



(12)发明专利申请

(10)申请公布号 CN 110647632 A
(43)申请公布日 2020.01.03

(21)申请号 201910722448.7

(22)申请日 2019.08.06

(71)申请人 上海孚典智能科技有限公司
地址 200082 上海市杨浦区国定路323号15层(集中登记地)

(72)发明人 赵继胜 吴宇

(74)专利代理机构 上海精晟知识产权代理有限公司 31253
代理人 肖爱华

(51) Int. Cl.
G06F 16/36(2019.01)
G06F 16/35(2019.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

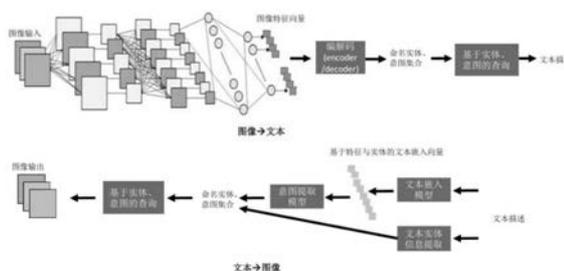
权利要求书2页 说明书3页 附图2页

(54)发明名称

基于机器学习的图像与文本映射技术

(57)摘要

本发明提供了一种基于语义分析与图像特征提取技术相结合的图像到说明文本映射技术,通过该技术,可以由输入图像得出相关的文本描述。本项专利建立了图像与文本描述的映射关系,通过卷积神经网络提取图像特征向量,并与文本信息中的语义信息建立关联。这里的语义信息由时间卷积网络结合条件随机场技术,可以提取文本中的意图信息和实体信息,再通过对“意图”和“实体”两类信息进行编码并与图像特征向量建立对应模型,由该模型实现图像特征向量到文本的预测能力。即由输入图像得出基本的文本描述。该技术应用于医疗图像和智慧城市中的数据分析,可以有效减轻操作和维护人员的工作量。



1. 基于机器学习的图像与文本映射技术,包括:

随着人工智能技术的飞速发展,不同领域的各种基于人工智能的智能应用层出不穷,特别是在图形图像识别和自然语言处理领域,人工智能所提供的智能化服务正逐步占据主流应用的地位。在医疗健康领域,医疗工作者往往需要针对医疗图像快速给出说明性描述,为病理分析提供支撑,因此引出了从图像到文本的映射技术的需求。同时,在智慧城市领域,往往需要根据一段语言描述,快速定位相应的地标性建筑或场馆,因此需要提供从文本到图像的高效索引技术。实现上述需求,需要具备从图像和文本中提取关键特性和表达意图的能力。例如从医疗图像中提取可疑病灶的能力,从文本说明中提取作者的表达意图的能力。根据目前神经网络技术的发展,对应于特征提取和意图定位所涉及的技术包括:深度卷积网络、词向量、文本向量技术和向量到集合映射技术。

本项发明设计了一种基于多种神经网络技术的框架,该框架用以实现对不同数据类型进行映射,从而实现以非结构化数据作为输入的查询能力,即通过图片查询文本和通过文本查询图片(如说明书附图1所示),具体包括:

A. 设计了对文本类型的非结构化信息的特征提取和向量化表达,所用算法可有两种选择:

I. 以doc2vec为基础,该算法是google词向量技术的扩展,通过采用针对宽采样窗体(采样宽度为200)来实现对一般医疗图片和建筑场景描述的文本信息的精确特征捕捉和特征向量生成(参见说明书附图2(b));

II. 以google bert模型为基础,作为一种高效和通用的文本嵌入模型,bert可以更为通用的场景描述提供特征捕捉和向量生成(参见说明书附图2(b))。

B. 设计了从文本特征向量映射到文本意图集合的机器学习模型(参见说明书附图2(b));

C. 设计了对图片类型的非结构化信息的特征提取和向量化表达,所用算法以残差网络resnet为基础,该算法已经被广泛应用于面部识别、复杂图形的特征分析(例如alphago的棋盘状态分析),该网络以深度卷积网络为基础,通过引入残差连接,避免了因增加网络深度导致的信息丢失。采用深度残差网络可以更有效的提取复杂图片的特征,特别是网络深度可以有效满足不同类型的信息提取需求(参见说明书附图2(a));

D. 图片特征向量到命名实体映射技术,所用算法为基于卷积神经网络的建模技术(参见说明书附图2(a));

E. 图谱特征向量到文本意图的映射技术,所用算法为基于卷积神经网络的建模技术(参见说明书附图2(a));

查询接口,通过输入图片/文本,为用户提供对应的文本/图片输出,在输出方面采用top-K方式,即提供相似度最高的K个查询结果供用户选择。

2. 根据权利要求1. 基于机器学习的图像与文本映射技术,其特征在于以应用框架的方式实现图像到文本与文本到图像的映射,用户可用灵活使用不同的特征提取机制还针对应用场景提升映射精度。

3. 根据权利要求1. 基于机器学习的图像与文本映射技术,其特征在于以在文本特征向量的生成可选择使用doc2vec或bert预训练模型,上述两种方式的输出特征向量长度均设定为512,并支持新的文本向量预训练模型。

4. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以ResNet50为基础生成图像特征向量,输出特征向量长度设定为256,全连接层神经元数量为256。

5. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以采用深度卷积神经网络GoogleNet实现图像和文本特征向量到命名实体和意图集合的映射。

6. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以采用top-k相似度,即最为相似的k个查询结果作为通过文本查询图像或通过图像查询文本的输出。

7. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以文本数据的命名实体提取采用前述文本特征向量通过条件随机场技术实现。

8. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以文本数据的意图提取采用前述文本特征向量通过卷积神经网络进行映射,这里采用GoogleNet作为卷积神经网络架构。

9. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以图像数据的命名实体提取采用前述图像特征向量通过卷积神经网络进行映射,这里采用GoogleNet作为卷积神经网络架构。

10. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以图像数据的意图提取采用前述图像特征向量通过卷积神经网络进行映射,这里采用GoogleNet作为卷积神经网络架构。

11. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以输入文本,转换文本的特征向量,再获取文本命名实体和意图集合,通过命名实体和意图集合与图片库中的实体和意图进行比对,提取最为接近的k张图片作为返回(top-K策略)。

12. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以输入图片,转换图片的特征向量,再获取文本命名实体和意图集合,通过命名实体和意图集合进入文本库并对进行实体和意图进行比,提取最为接近的k张文本作为返回(top-K策略)。

13. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以数据存储:命名实体与意图集合→文本/图片数据的键/值存储方式(参见图3),数据表可以看作是键/值数据的列表,排列顺序以命名实体与意图以字符顺序进行排序。

14. 根据权利要求1基于机器学习的图像与文本映射技术,其特征以数据查询:根据输入命名实体与意图集合进行在键/值数据列表中查询与之相似度最高的k条记录,计算相似度的算法可采用集合重合率进行比对: p (命名实体重合率)+ q (意图重合率), p 和 q 代表相似度换算方程,通过重合率乘以用户定义的固定系数来计算,通常 $p=0.4$, $q=0.6$,计算值越接近1表示越接近。

基于机器学习的图像与文本映射技术

技术领域

[0001] 本发明属于涉及信息技术领域,尤其是涉及采用深度神经网络对图像和文本之间建立关联和映射的技术。该项技术可以用来进行图像描述生成,依据描述搜索图像信息,因此可以在医疗领域(例如根据医疗图像获得相关评述)或智慧城市领域(例如依据基本描述定位地标建筑)作为智能应用的算法模块提供服务。

背景技术

[0002] 随着人工智能技术的飞速发展,不同领域的各种基于人工智能的智能应用层出不穷,特别是在图形图像识别和自然语言处理领域,人工智能所提供的智能化服务正逐步占据主流应用的地位。在医疗健康领域,医疗工作者往往需要针对医疗图像快速给出说明性描述,为病理分析提供支撑,因此引出了从图像到文本的映射技术的需求。同时,在智慧城市领域,往往需要根据一段语言描述,快速定位相应的地标性建筑或场馆,因此需要提供从文本到图像的搞笑索引技术。

[0003] 实现上述需求,需要具备从图像和文本中提取关键特性和表达意图的能力。例如从医疗图像中提取可疑病灶的能力,从文本说明中提取中作者的命名实体和表达意图的能力。根据目前深度神经网络技术为技术的发展,对应于特征提取和意图定位所涉及的技术包括:深度卷积网络、词向量、文本向量技术和序列映射技术。

[0004] 本项发明通过提供一种基于多种深度神经网络技术的框架,来实现基于非结构化信息作为查询输入的查询系统,实现图像到文本和文本到图像查询的技术能力,为满足医疗健康领域的辅助诊断和智慧城市领域的智能索引技术需求提供有效支撑。

发明内容

[0005] 本项发明设计了一种基于多种深度神经网络技术的框架,该框架用以实现对不同数据类型进行映射,从而实现以非结构化数据作为输入的查询能力,即通过图片查询文本和通过文本查询图片(如图1所示),具体包括:

[0006] 1. 设计了对文本类型的非结构化信息的特征提取和向量化表达,所用算法可有两种选择:

[0007] I. 以doc2vec为基础,该算法是google词向量技术的扩展,通过采用针对宽采样窗体(采样宽度为200)来实现对一般医疗图片和建筑场景描述的文本信息的精确特征捕捉和特征向量生成(参见图2(b));

[0008] II. 以google bert模型为基础,作为一种高效和通用的文本嵌入模型,bert可以更为通用的场景描述提供特征捕捉和向量生成(参见图2(b)).

[0009] 2. 设计了对文本特征向量映射到文本意图集合的机器学习模型(参见图2(b));

[0010] 3. 设计了对图片类型的非结构化信息的特征提取和向量化表达,所用算法以残差网络resnet为基础,该算法已经被广泛应用于面部识别、复杂图形的特征分析(例如alphago的棋盘状态分析),该网络以深度卷积网络为基础,通过引入残差连接,避免了因增

加网络深度导致的信息丢失。采用深度残差网络可以更有效的提取复杂图片的特征,特别是网络深度可以有效满足不同类型的信息提取需求(参见图2(a));

[0011] 4. 图片特征向量到命名实体映射技术,所用算法为基于卷积神经网络的建模技术(参见图2(a));

[0012] 5. 图谱特征向量到文本意图的映射技术,所用算法为基于卷积神经网络的建模技术(参见图2(a));

[0013] 6. 查询接口,通过输入图片/文本,为用户提供对应的文本/图片输出,在输出方面采用top-K方式,

[0014] 即提供相似度最高的K个查询结果供应用方选择。

[0015] 基于非结构化信息(文本、图片)查询系统的构建包括如下步骤:

[0016] 1. 构建特征提取训练模型:

[0017] a. 对于文本类型的特征提取模型:i. 可以直接使用bert文本向量生成模型;ii. 使用doc2vec技术在收集的医疗图片描述和建筑场景描述素材上构建文本向量化模型(参见图4(b));

[0018] b. 对于图片类型的特征提取模型:收集图片和分类标注信息作为训练样本,通过resnet网络架构训练深度神经网络,并以训练好的网络的全连接层输出作为特征提取向量输出(参见图4(a))。

[0019] 2. 特征映射模型构建:

[0020] 采用卷积神经网络,通过采用GoogleNet作为网络架构,分别训练:文本特征向量到命名实体集合,文本特征向量到文本意图集合,图片特征向量到命名实体集合,图片特征向量到文本意图集合的映射模型(参见图4(a)(b))。

[0021] 3. 基于非结构化数据的查询:

[0022] a. 输入文本,转换文本的特征向量,再获取文本命名实体和意图集合,通过命名实体和意图集合与图片库中的实体和意图进行比对,提取最为接近的k张图片作为返回(top-K策略);

[0023] b. 输入图片,转换图片的特征向量,再获取文本命名实体和意图集合,通过命名实体和意图集合进入文本库并对进行实体和意图进行比,提取最为接近的k张文本作为返回(top-K策略)。本项发明的上述技术方案有益结果如下:

[0024] 在医疗图像处理领域,越来越多智能应用需要快速的为给定图片提供文字说明,因此需要一种从图像到文本的查询和映射能力。在智慧城市领域,则需要提供一种从文本到建筑景观图片的查询能力,因此需要一种从文本到图像的查询映射能力。基于上述两点需求,需要通过非结构化数据的查询能力,本项发明提供了可实现图片→文本、文本→图片的以非结构化数据为输入进行查询的技术框架。该框架以机器学习技术为基础,特别是使用深度神经网络进行特征提取和分析,到特征映射建模,最终实现在文本和图片两种非结构化数据之间建立映射关系。该框架可作为为智能应用实现文本/图片相互查询工作所需的基础平台,为满足医疗和智慧城市领域的需求提供帮助。

附图说明

[0025] 图1由图像生成文本描述的样例

- [0026] 图2图像/文本映射的技术框架
- [0027] 图3命名实体、意图集合与非结构化数据存储结构
- [0028] 图4图像/文本映射的技术实现

具体实施方式

[0029] 根据发明内容中所阐述的构建通过文本和图像作为查询输入的非结构化信息查询系统的内容,其具体实现如下几节所述:

[0030] 构建特征提取训练模型:

[0031] a.对于文本类型的特征提取模型:i.可以直接使用bert文本向量生成模型;ii.使用doc2vec技术在收集的医疗图片描述和建筑场景描述素材上构建文本向量化模型;上述两种方式的输出特征向量长度均设定为512(参见图4(b));

[0032] b.对于图片类型的特征提取模型:收集图片和分类标注信息作为训练样本,通过resnet网络架构训练深度神经网络,并以训练好的网络的全连接层输出作为特征提取向量输出;上述方式的resnet层数为50,输出特征向量长度设定为256,全连接层神经元数量为256(参见图4(a)).

[0033] 图像数据的命名实体和意图提取:

[0034] a.图像数据的命名实体提取采用前述图像特征向量通过卷积神经网络进行映射,这里采用GoogLeNet作为卷积神经网络架构(参见图4(a));

[0035] b.图像数据的意图提取采用前述图像特征向量通过卷积神经网络进行映射,这里采用GoogLeNet作为卷积神经网络架构(参见图4(a)).

[0036] 文本数据的命名实体和意图提取:

[0037] c.文本数据的命名实体提取采用前述文本特征向量通过条件随机场技术实现(参见图4(b));

[0038] d.文本数据的意图提取采用前述文本特征向量通过卷积神经网络进行映射,这里采用GoogLeNet作为卷积神经网络架构(参见图4(b)).

[0039] 基于非结构化数据的查询:

[0040] a.数据存储:命名实体与意图集合→文本/图片数据的键/值存储方式(参见图3),数据表可以看作是键/值数据的列表,排列顺序以命名实体与意图以字符顺序进行排序;

[0041] b.数据查询:根据输入命名实体与意图集合进行在键/值数据列表中查询与之相似度最高的k条记录,计算相似度的算法可采用集合重合率进行比对: p (命名实体重合率)+ q (意图重合率), p 和 q 代表相似度换算方程,通过重合率乘以用户定义的固定系数来计算,通常 $p=0.4$, $q=0.6$,计算值越接近1表示越接近;

[0042] c.输入文本,转换文本的特征向量,采用文本到图片的特征向量映射模型得到对应图片的特征向量,然后计算得出命名实体与意图集合,进入图片库进行命名实体与意图的相似度比对,提取最为接近的k张图片作为返回(top-K策略);

[0043] d.输入图片,转换图片的特征向量,采用图片到文本的特征向量映射模型得到特征向量,通过特征向量到命名实体与意图集合的映射模型的到命名实体与意图集合,进入文本库进行命名实体与意图相似度比对,提取最为接近的k张文本作为返回(top-K策略)。

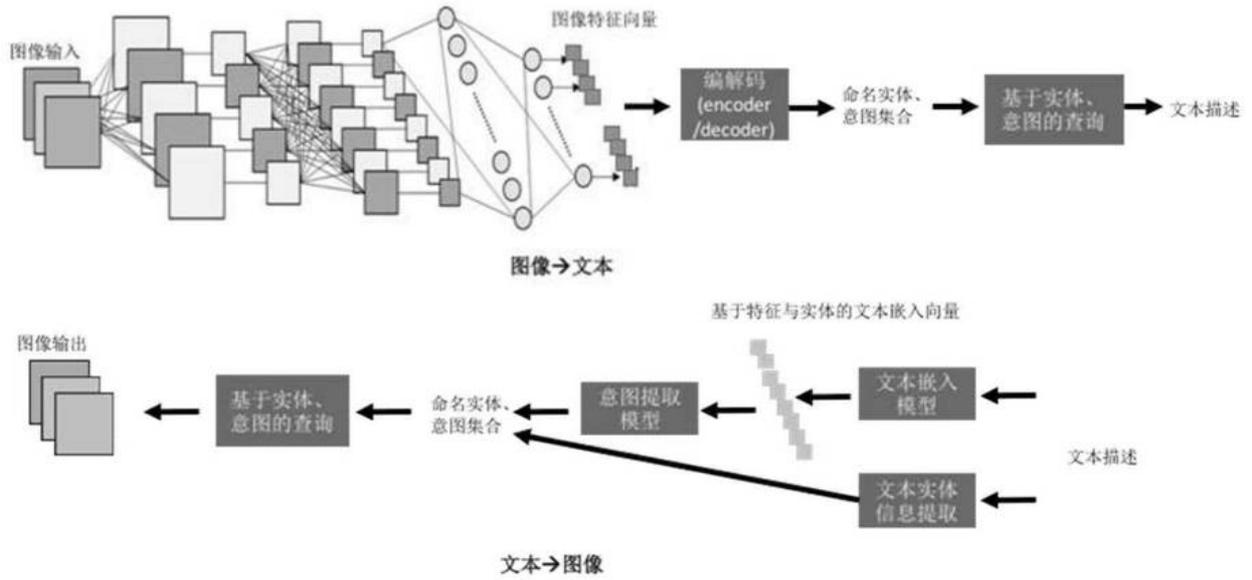


图1

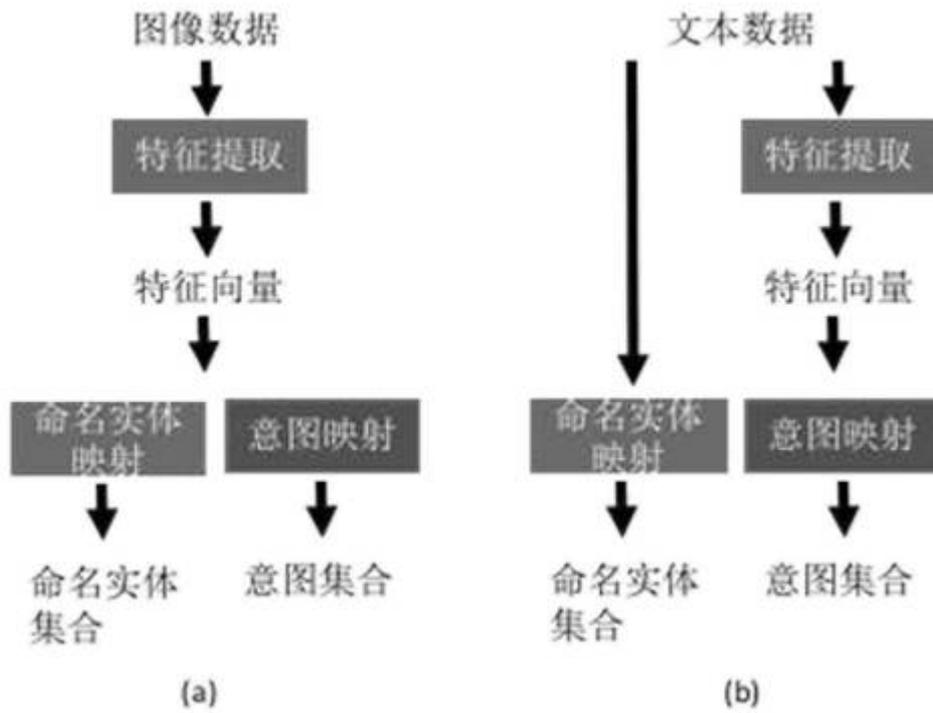


图2

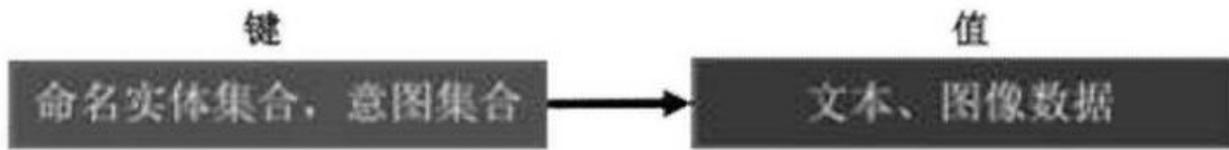


图3

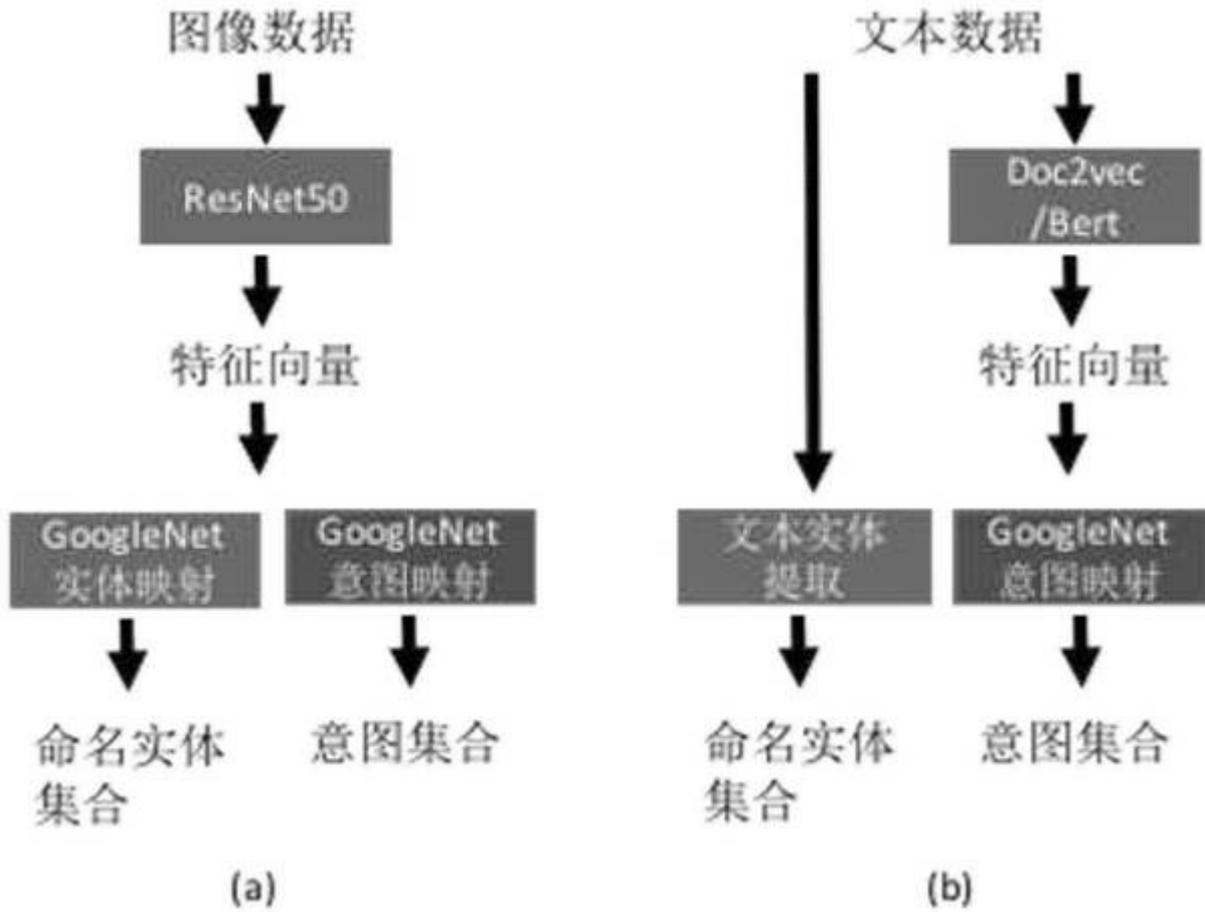


图4