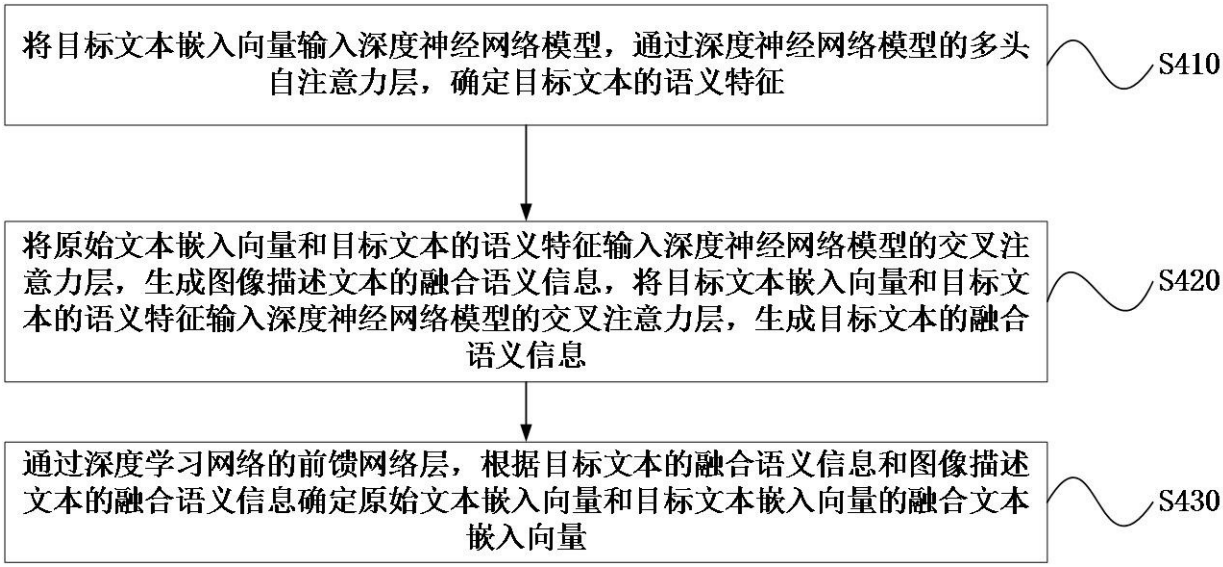

将目标文本嵌入向量输入深度神经网络模型，通过深度神经网络模型的多头自注意力层，确定目标文本的语义特征
S410
将原始文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层，生成图像描述文本的融合语义信息，将目标文本嵌入向量和目标文本的语义特征输入深度神经网络模型的交叉注意力层，生成目标文本的融合语义信息
S420
通过深度学习网络的前馈网络层，根据目标文本的融合语义信息和图像描述文本的融合语义信息确定原始文本嵌入向量和目标文本嵌入向量的融合文本嵌入向量
S430

图4

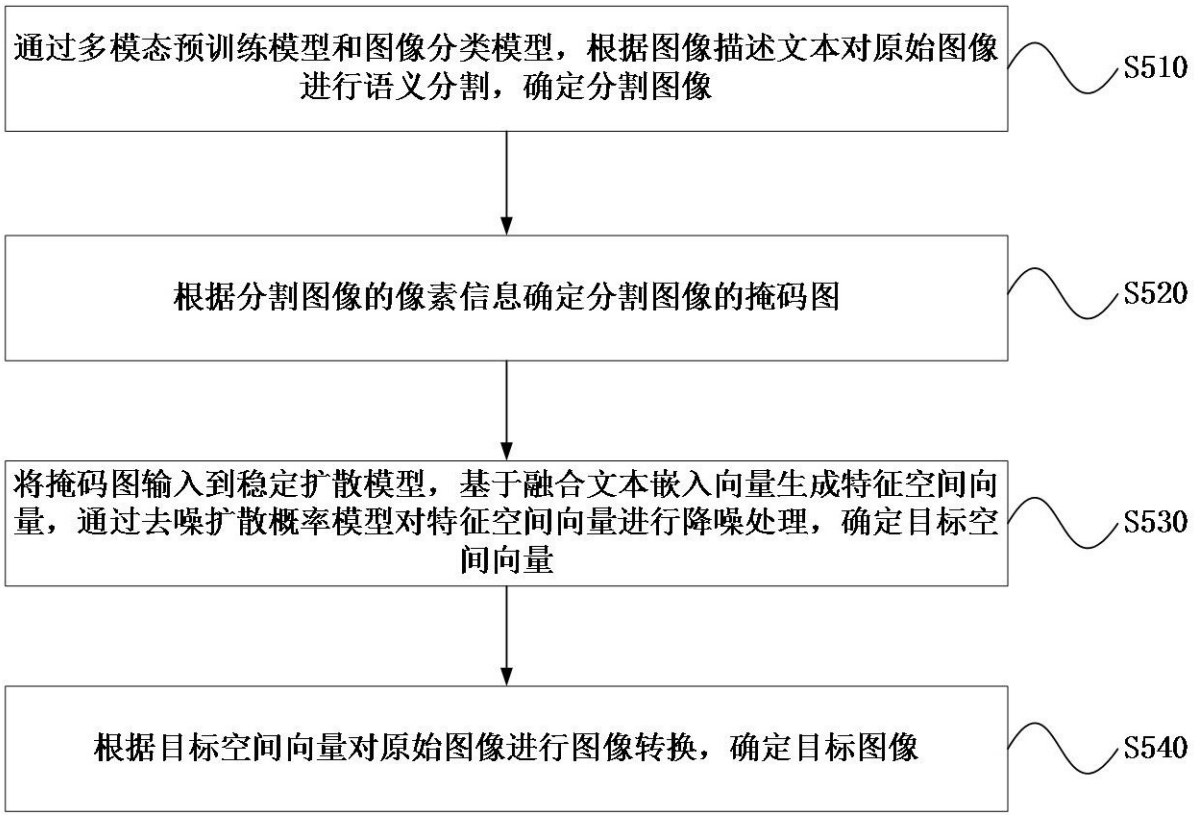

通过多模态预训练模型和图像分类模型，根据图像描述文本对原始图像进行语义分割，确定分割图像
S510
根据分割图像的像素信息确定分割图像的掩码图
S520
将掩码图输入到稳定扩散模型，基于融合文本嵌入向量生成特征空间向量，通过去噪扩散概率模型对特征空间向量进行降噪处理，确定目标空间向量
S530
根据目标空间向量对原始图像进行图像转换，确定目标图像
S540

图5

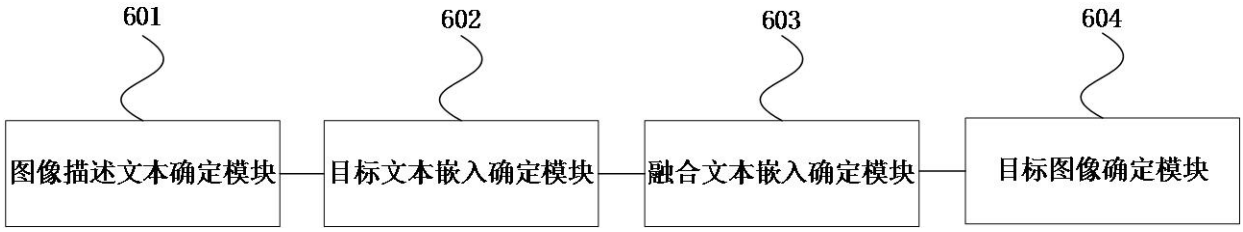

601
602
603
604
图像描述文本确定模块
目标文本嵌入确定模块
融合文本嵌入确定模块
目标图像确定模块

图6

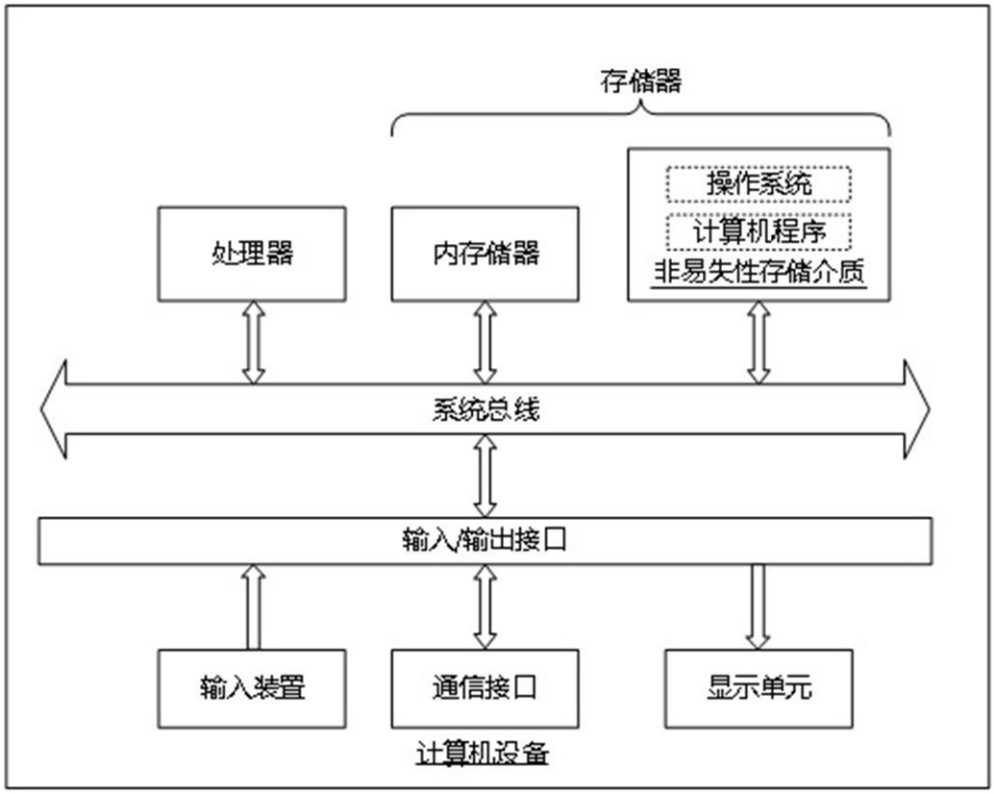

存储器
操作系统
计算机程序
非易失性存储介质
处理器
内存储器
系统总线
输入/输出接口
输入装置
通信接口
显示单元
计算机设备

图7