



(12) 发明专利

(10) 授权公告号 CN 106980868 B

(45) 授权公告日 2022. 03. 11

(21) 申请号 201610995335.0
(22) 申请日 2016.11.11
(65) 同一申请的已公布的文献号
 申请公布号 CN 106980868 A
(43) 申请公布日 2017.07.25
(30) 优先权数据
 14/997,011 2016.01.15 US
(73) 专利权人 奥多比公司
 地址 美国加利福尼亚州
(72) 发明人 金海琳 任洲 林哲 方晨
(74) 专利代理机构 北京市金杜律师事务所
 11256
 代理人 鄢迅 辛鸣
(51) Int.Cl.
 G06K 9/62 (2022.01)
 G06V 10/774 (2022.01)
 G06V 10/778 (2022.01)
(56) 对比文件
 CN 104156433 A, 2014.11.19

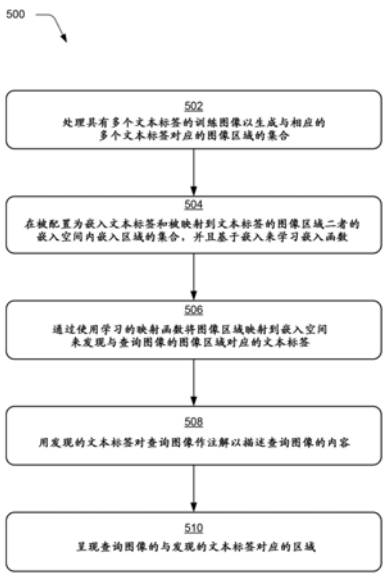
CN 102782678 A, 2012.11.14
CN 102057371 A, 2011.05.11
US 2015067471 A1, 2015.03.05
Liwei Wang 等. Learning Deep
Structure-Preserving Image-Text
Embeddings.《arXiv》.2015,
Andrej Karpathy 等. Deep Visual-
Semantic Alignments for Generating Image
Descriptions.《arXiv》.2015,
Shaoqing Ren 等. Faster R-CNN: Towards
Real-Time Object Detection with Region
Proposal Networks.《arXiv》.2015,
Jeffrey Pennington 等. GloVe: Global
Vectors for Word Representation.
《Proceedings of the 2014 Conference on
Empirical Methods in Natural Language
Processing (EMNLP)》.2014, (续)
审查员 郭妍妍

权利要求书3页 说明书19页 附图8页

(54) 发明名称
用于具有多个文本标签的图像的嵌入空间
(57) 摘要

描述用于具有多个文本标签的图像的嵌入空间。在嵌入空间中，嵌入了文本标签和图像区域二者。嵌入的文本标签描述可以在图像内容中被展示的语义概念。嵌入空间被训练以将嵌入的文本标签语义地相关，从而使得如“太阳”和“日落”之类的标签比“太阳”和“鸟”被更接近地相关。训练嵌入空间也包括将具有举例说明语义概念的图像内容的代表性图像映射到相应的文本标签。不同于对于与训练图像关联的每个文本标签将整个训练图像嵌入到嵌入空间中的常规技术，这里描述的技术处理训练图像以生成与多个文本标签对应的区域。训练图像的区域然后以将该区域映射到对应的文本标签的方式被嵌入到

训练空间中。



CN 106980868 B

[接上页]

(56) 对比文件

Philipp Krahenbuhl 等.Geodesic Object
Proposals.《European Conference on
Computer Vision (ECCV) 》.2014,

Mohammad Norouzi 等.Zero-Shot
Learning by Convex Combination of
Semantic Embeddings.《arXiv》.2014,

1. 一种由计算设备实施以用多个文本标签对个体图像作注解以描述所述图像的内容的方法,所述方法包括:

处理具有多个文本标签的训练图像以生成与相应的所述多个文本标签对应的图像区域的集合;

部分基于与所述训练图像的所述图像区域对应的所述多个文本标签在嵌入空间中嵌入的位置,在被配置为嵌入文本标签和被映射到所述文本标签的图像区域二者的所述嵌入空间内嵌入图像区域的所述集合;

学习将图像区域映射到在所述嵌入空间中嵌入的所述文本标签的映射函数,所述学习部分基于在所述嵌入空间内所述嵌入图像区域的所述集合;

确定针对查询图像的语义有意义图像区域的集合;

处理所述查询图像的语义有意义图像区域的所述集合以丢弃所述查询图像中的未满足至少一个预定义标准的所述语义有意义图像区域并且获取所述查询图像中的满足所述至少一个预定义标准的语义有意义图像区域的子集;

通过使用学习的所述映射函数将所述满足标准的图像区域映射到所述嵌入空间来发现与所述查询图像的满足标准的图像区域对应的所述嵌入空间的文本标签;

计算所述嵌入空间中的发现的所述文本标签和所述满足标准的图像区域之间的距离;

基于所计算的所述距离来对发现的所述文本标签进行排名;

至少部分地基于所述排名,用发现的所述文本标签中的至少两个文本标签对所述查询图像作注解。

2. 根据权利要求1所述的方法,其中发现的所述文本标签描述所述查询图像的所述图像内容。

3. 根据权利要求1所述的方法,其中处理所述训练图像以生成与相应的所述多个文本标签对应的图像区域的所述集合包括:

确定用于图像区域的所述集合的候选图像区域;以及

使用至少一种后处理技术来减少确定的所述候选图像区域的数目。

4. 根据权利要求3所述的方法,其中所述候选图像区域使用测地对象提议被确定。

5. 根据权利要求1所述的方法,其中所述至少一个预定义标准包括阈值大小,并且所述处理有效地丢弃所述查询图像中的具有少于所述阈值大小的所述语义有意义图像区域。

6. 根据权利要求1所述的方法,其中所述至少一个预定义标准包括可允许纵横比的预定义集合,并且所述处理有效地丢弃所述查询图像中的具有可允许纵横比的所述预定义集合以外的纵横比的所述语义有意义图像区域。

7. 根据权利要求3所述的方法,其中所述至少一种后处理技术使用单标签嵌入模型来向所述训练图像的所述多个文本标签中的每个文本标签指派单个候选图像区域。

8. 根据权利要求1所述的方法,还包括:

发现针对所述满足标准的图像区域的每个满足标准的图像区域的所述嵌入空间的多个文本标签;以及

针对给定的满足标准的图像区域:

计算针对所述给定的满足标准的图像区域发现的所述文本标签中的每个文本标签和所述嵌入空间中的所述满足标准的图像区域之间的距离;

针对所述给定的满足标准的图像区域发现的所述文本标签进行排名;以及
用最高排名文本标签来注释所述给定的满足标准的图像区域。

9. 根据权利要求1所述的方法, 处理所述训练图像以生成对应于相应的所述多个文本标签的图像区域的所述集合包括丢弃未满足所述至少一个预定义标准的图像区域。

10. 根据权利要求1所述的方法, 其中所述距离使用代表所述查询图像的相应的图像区域的矢量被计算, 所述矢量利用卷积神经网络CNN从所述查询图像的所述图像区域被提取。

11. 根据权利要求1所述的方法, 其中所述查询图像的语义有意义图像区域的所述集合是通过使用测地对象提议技术代替聚焦于前景对象检测的对象提议技术而被确定的。

12. 根据权利要求1所述的方法, 还包括呈现所述查询图像的与用来对所述查询图像作注解的发现的所述文本标签对应的所述图像区域。

13. 根据权利要求1所述的方法, 还包括训练所述嵌入空间以实现将所述图像区域映射到所述文本标签, 包括:

将文本词汇表的文本标签语义地相关以确定用于在所述嵌入空间中嵌入所述文本标签的位置;

处理各自具有多个文本标签的多个训练图像以生成与相应的所述多个文本标签对应的图像区域的集合; 以及

部分基于所述嵌入空间中的第一计算的距离与所述嵌入空间中的第二计算的距离的差值来在所述嵌入空间内嵌入图像区域的所述集合, 所述第一计算的距离在对图像区域的集合的个体图像区域的嵌入与所述个体区域被生成用于的相应的文本标签之间, 所述第二计算的距离在对所述个体区域的所述嵌入与负文本标签矢量的集合之间。

14. 根据权利要求13所述的方法, 其中所述文本词汇表的所述文本标签基于Glove模型被语义地相关。

15. 一种用于用多个文本标签对个体图像作注解以描述所述图像的内容的系统, 所述系统包括:

一个或者多个处理器; 以及

计算机可读存储介质, 所述计算机可读存储介质具有存储于其上的指令, 所述指令由所述一个或者多个处理器可执行以执行操作, 所述操作包括:

训练其中图像和文本标签二者被嵌入的嵌入空间, 所述训练将被配置为描述在图像内容中展示的语义概念的文本标签语义地相关并且将具有举例说明所述语义概念的图像内容的代表性图像映射到相应的文本标签;

基于所述训练来学习将图像区域映射到在所述嵌入空间中嵌入的所述文本标签的映射函数;

获得待注解的查询图像;

确定所述查询图像的语义有意义图像区域的集合;

处理所述查询图像的所述语义有意义图像区域的所述集合以将所述集合减少到所述查询图像中的满足至少一个预定义标准的语义有意义图像区域的子集;

在减少到所述子集之后, 使用所述映射函数将所述查询图像的满足标准的图像区域映射到所述嵌入空间中的对应的文本标签, 所述对应的文本标签描述在所述查询图像的所述满足标准的图像区域的图像内容中展示的语义概念; 以及

用所述对应的文本标签中的至少两个文本标签对所述查询图像作注解。

16. 根据权利要求15所述的系统,其中所述查询图像的语义有意义图像区域的所述子集的至少一个区域包括少于所述查询图像的全部。

17. 一种由计算设备实施以用多个文本标签对个体图像作注解以描述所述图像的内容的方法,所述方法包括:

训练其中图像和文本标签二者被嵌入的嵌入空间,所述训练将被配置为描述在图像内容中展示的语义概念的文本标签语义地相关并且将具有举例说明所述语义概念的图像内容的代表性图像映射到相应的文本标签;

确定针对查询图像的语义有意义图像区域的集合;

处理所述查询图像的语义有意义图像区域的所述集合以将所述集合减少到满足所述至少一个预定义标准的语义有意义图像区域的子集;

通过将所述满足标准的图像区域映射到训练的所述嵌入空间来发现训练的所述嵌入空间中与所述查询图像的满足标准的图像区域相对应的文本标签;以及

基于排名将至少两个发现的文本标签与所述查询图像关联,所述排名根据在所述嵌入空间中在发现的文本标签和所述查询图像的相应满足标准的图像区域之间的计算的距离,来相对于至少第三发现的文本标签来对所述至少两个发现的文本标签进行排名。

18. 根据权利要求17所述的方法,其中与所述查询图像关联的所述至少两个发现的文本标签之一用来将所述查询图像标识为对应于搜索。

19. 根据权利要求17所述的方法,其中对应于所述至少两个发现的文本标签的至少两个相应的满足标准的图像区域包括少于所述查询图像的全部。

用于具有多个文本标签的图像的嵌入空间

技术领域

[0001] 本申请的各实施例涉及用于具有多个文本标签的图像的嵌入空间。

背景技术

[0002] 图像分类是计算机视觉中的基本问题。广而言之，图像分类尝试从图像提取语义信息，从而使得图像可以被标注以描述图像的内容。语义信息可以例如包括在图像中描绘的物体(和图像的其中描绘物体的位置)、在图像中描绘的景象(例如，图像是否描绘海滩或者日落)、与在图像中描绘的人脸或者脸部表情关联的情绪、图像审美(例如，良好组成、不良组成、遵守三分法(rule-of-thirds)等)、图像感情(例如，恐惧、愤怒等)等等。

[0003] 一些常规图像分类技术通过训练多类分类器将图像分类成代表语义信息的固定类集合。然而，由于在类之间的语义关系可能复杂(例如，分级、不相交等)，所以难以定义对语义关系中的许多语义关系编码的分类器。为了解决这些缺点，已经开发了可视语义嵌入技术。常规可视语义嵌入技术利用来自无注解文本数据的语义信息来学习在文本标签之间的语义关系，并且将图像显式地映射到丰富语义嵌入空间中。然而，这些常规可视语义嵌入技术限于用单个文本标签对图像作注解。因而，用于自动地将文本标签与图像关联以描述它们的内容的常规技术对于一些图像标注任务是不够的。

发明内容

[0004] 描述了用于具有多个文本标签的图像的嵌入空间。在嵌入空间中，可以嵌入文本标签和图像区域二者。在嵌入空间中嵌入的文本标签被配置为描述在图像内容中展示的语义概念，例如，图像是否包括海滩或者日落、图像审美(比如图像是否根据摄影“规则”备良好地组成)等等。嵌入空间被训练以将嵌入的文本标签语义地相关，从而使得如“太阳”和“日落”的文本标签比如“太阳”和“鸟”的文本标签被更接近地相关。训练嵌入空间也包括将具有举例说明语义概念的图像内容的代表性图像映射到相应的文本标签。

[0005] 不同于常规可视语义嵌入技术，这里描述的技术使用各自与多个文本标签关联的训练图像来训练嵌入空间。并非对于与训练图像关联的多个文本标签中的每个文本标签将整个训练图像嵌入到嵌入空间中，这里描述的技术处理训练图像以生成与多个文本标签对应的区域。训练图像的区域然后以将区域映射到对应的文本标签的方式被嵌入到嵌入空间中。在训练图像的区域与对应的文本标签之间的映射被学习作为映射函数。一旦嵌入空间被训练，就可以通过应用映射函数来使用嵌入空间以用多个文本标签对图像作注解以描述图像的内容。为了对图像作注解，为图像提议了多个语义有意义区域。然后使用在训练期间学习的映射函数对于多个区域中的每个区域从嵌入空间发现至少一个文本标签。学习的映射函数取得待注解的图像的图像区域作为输入，并且将图像区域中的每个图像区域映射到嵌入空间中的文本标签。对于图像的多个区域而发现的文本标签然后可以用来对图像作注解。

[0006] 这一发明内容以简化形式介绍了以下在具体实施方式中被进一步描述的概念的

选集。这样,这一发明内容未旨在于标识要求保护的主体内容的实质特征、也未旨在于在确定要求保护的主体内容的范围时用作辅助。

附图说明

[0007] 参照附图描述具体实施方式。在各图中,标号的最左数字标识该标号在其中首次出现的图。在描述和各图中在不同实例中使用相同标号可以指示相似或者相同项目。在各图中代表的实体可以指示一个或者多个实体,并且因此可以在讨论中可互换地指代实体的单数或者复数形式。

[0008] 图1是示例实现方式中的数字环境的图示,该数字环境可操作用于运用这里描述的技术。

[0009] 图2是描绘具有多个文本标签的图像和图像的与多个文本标签对应的区域的指示的示例实现方式。

[0010] 图3是被配置为嵌入文本标签和被映射到文本标签的图像区域二者的嵌入空间的示例实现方式。

[0011] 图4是嵌入空间的另一示例实现方式,其中由文本标签描述的语义概念被建模为嵌入空间中的分布。

[0012] 图5是描绘示例实现方式中的过程的流程图,在该过程中,与多个文本标签关联的训练图像的区域被映射到嵌入空间中的对应的文本标签以训练嵌入空间,并且在该过程中,使用嵌入空间对于查询图像发现文本标签。

[0013] 图6是描绘另一示例实现方式中的过程的流程图,在该过程中,训练的嵌入空间被用来将多个文本标签与待注解的图像的相应的区域关联。

[0014] 图7是描绘另一示例实现方式中的过程的流程图,在该过程中,嵌入空间被生成以将由文本标签描述的语义概念建模为嵌入空间中的分布。

[0015] 图8图示了包括示例设备的各种部件的示例系统,该设备可以被运用于用于这里描述的用于具有多个文本标签的图像的嵌入空间的技术的一个或者多个实现方式。

具体实施方式

[0016] 概述

[0017] 常规可视语义嵌入技术利用来自无注解文本数据的语义信息以学习在文本标签之间的语义关系并且将图像显式地映射到丰富语义嵌入空间中。虽然这些常规可视语义嵌入技术用于对具有单个标签的图像的图像任务,但是这样的常规技术尚未被扩展用于对具有多个文本标签的图像使用。这些缺点使常规可视语义技术不适合用于多种应用,因为图像经常与多个文本标签关联。此外,将单标签可视语义模型扩展成多标签可视语义模型是不平凡任务。常规可视语义技术的实现方式部分基于图像标签与图像的全部对应的假设。然而,这对于与多个文本标签关联的图像完全不是这种情况。实际上,多标签图像的不同文本标签可以对应于图像的不同区域。

[0018] 用于具有多个文本标签的图像的嵌入空间被描述。如这里所用,“嵌入空间”是指能够代表已经在模型中嵌入的文本标签和图像的计算机生成的模型。嵌入空间也能够代表在嵌入的文本标签与图像之间的关系,比如在不同文本标签之间的语义关系和在图像与文

本标签之间的对应性。使用这里描述的技术,各自与多个文本标签关联的训练图像被用来训练嵌入空间。一旦嵌入空间被训练,嵌入空间就可以用来对图像作注解。具体而言,嵌入空间可以用来发现用于图像的多个文本标签。被发现的多个文本标签然后可以与图像关联(例如,通过与图像包括作为元数据或者以其它方式)以对图像作注解。

[0019] 再次关于这里使用的嵌入空间,嵌入空间被配代表联合图像-文本嵌入空间,该嵌入空间代表已经在嵌入空间中嵌入的文本标签和图像(例如,图像的区域)二者。具体而言,被嵌入的文本标签来自为了训练嵌入空间而被选择的文本词汇表,并且可以用来描述在图像中描绘的语义概念。文本标签也在嵌入空间中被语义地相关作为训练嵌入空间的部分。在嵌入空间中将文本标签语义地相关涉及对在文本标签之间的关系编码,这些关系指示由文本标签描述的语义概念如何接近地被相关。举例而言,文本词汇表的文本标签被语义地相关,从而使得如“太阳”和“日落”的文本标签比“太阳”和“鸟”的文本标签被更接近地相关。

[0020] 除了将文本标签语义地相关之外,还训练嵌入空间,从而使得举例说明由文本词汇表的文本标签描述的语义概念的图像区域被嵌入在嵌入空间中并且被映射到文本标签。从已经出于训练嵌入空间的目的而选择的并且各自与多个文本标签关联的训练图像提取被嵌入作为训练嵌入空间的部分的图像区域。在训练中,图像区域被映射到相应的文本标签,例如,描绘太阳的训练图像的区域可以被映射到嵌入空间中的文本标签“太阳”,而描绘鸟的训练图像的不同区域可以被映射到嵌入空间中的文本标签“鸟”。

[0021] 如以上指出的那样,用来训练嵌入空间的训练图像中的每个训练图像在被用来训练嵌入空间之前与多个文本标签关联,例如,文本标签与训练图像关联作为元数据。作为训练嵌入空间的部分,每个训练图像被处理以生成训练图像的与关联于训练图像的多个文本标签对应的区域的集合。为了生成区域的集合,根据区域提议技术(例如,测地对象提议)初始地处理训练图像,其结果是候选区域的集合。对于与训练图像关联的多个标签的每个个体标签,选择被确定与由文本标签描述的语义概念最佳地匹配的提议的候选区域。这对于形成用于训练图像的区域到标签对应性是有效的。另外,这一方式防止图像以将图像与最少地相关于图像的文本标签相关的方式被嵌入在嵌入空间中。

[0022] 在为训练图像的文本标签生成区域的集合之后,在嵌入空间中嵌入区域。关于在嵌入空间中“嵌入”区域,如这里所用,在嵌入空间中“嵌入”区域涉及提取图像区域的特征、将提取的特征转换成代表提取的数据的数据(例如,矢量)以及在嵌入空间中插入代表以使得区域由嵌入空间代表。图像区域的嵌入也涉及形成嵌入空间中的在插入的代表与对应的文本标签之间的关系。例如,基于与训练图像区域对应的多个文本标签被嵌入的位置来嵌入训练图像的区域。举例而言,基于文本标签“太阳”被嵌入的位置在嵌入空间中嵌入描绘太阳的图像区域。对于为训练图像而生成的区域的集合执行这一过程。

[0023] 嵌入空间一旦被训练就然后可以用来用多个文本标签对图像作注解。考虑如下示例,在该示例中接收尚未与文本标签关联的图像作为用于注解的输入。可以对于查询图像从训练的嵌入空间发现文本标签以描述查询图像的内容。为了从训练的嵌入空间发现文本标签,(例如,使用能够确定图像的语义有意义区域的一种或者多种已知技术,比如测地对象提议来)确定查询图像的语义有意义区域。一旦确定了查询图像的语义有意义区域,然后就例如通过比较指示嵌入的区域的可视特性的矢量和查询图像的确定的区域来在训练期

间比较语义有意义区域与在嵌入空间中嵌入的图像区域。基于比较,查询图像的确定的区域可以被映射到可视地相似的嵌入的图像区域,并且因此也被映射到相似的嵌入的图像区域对应于的文本标签。

[0024] 然后用在这一过程中发现的多个(例如,两个或者更多个)文本标签对查询图像作注解。例如,可以用预定数目的如下发现的标签对查询图像作注解,对于这些发现的标签,在查询图像的确定的区域与嵌入的区域之间的相似性最大,例如,用于查询图像的与嵌入的区域最相似的五个确定的区域的文本标签。备选地或者附加地,可以用如下发现的文本标签中的每个发现的文本标签对查询图像作注解,对于这些发现的文本标签,在查询图像的确定的区域与嵌入的区域之间的相似性超过阈值,例如,与嵌入的区域至少百分之八十五相似的任何确定的区域的文本标签。

[0025] 不同于常规可视语义嵌入技术,这里描述的技术实现使用具有多个文本标签的训练图像来训练联合图像-文本嵌入空间,并且实现使用联合图像-文本嵌入空间用多个文本标签对图像作注解。这里描述的技术也实现对于与图像关联的多个文本标签中的每个文本标签发现语义有意义区域。附加地,这里描述的技术在零射(zero shot)学习中提供较常规技术而言的优点,例如,在嵌入空间的文本标签先前没有描述的新类别中对图像分类。

[0026] 在以下讨论中,首先描述可以运用这里描述的技术的示例环境。然后描述可以在示例环境以及其它环境中执行的示例实施细节和过程。因而,对示例过程的执行不限于示例环境并且示例环境不限于执行示例过程。

[0027] 示例环境

[0028] 图1是示例实现方式中的环境100的图示,环境100可操作用于运用这里描述的技术。图示的环境100包括计算设备102,计算设备102具有包括一个或者多个处理设备(例如,处理器)的处理系统104和一个或者多个计算机可读存储介质106。图示的环境100也包括在计算机可读存储介质106上体现并且经由处理系统104可操作用于实施这里描述的对应的功能的图像数据108、文本标签110、多实例嵌入模型112、多实例嵌入模块114(“MIE模块114”)和基于区域的多标签图像116。在至少一些实现方式中,计算设备102包括如下进一步具体描述的用于访问各种种类的基于web的资源(内容和服务)、与在线提供方交互等等的功能。

[0029] 计算设备102可被配置为任何适当类型的计算设备。例如,计算设备102可以被配置为服务器、台式计算机、膝上型计算机、移动设备(例如,假设手持配置,比如平板计算机或者移动电话)、平板计算机、被配置为接收手势输入的设备、被配置为接收三维(3D)手势作为输入的设备、被配置为接收语音输入的设备、被配置为接收基于触笔的输入的设备、被配置为接收那些输入的组合作为输入的设备等等。因此,计算设备102可以范围从具有大量存储器和处理器资源的全资源设备(例如,服务器、个人计算机、游戏控制台)到具有有限存储器和/或处理资源的低资源设备(例如,移动设备)。附加地,虽然示出了单个计算设备102,但是计算设备102可以代表用于如关于图8进一步描述的那样“在云之上”执行操作的多个不同设备。

[0030] 环境100还描绘被配置为通过网络120(比如因特网)与计算设备102通信以提供“基于云”的计算环境的一个或者多个服务提供方118。一般而言,服务提供方118被配置为使各种资源122通过网络120可用于客户端。在一些场景中,用户签注被运用以访问来自提

供方的对应的资源的账户。提供方在批准对账户和对应的资源122的访问之前认证用户的证书(例如,用户名和口令)。使其它资源122自由地可用(例如,无认证或者基于账户的访问)。资源122可以包括通常地由一个或者多个提供方通过网络使其可用的服务和/或内容的任何适当组合。服务的一些示例包括但不限于图像存储和/或共享服务(例如, **Flickr®**)、社交网络服务(例如, **Facebook®**、**Twitter®**、**Instagram®**、**Hyperlapse®**)、能够返回图像作为搜索结果的搜索引擎服务等等。

[0031] 这些服务用作为大量图像数据的来源。图像数据108代表可以在包括但不限于JPEG、TIFF、RAW、GIF、BMP、PNG等的多种图像格式中的任何图像格式中被格式化的这样的图像数据。通过服务使其可用的图像数据可以由具有与那些服务的账户的用户发表。例如,具有与图像存储和/或共享服务的账户的用户上传图像,比如利用用户的数字相机拍摄的图像,或者经由电子手段向用户发送的图像。图像存储和/或共享服务的用户然后可以比如通过提供指向图像的相册或者指向用户的简档的链接来与其他人共享他们的上传的图像。

[0032] 附加地,图像能够与附加信息关联以描述图像。例如,由图像数据108代表的图像可以包括描述图像的元数据。举例而言,图像的元数据可以描述图像的按像素的大小、图像的存储大小、用来捕获图像的相机设备、捕获图像的日期、上次修改图像的日期等等。附加地,由图像数据108代表的图像可以与描述在图像中描绘的内容的元数据关联。例如,元数据可以被配置为文本标签,并且可以用描述图像的内容的那些文本标签对图像作注解。这样的文本标签可以描述在图像中描绘的对象(以及图像的描绘对象的位置)、在图像中描绘的景象(例如,图像是否描绘海滩或者日落)、与在图像中描绘的人脸或者脸部表情关联的情绪、图像审美(例如,良好组成、不良组成、遵守三分法等)、图像感情(例如,恐惧、愤怒等)等等。广而言之,文本标签110代表可以用来对图像作注解的多种文本标签。

[0033] MIE模块114代表用来实施如这里描述的用于具有多个文本标签的图像的嵌入空间的技术的功能。作为这样做的部分,MIE模块114创建被配置为联合图像-文本嵌入空间的嵌入空间。“联合图像-文本嵌入空间”意味着可以在其中嵌入文本标签110和来自图像数据108的代表性图像(例如,图像的区域)二者。多实例嵌入模型112代表这样的嵌入空间。另外,MIE模块114代表如下功能,该功能用于训练多实例嵌入模型112,从而使得在图像区域给定时,多实例嵌入模型112可以用来发现文本标签,该文本标签描述在图像区域中描绘的内容。

[0034] 为了训练多实例嵌入模型112,MIE模块114根据由文本标签描述的语义概念来将在多实例嵌入模型112中嵌入的文本标签语义地相关。例如,MIE模块114将文本标签语义地相关,从而使得文本标签“太阳”和“日落”在多实例嵌入模型112中比文本标签“太阳”和“鸟”被更接近地定位。

[0035] MIE模块114也使用可以作为图像数据108的部分而被包括的训练图像来训练多实例嵌入模型112。训练图像各自在训练之前已经与描述描绘的内容的多个文本标签关联。对于每个训练图像,MIE模块114处理图像以生成图像的与训练图像的多个文本标签对应的区域的集合。因此,MIE模块114形成用于每个训练图像的区域到标签对应性,从而使得训练图像的区域被选择以代表多个文本标签中的每个文本标签。

[0036] MIE模块114然后将区域的集合嵌入到多实例嵌入模型112中。具体而言,MIE模块

114基于区域的对应的标签被嵌入的位置来将区域嵌入到多实例嵌入模型112中。例如,在训练图像与文本标签“太阳”关联时,选择训练图像的与“太阳”对应的区域。MIE模块114然后至少部分地基于多实例嵌入模型112中的其中嵌入文本标签“太阳”的位置来嵌入训练图像的与“太阳”对应的区域。MIE模块114可以对于训练图像的集合中的每个训练图像执行这一过程。通过用训练图像训练多实例嵌入模型112,学习映射函数。在图像(例如,图像区域)给定时,映射函数能够将给定的图像映射到多实例嵌入模型112中的位置。然后确定给定的图像对应于在多实例嵌入模型112中嵌入的与给定的图像被映射到的位置最近的文本标签。

[0037] 在一个或者多个实现方式中,多实例嵌入模型112被配置为将由文本标签描述的和由代表性图像描绘的语义概念建模为分布,例如,高斯分布、高斯混合等等。这与将语义概念代表为单个点的技术形成对照。将与多个文本标签关联的图像嵌入到可视语义嵌入空间中的单个点可能使可视语义映射函数被容易地混淆,例如,使技术在预测在图像中描绘的内容时更不准确。

[0038] 附加地,MIE模块114表示用于对图像作注解的功能。“注解”意味着描述在图像中描绘的内容的文本标签可以被选择并且与图像关联。为了对图像作注解,MIE模块114确定图像的区域集合。具体而言,MIE模块114例如使用已知技术(比如测地对象提议,因为它覆盖前景区域和背景区域二者)来从图像提取被确定为语义有意义的区域。MIE模块114然后通过使用学习的映射函数将区域映射到文本标签来发现用于区域的文本标签。使用映射函数,MIE模块114可以确定在多实例嵌入模型112中嵌入的图像中的哪些图像与待注解的区域相似,并且关联与相似的嵌入的图像对应的文本标签与待注解的区域。

[0039] 以这一方式,图像可以初始地与大量文本标签关联,例如,每个文本标签用于确定的区域中的一个区域。然而,可以减少文本标签的数目,从而使得用被确定为最代表图像内容的标签对图像作注解。例如,MIE模块114可以用预定数目的文本标签对图像作注解。在这一场景中,MIE模块114选择预定数目的如下文本标签,对于这些文本标签,图像的区域与在多实例嵌入模型112中嵌入的图像最相似。备选地,MIE模块114选择用于待注解的图像的可变数目的文本标签。在这一备选场景中,选择的文本标签的数目可以基于阈值相似度,从而使得对于图像的任何如下区域(对于这些区域,与嵌入的图像的相似度在阈值相似度以上),发现的文本标签用来对图像作注解。在任何情况下,基于区域的多标签图像116代表MIE模块114通过确定图像区域和使用多实例嵌入模型112以发现用于图像区域的标签来用多个文本标签作注解的图像。

[0040] 在一个或者多个实现方式中,MIE模块114可实施为软件模块、硬件设备或者使用软件、硬件、固件、固定逻辑电路装置等的组合可实施。另外,MIE模块114可以可实施为如图所示的计算设备102的单独部件。附加地或者备选地,MIE模块114可以被配置为web服务的部件、应用、计算设备102的操作系统、插件模块或者如关于图8进一步描述的其它设备应用。

[0041] 已经考虑了示例环境,现在考虑对根据一个或者多个实现方式的用于具有多个文本标签的图像的嵌入空间的技术的一些示例细节的讨论。

[0042] 用于有多个文本标签的图像的嵌入空间

[0043] 这一节描述根据一个或者多个实现方式的用于针对具有多个文本标签的图像的

嵌入空间的技术的一些示例细节。该讨论参照图2,图2图示了具有多个文本标签的图像并且图示了与这里描述的功能有关的概念。

[0044] 具体而言,图2大体地在200处描绘了具有文本标签204、206、208、210、212的图像202。图2也描绘了图像区域214、216、218、220。图像202可以代表训练图像,该训练图像已经与文本标签204、206、208、210、212关联,并且对于该训练图像,MIE模块114选择图像区域214、216、218、220与文本标签对应,例如,图像区域214与文本标签208对应、图像区域216与文本标签206对应、图像区域218与文本标签204对应、图像区域220与文本标签210对应并且代表图像202的全部的图像区域与文本标签212对应。备选地,图像202可以代表由MIE模块114作注解的图像,从而使得MIE模块114生成用于图像202的提出的区域、使用多实例嵌入模型112来发现用于提出的区域的文本标签并且从发现的文本标签选择与图像区域214、216、218、220(以及与图像的全部对应的区域)对应的文本标签204、206、208、210、212以用于描述图像202的内容。

[0045] 图3大体地在300处描绘了其中文本标签和被映射到文本标签的图像区域二者可以被嵌入的示例嵌入空间。具体而言,图3包括联合图像-文本嵌入空间302(其可以对应于图1的多实例嵌入模型112)、图像数据库304(其包括由MIE模块114用来训练联合图像-文本嵌入空间302的图像)和文本词汇表306(其可以对应于图1的可用来描述在图像中描绘的语义概念的文本标签110)。

[0046] 联合图像-文本嵌入空间302被配置为让文本标签和图像(全部图像和图像的部分)二者嵌入在其中。图像数据库304和文本词汇表306分别代表图像和文本标签,MIE模块114使用这些图像和文本标签以训练联合图像-文本嵌入空间302,例如,从而使得联合图像-文本嵌入空间302可以用来发现描述图像的内容的文本标签。组成文本词汇表306的文本标签被配置为描述可以在图像内容中展示的语义概念。图像数据库304的图像描绘举例说明由文本词汇表306的文本标签描述的语义概念的图像内容。

[0047] 在训练中,MIE模块114将在文本词汇表306内的文本标签语义地相关。例如,取得用于“树木”的文本标签308。MIE模块114表示如下功能,其用于例如通过在联合图像-文本嵌入空间302中的位置处嵌入文本标签308、310、312来将用于“树木”的文本标签308语义地相关到用于“女人”的文本标签310和用于“建筑”的文本标签312。文本标签在联合图像-文本嵌入空间302中被嵌入的位置依赖于由文本标签描述的语义概念多么接近地被相关。举例而言,用于“树木”的文本标签308可以被嵌入在联合图像-文本嵌入空间302中,从而使得它被定位得比与用于“女人”的文本标签310和用于“建筑”的文本标签312更接近用于“松树”、“森林”和“植物”(未示出)的文本标签。

[0048] 在图3中,在表示相应的文本标签的填入形状旁边图示了文本标签308、310、312中的每个文本标签,例如,在填入星号旁边图示了用于“树木”的文本标签308,在填入十字形旁边图示了用于“女人”的文本标签310,并且在填入X旁边图示了用于“建筑”的文本标签312。也在联合图像-文本嵌入空间302中包括这些填入形状中的每个填入形状以代表对文本标签308、310、312的嵌入。应当认识到联合图像-文本嵌入空间302的描绘的形状和填入形状的描绘的位置可能没有准确地反映联合图像-文本嵌入空间302的实际形式或者MIE模块114实际地嵌入文本标签308、310、312的位置。实际上,联合图像-文本嵌入空间302的描绘的形状和填入形状的描绘的位置对于说明这里描述的技术有用。

[0049] MIE模块114也将图像数据库304中的图像相关到文本词汇表306的文本标签。如以上提到的那样,图像数据库304的训练图像可以描绘举例说明由文本词汇表306的文本标签描述的语义概念的图像内容。为了将图像相关到文本标签,MIE模块114部分基于对应的文本标签的位置在联合图像-文本嵌入空间302中的位置处嵌入每个图像。换言之,MIE模块114将每个图像映射到描述图像的图像内容的文本标签,并且基于该文本标签的位置嵌入图像。

[0050] 例如,图像数据库304的描绘树木的图像基于用于“树木”的文本标签308被嵌入的位置而被嵌入在联合图像-文本嵌入空间302中,例如,MIE模块114将描绘树木的图像映射到用于“树木”的文本标签308。以相似方式,描绘女人的图像基于用于“女人”的文本标签310被嵌入的位置被嵌入,并且描绘建筑的图像基于用于“建筑”的文本标签312被嵌入的位置被嵌入。换言之,图像在比与其它文本标签更接近图像对应于的文本标签的位置被嵌入在联合图像-文本嵌入空间302中。图像也被嵌入在联合图像-文本嵌入空间302中比与更不接近地相关的语义概念对应的图像更接近与接近地相关的语义概念对应的图像。举例而言,树木的图像一般地比与建筑的图像或者妇女的图像更接近松树和森林的图像被嵌入在联合图像-文本嵌入空间302中。

[0051] 在图3中,图像集合314代表图像数据库304的描绘树木的图像,图像集合316代表图像数据库304的描绘妇女的图像,并且图像集合318代表图像数据库304的描绘建筑的图像。用表示图像的未填充形状图示每个图像集合314、316、318的图像,例如,用未填充星形图示图像集合314的图像(其描绘树木),用未填充十字形图示图像集合316的图像(其描绘女人),并且用未填充“X”图示图像集合318的图像(其描绘建筑)。

[0052] 也在联合图像-文本嵌入空间302中包括这些未填充形状以代表对来自图像集合314、316、318的图像的嵌入,这些图像的位置基于对应的文本标签308、310、312被嵌入的位置。因而,在填入星形周围图示了代表描绘树木的图像的未填充星形,这些未填充星形代表用于“树木”的文本标签308,在填入十字周围图示了代表描绘女人的图像的未填充十字,这些未填充十字代表用于“女人”的文本标签310,并且在填入X周围图示了代表描绘建筑的图像的未填充X,这些未填充X代表用于“建筑”的文本标签312。

[0053] 关于在联合图像-文本嵌入空间302中嵌入的代表性图像,MIE模块114从训练图像提取它们作为区域。因此,描绘女人的图像的集合314可以对应于从训练图像提取的区域(例如,图像的少于全部的部分)。广而言之,为了训练联合图像-文本嵌入空间302,MIE模块114从各自与多个文本标签关联的训练图像开始。举例而言,MIE模块114从训练图像(比如已经与文本标签204、206、208、210、212关联的图像202)开始。

[0054] 对于每个训练图像,MIE模块114生成与训练图像的文本标签对应的区域的集合。考虑图2的示例,在该示例中,图像202与文本标签204、206、208、210、212关联。在这一示例中,MIE模块114可以生成与文本标签204、206、208、210、212中的每个文本标签对应的区域,从而使得图像区域214对应于用于“女人”的文本标签208,图像区域216对应于用于“埃菲尔铁塔”的文本标签206,图像区域218对应于用于“包”的文本标签204,图像区域220对应于用于“建筑”的文本标签210,并且代表图像202的全部的图像区域对应于用于“都市风景”的文本标签212。

[0055] 为了生成与训练图像的文本标签对应的区域,MIE模块114初始地将区域提议技术

应用于训练图像。例如，MIE模块114应用测地对象提议技术、随机区域提议技术等中的至少一种技术。图像的语义有意义区域并非仅包括在图像的前景中描绘的对象。实际上，图像的背景部分也可以语义有意义。与一般对象提议技术形成对照，由MIE模块114使用的区域提议技术覆盖图像的前景区域和背景区域二者以提议图像区域。

[0056] 一旦为训练图像提议了候选区域的集合，MIE模块114就丢弃提议的候选区域中的至少一些提议的候选区域。MIE模块114使用的由对象提议技术最初提议的候选区域的数量可以是数千个，并且使用提议的候选区域中的每个提议的候选区域用于训练将使计算资源有大量负担。因而，MIE模块114可以应用一种或者多种后处理技术以丢弃提出的候选区域中的至少一些提出的候选区域。例如，MIE模块114可以应用一种或者多种后处理技术以丢弃未能满足预定义的大小标准或者纵横比标准的提议的候选区域。举例而言而非限制，预定义的大小标准可以指定为了使用提议的候选区域，提议的候选区域的边将是图像的对应的边的长度的至少百分之三十，例如，对于400像素×500像素的图像，MIE模块114通过丢弃具有小于120像素×150像素的大小的任何提出的候选区域来实行标准。关于示例纵横比，预定义的纵横比标准可以指定为了使用提出的候选区域，提出的候选区域的纵横比将在1:4与4:1之间。

[0057] 从剩余的提出的候选区域，MIE模块114然后向训练图像的每个文本标签联合地指派最佳匹配的候选区域以构建区域到标签对应性。在这样做时，MIE模块114被配置为使用单标签嵌入模型。这具有丢弃甚至更大数目的提出的候选区域以使得剩余区域的数量对应于与图像关联的文本标签的数量的效果。MIE模块114然后将这些区域嵌入到联合图像-文本嵌入空间302中。参照图2中的示例并且在其中图像202用来训练联合图像-文本嵌入空间302的场景中，MIE模块可以将由区域提议技术提议的数千个候选区域减少成图像区域214、216、218、220以及与图像202的全部对应的区域。

[0058] 为了在联合图像-文本嵌入空间302中嵌入训练图像的区域，MIE模块114计算用于区域的d维特征矢量。MIE模块114可以例如通过用卷积神经网络(CNN)(例如，包括卷积、汇合(pooling)和起始层的CNN)提取图像特征来计算用于区域的特征矢量。在一个具体示例中，MIE模块114用快速的基于区域的CNN(RCNN)技术来提取图像特征。换言之，在训练图像和将在联合图像-文本嵌入空间302中嵌入的区域给定时，MIE模块114单次通过CNN传递训练图像。MIE模块114然后可以将生成的区域的集合的每个区域汇合成固定大小的特征映射以获得最终特征矢量。

[0059] 考虑如下示例，在该示例中，MIE模块114使用各自与多个文本标签关联的训练图像的集合来训练联合图像-文本嵌入空间302。这样的训练图像的集合可以被定义为 $D \equiv \{(x_i, y_i)\}_{i=1}^N$ ，其中 x_i 代表集合中的第i个图像，并且N代表集合中的训练图像的数量。

另外，每个图像可以由可以如以上讨论的那样计算的d维特征矢量 $\mathbb{R}^d$ 代表，从而使得对于集合的第i个图像， $x_i \in \mathbb{R}^d$ 。另外，每个训练图像与文本标签的集合关联，从而使得对于第i个图像 x_i ，文本标签的集合 y_i 被定义为 $(y_i^1, \dots, y_i^t)$ ，其中t代表与第i个图像关联的文本标签的数量。应当注意到，用于文本标签的集合的这一定义允许与训练图像关联的标签数目从图像到图像变化。

[0060] 为了简化,待嵌入的训练图像的集合可以被表示为 $X \stackrel{\text{def}}{=} \mathbb{R}^d$ 。常规技术将用于与多个文本标签关联的图像的分类或者注解问题用公式表示为多类分类。与这里描述的技术形成对照,这一常规公式表示预定义固定的类标签的集合(比如

$y_i^1 \in Y \equiv \{1, \dots, m\}$), 学习 m 向分类器或者 m 个一比全部型分类器,并且使用一个或者多个学习的分类器以将训练图像的集合 X 分类成固定的类标签的集合 Y ,从而使得 $(X \rightarrow Y)$ 。然而,这样的常规技术的可伸缩性就标签的数目而言有限制,因为常规技术在新标签显现时重新训练它们利用的模型。

[0061] 取代如这些常规技术所做的那样将文本标签定义为离散集合,可视语义模型学习连续语义空间 S ,该连续语义空间 S 捕获在文本标签之间的语义关系并且显式地学习从训练图像的集合 X 到连续语义空间 S 的映射函数 $(X \rightarrow S)$ 。因此,MIE模块114被配置为构造图像标签训练数据对 $\{(x_i, s(y_i))\}$ 以学习用于联合图像-文本嵌入空间302的映射函数 $f: X \rightarrow S$ 。MIE模块114学习映射函数以便将训练图像的集合的第 i 个图像 x_i 映射到联合图像-文本嵌入空间302。MIE模块114将这学习为与第 i 个图像 x_i 关联的文本标签的集合 y_i 的函数 $s(y_i)$ 。函数 $s(y_i)$ 被学习,从而使得使用的标签的集合 Y 被映射到连续语义空间 S ,从而使得 $s: Y \rightarrow S$ 。

[0062] 在一个或者多个实现方式中,MIE模块114被配置为使用未注解文本数据(例如,来自因特网的未注解文本数据)训练的文本建模架构技术来学习连续语义标签空间 S (例如,在文本标签之间的关系)。文本建模架构使MIE模块114能够学习用于语义相关的文本标签的相似嵌入位置,因为同义词更可能出现在相似语义情境中。举例而言,MIE模块114使用Glove模型以构造将在文本标签之间的语义关系具体化的300维文本标签空间 S 。

[0063] 关于学习映射函数,一些常规可视语义方式用 L_2 损耗函数或者排行损耗函数学习用于嵌入空间的映射函数 $f: X \rightarrow S$ 。一般而言,这样的函数鼓励用于给定的图像的投影的嵌入点更接近它在嵌入空间中的对应的标签。以下是由常规可视语义方式使用的这样的排行损耗函数的示例,具体而言,它对应于按对铰接排行损耗函数:

$$[0064] \quad l_{\text{rank}}(x_i, y_i) = \sum_{j \in \tau_+} \sum_{k \in \tau_-} \max(0, m + D_{f(x_i), y_i} - D_{f(x_i), y_k})$$

[0065] 这里,项 $f(\cdot)$ 代表待学习的映射函数,项 m 代表裕度。为了符号表示简化,使用 y 取代 $s(y)$ 以表示连续语义标签空间 S 中的标签矢量。另外, τ_+ 表示正文本标签的集合, τ_- 表示负文本标签的集合, $\{y_j\}_{j \in \tau_+}$ 表示正标签矢量,并且 $\{y_k\}_{k \in \tau_-}$ 表示负标签矢量。项 $D_{f(x_i), y_i}$ 指示在用于第 i 个图像 x_i 的嵌入矢量 $f(x_i)$ 与用于图像的标签的矢量 y_i 之间的距离。

[0066] 然而,使用以上指出的按对铰接排行函数的常规可视语义方式具有缺点。例如,虽然每个图像 x_i 可以对应于文本标签的集合 y_i 中的多个标签,但是那些标签的一对或者多对可以被嵌入在连续语义标签空间 S 中的相互相对地远离的位置中。因此,推动对整个图像 $f(x_i)$ 的嵌入,从而使得第 i 个图像 x_i 被嵌入在嵌入空间中的与多个文本标签(这些文本标签被嵌入在相互相对地远离的位置)接近的位置可能引起映射函数的误差。附加地或者备选地,这样做可能造成将图像 x_i 映射到标签矢量的平均位置,从而使得平均位置对应于与由个体文本标签描述的概念不同的语义概念。

[0067] 与常规可视语义嵌入技术形成对照,MIE模块114被配置为在一个或者多个实现方

式中使用以下损耗函数以学习用于联合图像-文本嵌入空间302的多实例嵌入:

$$[0068] \quad l_{\text{MIE}}(x_i, y_i) = \sum_{j \in \tau_+} \sum_{k \in \tau_-} \max \left(0, m + \min_{c \in C} D_{f(x_i^c), y_i} - \min_{c \in C} D_{f(x_i^c), y_k} \right)$$

[0069] 这里,项 x_i^c 代表第 i 个图像 x_i 的特定区域 c ,例如,用以上描述的方式生成的与关联于第 i 个图像 x_i 的文本标签之一对应的区域。项 C 代表为第 i 个图像 x_i 生成的区域的集合。

另外, $\{y_j\}_{j \in \tau_+}$ 表示正标签矢量, $\{y_k\}_{k \in \tau_-}$ 表示负标签矢量, 并且 $D_{f(x_i^c), y_j}$ 表示在对图像区域 $f(x_i^c)$ 的嵌入之间与该图像区域被生成用于的对应的文本标签 y_j 的距离。在一个或者多个实现方式中, MIE模块114将距离计算为平方欧几里得距离。另外并且如用于学习多实例嵌入的损耗等式中所示, MIE模块114可以通过对从标签到区域的集合的每个区域的距离应用最小值运算 (例如, $\min$ 项) 对区域到标签的对应性建模。在这样做时, MIE模块114选择训练图像的至少一个区域以说明由训练图像的每个文本标签描述的语义概念。附加地, MIE模块114更可能为文本标签选择在嵌入空间中具有与文本标签最接近的嵌入位置的区域。

[0070] 在一个或者多个实现方式中, MIE模块114优化对用于训练图像的区域预测的文本标签的排行。为了这样做, MIE模块114使用调整的多实例嵌入损耗公式, 该公式鼓励正文本标签 (例如, 与训练图像关联的文本标签) 具有比大多数负文本标签更小的 $\min$ 距离。以这一方式, MIE模块114将正文本标签排行在用于训练图像的给定的区域的预测的文本标签的列表的顶部。换言之, MIE模块114惩罚将正文本标签排行在列表的底部的错误预测。举例而言, MIE模块114可以被配置为根据以下调整的公式优化对预测的文本标签的排行:

$$[0071] \quad \widehat{l}_{\text{MIE}}(x_i, y_i) = \sum_{j \in \tau_+} \sum_{k \in \tau_-} w(r_j) \cdot \max \left(0, m + \min_{c \in C} D_{f(x_i^c), y_i} - \min_{c \in C} D_{f(x_i^c), y_j} \right)$$

[0072] 这里, 项 $w(\cdot)$ 是权值函数, 并且项 r_j 是正文本标签在用于训练图像的给定的区域的预测的文本标签的列表中的排行。在一个或者多个实现方式中, MIE模块114根据以下公式确定项 r_j :

$$[0073] \quad r_j = \sum_{t \neq j} \mathbb{1} \left(\min_{c \in C} D_{f(x_i^c), y_t} \leq \min_{c \in C} D_{f(x_i^c), y_k} \right)$$

[0074] 这里, $\mathbb{1}(\cdot)$ 是指示符函数, MIE模块114对于第 i 个图像 x_i 使用该指示符函数以根据预测的文本标签 y_t (例如, 在联合图像-文本嵌入空间302中) 与区域的集合 C 中的每个区域的距离来对它排行。具体而言, MIE模块114将排行计算为由项 $\min_{c \in C} D_{f(x_i^c), y_t}$ 代表的与每个区域的最小距离的函数。另外, MIE模块114可以被配置为使用定义如下的权值函数:

$$[0075] \quad w(r) = \begin{cases} 1 & \text{如果 } r < \#(\tau_+), \\ r & \text{其它} \end{cases}$$

[0076] 这里, 项 $\#(\tau_+)$ 代表与第 i 个图像 x_i 关联的文本标签的数目。因此, 如果与第 i 个图像 x_i 关联的文本标签之一是预测列表中的前 $\#(\tau_+)$ 个文本标签之一, 则MIE模块114通过向权值指派更小小数来惩罚损耗。例如, 如果第 i 个图像 x_i 与五个文本标签关联并且关联的文本标签

之一是为图像而预测的文本标签的列表中的前五个排行的标签之一,则MIE模块114通过向权值指派更小数来惩罚计算出的损耗。对照而言,如果与第 i 个图像 x_i 关联的标签之一不是预测列表中的前 $\#(\tau_+)$ 个文本标签之一,则MIE模块114通过向权值指派更大数来惩罚损耗。

[0077] 继续其中第 i 个图像 x_i 与五个文本标签关联的示例,考虑如下场景,在该场景中,关联的文本标签之一不是为图像而预测的文本标签的列表中的前五个排行的标签之一。在这一情况中,MIE模块114通过向权值指派更大数来惩罚计算出的损耗。通过以这一方式向损耗函数应用惩罚,MIE模块114将语义有意义标签映射得更接近它们在联合图像-文本嵌入空间302中的对应的标签。在训练图像(例如,在图像数据库304中包括的训练图像)的区域被映射到联合图像-文本嵌入空间302中的文本标签(例如,在文本词汇表306中包括的文本标签)之后,它可以由MIE模块114用来发现用于未注解图像的文本标签。

[0078] 考虑如下示例,其中例如结合对图像编索引以用于搜索来向MIE模块114给定待注解的图像 x' 。以与为训练图像生成区域的相似方式,MIE模块114生成用于待注解的图像 x' 的区域的集合 C' ,例如,MIE模块114将区域提议技术(例如,测地对象提议或者随机区域提议)应用于待注解的图像,然后将后处理技术应用于提议的区域以减少用于待注解的图像的提议的区域的数目。

[0079] MIE模块114然后通过联合图像-文本嵌入空间302传递待注解的图像和从其生成的区域的集合以获得用于区域中的每个区域的嵌入矢量。在以上讨论的映射函数给定时,MIE模块114计算用于待注解的图像的区域的集合的嵌入矢量,从而使得为给定的区域 c 获得嵌入矢量 $f(x'^c)$ 。然后对于联合图像-文本嵌入空间302中的文本标签 y' 中的任何文本标签,MIE模块114通过 $\min_{c \in C'} D_f(x'^c, y')$ 计算在待注解的图像 x' 与联合图像-文本嵌入空间302中的文本标签 y' 的位置之间的距离。因此,MIE模块114为待注解的图像的每个区域例如基于在为区域确定的嵌入矢量与联合图像-文本嵌入空间302中的文本标签的位置之间的距离来确定区域在联合图像-文本嵌入空间302中最接近的文本标签。

[0080] MIE模块114然后对为待注解的图像确定的文本标签排行。具体而言,MIE模块114根据在区域的嵌入矢量与相应的文本标签之间的距离对确定的文本标签排行,从而使得与相应的区域的距离更少的文本标签被更有利地排行。因而,与相应的区域的距离最小的确定的文本标签是最高排行的文本标签。使用排行,MIE模块114生成被确定为与待注解的图像的区域对应的文本标签的排行的列表。MIE模块114然后可以用来自列表的预定数目的标签对待注解的图像作注解。举例而言而非限制,MIE模块114可以用来自与区域对应的文本标签的列表的前五个排行的文本标签对待注解的图像作注解。无论预定数目如何,MIE模块114都被配置为使用联合图像-文本嵌入空间302以用多个文本标签对图像作注解。

[0081] 如以上简要地描述的那样,联合图像-文本嵌入空间302也被配置为将由文本标签描述的并且代表性图像被映射到的语义概念建模为分布,例如,高斯分布、高斯混合等。

[0082] 图4在400处描绘了嵌入空间的示例,该嵌入空间将由文本标签描述的语义概念代表为分布。具体而言,示例400描绘了包括分布404、406、408的联合图像-文本嵌入空间402。分布404、406、408可以例如分别代表树木、苹果和水果这些语义概念。应当注意到,在分布404、406、408之间有一些重叠,该重叠代表由文本标签代表的语义概念可以重叠。换言之,由用于“树木”、“苹果”和“水果”的文本标签代表的语义概念可以重叠,部分因为苹果可以对应于树木并且也可以对应于水果。因而,语义概念被建模为联合图像-文本嵌入空间

402中的分布而不是被建模为单个点。

[0083] MIE模块114学习联合图像-文本嵌入空间402以通过对由文本标签描述的每个语义概念建模来生成它。举例而言,MIE模块114将语义概念中的每个语义概念建模为高斯分布如下:

$$[0084] \quad t_k \sim N(\mu_k, \Sigma_k), \forall k \in (1, \dots, M)$$

[0085] 这里,项M代表用来训练联合图像-文本嵌入空间402的文本词汇表中的文本标签的数目。项 t_k 代表文本词汇表中的第k个文本标签t,项 μ_k 代表第k个文本标签t的矢量(例如,从用于文本标签代表的Glove模型获得的文本标签的Glove特征),项 Σ_k 代表被初始化到联合图像-文本嵌入空间402的单位矩阵,并且 $N(\cdot)$ 代表分布建模函数(例如,标准高斯分布,但是可以使用其它分布)。在一个或者多个实现方式中,MIE模块114通过深度神经网络来学习用于联合图像-文本嵌入空间402的分布。MIE模块114可以通过卷积神经网络(CNN)来代表训练图像的集合中的给定的图像I,并且可以根据损耗函数(例如,排行损耗)来初始化用于图像的图像嵌入权值。关于项 Σ_k ,MIE模块114通过最小化用于训练图像的集合中的每个图像I的损耗来如下学习它:

$$[0086] \quad loss(I) = \sum_{\substack{n=1 \\ n \neq i}}^M \max(0, m + (I - \mu_i)^T \mu_i^{-1} (I - \mu_i) - (I - \mu_n)^T \mu_n^{-1} (I - \mu_n))$$

[0087] 参照这一等式,MIE模块114被配置为将用于给定的图像I的正文本标签 t_i 建模为 $N(\mu_i, \Sigma_i)$,并且将用于给定的图像的负文本标签 t_n 建模为 $N(\mu_n, \Sigma_n)$ 。如在以上等式中那样,m代表裕度项。

[0088] 虽然主要地参照高斯分布描述了将语义概念建模为嵌入空间中的分布的实现方式,但是应当认识到,MIE模块114被配置为使用多种其它分布对由嵌入空间中的文本标签描述的语义概念建模而未脱离这里描述的技术的精神实质或者范围。举例而言而非限制,MIE模块114也能够将由嵌入空间中的文本标签描述的语义概念建模为高斯混合、对数正态分布、Pareto分布等等。

[0089] 通过将由文本标签描述的语义概念建模为分布而不是单个点,这里描述的技术使语义概念能够重叠。这造成生成嵌入空间,该嵌入空间比将语义概念建模为单个点的嵌入空间更准确地代表在语义概念之间的关系,例如,因为许多语义概念重叠。例如,太阳和日落这些语义概念重叠,但是太阳和日出这些语义概念也是这样。在任何情况下,MIE模块114被配置为生成将由文本词汇表的文本标签描述的语义概念建模为分布的嵌入空间。

[0090] 已经讨论了用于具有多个文本标签的图像的嵌入空间的技术的示例细节,现在考虑一些示例过程以举例说明技术的附加方面。

[0091] 示例过程

[0092] 这一节描述了一个或者多个实现方式中的用于针对具有多个文本标签的图像的嵌入空间的示例过程。可以在硬件、固件或者软件或者它们的组合中实施过程的方面。过程被示出为指定由一个或者多个设备执行的操作的块的集合而未必限于所示的用于由相应的块执行操作的顺序。在至少一些实现方式中,过程由适当地配置的设备(比如图1的利用MIE模块114的示例计算设备102)执行。

[0093] 图5描绘了示例过程500,在该过程中,与多个文本标签关联的训练图像的区域被映射到嵌入空间中的对应的文本标签以训练嵌入空间,并且在该过程中,使用嵌入空间为查询图像发现文本标签。与多个文本标签关联的训练图像被处理以生成与相应的多个文本标签对应的图像区域的集合(块502)。

[0094] 例如,假设图像202对应于训练图像并且已经与文本标签204、206、208、210、212关联。在这一场景中,MIE模块114处理图像202以生成图像区域214、216、218、220以对应于文本标签,从而使得图像区域214被生成以对应于文本标签208,图像区域216被生成以对应于文本标签206,图像区域218被生成以对应于文本标签204,图像区域220被生成以对应于文本标签210,并且代表图像202的全部的图像区域被生成以对应于文本标签212。如以上更具体讨论的那样,MIE模块114通过应用区域提议技术并且对提议的候选区域执行后处理技术来生成图像区域。MIE模块114也通过例如基于在区域将在嵌入空间中被嵌入的位置与嵌入空间中的文本标签的位置之间的距离选择用于文本标签204、206、208、210、212中的每个文本标签的最佳匹配区域来生成对应的区域。

[0095] 在被配置为嵌入文本标签和被映射到文本标签的图像区域二者的嵌入空间内嵌入区域的集合(块504)。另外,至少部分基于与训练图像的图像区域对应的多个文本标签在嵌入空间中被嵌入的位置来嵌入区域的集合。并且对区域集合的嵌入与嵌入其它训练图像的区域的集合一起对学习映射函数有效。举例而言,MIE模块114在联合图像-文本嵌入空间302中嵌入图像区域214、216、218、220和对应于图像202的全部的图像区域。MIE模块114也至少部分基于文本标签204、206、208、210、212在联合图像-文本嵌入空间302中被嵌入的位置来这样做。

[0096] 如以上更具体讨论的那样,MIE模块114例如通过用卷积神经网络(CNN)(比如快速的基于区域的CNN(RCNN)技术)提取图像特征来计算用于图像区域214、216、218、220和对应于图像202的全部的图像区域中的每个图像区域的特征矢量,以在联合图像-文本嵌入空间302中嵌入它们。附加地,在MIE模块114在联合图像-文本嵌入空间302中嵌入图像区域214、216、218、220和对应于图像202的全部的图像区域时,MIE模块114将区域映射到对应的文本标签以学习映射函数,例如,从而使得无文本标签的图像和图像区域可以被映射到联合图像-文本嵌入空间302中的位置。

[0097] 一旦嵌入空间被训练,学习的映射函数就可用来从用于图像的嵌入空间发现文本标签以例如用于图像加标记、用于基于多文本的图像搜索(以将图像标识为对应于搜索)、图像加字幕等等。使用训练的嵌入空间来发现与查询图像的图像区域对应的文本标签(块506)。为了发现文本标签,向学习的映射函数中输入图像区域,该映射函数将图像区域映射到嵌入空间的文本标签。

[0098] 现在假设图像202对应于查询图像并且文本标签204、206、208、210、212尚未与图像202关联。在这一场景中,图像区域214、216、218、220和对应于图像202的全部的图像区域到联合图像-文本嵌入空间302中的文本标签可以用作向学习的映射函数的输入。

[0099] 根据学习的映射函数将图像区域214、216、218、220和对应于图像202的全部的图像区域映射到联合图像-文本嵌入空间302允许执行与在联合图像-文本嵌入空间302中嵌入的图像区域的比较。在一个或者多个实现方式中,通过比较为图像区域214、216、218、220和对应于图像202的全部的图像区域提取的特征矢量与在联合图像-文本嵌入空间302中嵌

入的图像区域的特征矢量来作为训练的部分执行比较。MIE模块114向图像区域214、216、218、220和对应于图像202的全部的图像区域指派相似的嵌入的图像区域被映射到的嵌入的文本标签。以这一方式，MIE模块114为图像区域214、216、218、220和对应于图像202的全部的图像区域发现联合图像-文本嵌入空间302中的文本标签204、206、208、210、212。

[0100] 用发现的文本标签对查询图像作注解以描述查询图像的内容(块508)。举例而言，MIE模块114例如通过与图像202一起包括文本标签204、206、208、210、212作为元数据或者以某种其它方式将它们与图像202关联来用它们对图像202作注解。

[0101] 呈现查询图像的与文本标签对应的区域(块510)。举例而言，接收用于呈现图像202的与用来对图像202作注解的文本标签对应的区域的用户请求。MIE模块114比如在图像202的显示之上呈现图像区域214、216、218、220和对应于图像202的全部的图像区域中的一个或者多个图像区域的指示(例如，定界框)。应当认识到，MIE模块114能够以多种不同方式呈现图像区域的指示而未脱离这里描述的技术的精神实质或者范围。举例而言而非限制，可以与对应的文本标签一起呈现图像区域的指示，可以可视地改变区域以表现得不同于图像的其它部分(例如，在使图像的其它部分模糊之时聚焦)等等。

[0102] 图6描绘了示例过程600，在该过程中，使用训练的嵌入空间以将多个文本标签与待注解的图像的相应的区域关联。训练其中图像和文本二者被嵌入的嵌入空间(块602)。训练涉及将被配置为描述在图像内容中展示的语义概念的文本标签语义地相关以及将具有举例说明语义概念的图像内容的代表性图像映射到相应的文本标签。

[0103] 举例而言，MIE模块114训练联合图像-文本嵌入空间302。为了这样做，MIE模块114例如通过利用在因特网上可用的文本数据以学习可伸缩和按词典分布的字词的代表以在文本词汇表306的文本标签之中捕获语义含义来将文本词汇表306的文本标签语义地相关。在一个或者多个实现方式中，MIE模块114利用一种或者多种文本建模架构技术以这样做，比如word2vec模型、Glove模型等等。

[0104] 继续这一示例，MIE模块114通过将图像数据库304的代表性图像映射到联合图像-文本嵌入空间302中的文本词汇表306的相应的文本标签来训练联合图像-文本嵌入空间302。例如，MIE模块114在联合图像-文本嵌入空间302中将图像集合314的图像映射到用于“树木”的文本标签308、将图像集合316的图像映射到用于“女人”的文本标签310并且将图像集合318的图像映射到用于“建筑”的文本标签312。作为训练的结果，MIE模块114学习用于将图像区域映射到联合图像-文本嵌入空间302中的文本标签的映射函数。

[0105] 一旦联合图像-文本嵌入空间302被训练，学习的映射函数就可以用来对图像作注解。待注解的图像被获得(块604)。举例而言，图像202被获得。在这一示例中，假设图像202未被作注解并且尚未与文本标签204、206、208、210、212关联。

[0106] 图像的区域集合被生成(块606)。在一个或者多个实现方式中，生成涉及使用区域提议技术，该区域提议技术提议能够被映射到嵌入空间中的对应的文本标签的区域。举例而言，MIE模块114生成用于图像202的区域集合，该区域的集合包括图像区域214、216、218、220和对应于图像202的全部的图像区域。MIE模块114用以上更具体描述的方式生成用于图像202的区域集合。生成的区域在语义上有意义并且能够被映射到嵌入空间中的嵌入的文本标签。

[0107] 向学习的映射函数中输入区域的集合以将区域的集合映射到嵌入空间中的描述

在区域的集合中展示的语义概念的对应的文本标签(块608)。举例而言,MIE模块114使用通过在块602处训练联合图像-文本嵌入空间302学习的映射函数以在块606处发现用于为图像202生成的区域的集合的文本标签。具体而言,MIE模块114使用学习的映射函数以将区域的集合映射到在联合图像-文本嵌入空间302中嵌入的描述在图像区域214、216、218、220和对应于图像202的全部的图像区域中的每个图像区域中展示的语义概念的文本标签。具体而言,MIE模块114使用学习的映射函数以将描绘女人的图像区域214映射到用于“女人”的文本标签208、将描绘艾尔菲铁塔的图像区域216映射到用于“埃菲尔铁塔”的文本标签206、将描绘购物袋的图像区域218映射到用于“袋”的文本标签204、将描绘建筑的图像区域220映射到用于“建筑”的文本标签210并且将对应于图像202的全部的图像区域映射到用于“都市风景”的文本标签212。

[0108] 用对应的文本标签中的至少两个文本标签对图像作注解(块610)。例如,MIE模块114基于在块608处的映射用文本标签204、206、208、210、212对图像202注解作。如以上指出的那样,比图2中描绘的图像区域更大数目的图像区域可以被生成并且被映射到联合图像-文本嵌入空间302中。作为结果,可以生成与图像202对应的文本标签的列表。然而,可以用来自列表的选择的一组文本标签对图像作注解。举例而言而非限制,可以如以上更具体描述的那样对文本标签的列表排行,并且MIE模块114选择预定数目的最高排行的标签以用于图像202的注解。

[0109] 除了使用排行的列表之外,还可以用其它方式选择为对图像作注解选择的标签。例如,可以基于对应的区域与在联合图像-文本嵌入空间302中嵌入的代表性图像的阈值差值来选择标签。为了这样做,MIE模块114以例如通过提取图像区域的特征矢量和嵌入的代表性图像并且比较特征矢量来找到在为图像生成的每个区域与联合图像-文本嵌入空间302中的嵌入的代表性图像之间的距离(例如,欧几里得距离)。MIE模块114然后可以选择如下文本标签,对于这些文本标签,图像区域具有与嵌入的代表性图像的少于阈值差值的差值。MIE模块114可以用更多其它方式从被确定为与图像的区域对应的文本标签选择用于对图像作注解的文本标签而未脱离这里描述的技术的精神实质或者范围。

[0110] 图7描绘了示例过程700,在该过程中,嵌入空间被生成以将由文本标签描述的语义概念建模为嵌入空间中的分布。其中可以嵌入图像和文本标签二者的嵌入空间被生成,从而使得由文本词汇表中的文本标签描述的语义概念被建模为嵌入空间中的分布(块702)。对嵌入空间的生成也包括通过在嵌入空间中嵌入代表性图像来训练嵌入空间,从而使得代表图像性映射到相应分布。

[0111] 举例而言,MIE模块114通过将来自文本词汇表的文本标签描述的语义概念建模为分布(例如,高斯分布、高斯混合等等)来生成联合图像文本嵌入空间402。MIE模块114生成联合-图像文本嵌入空间402以相对于被建模为分布的语义概念来实现图像和文本标签嵌入。另外,MIE模块114通过为语义概念形成的相应的分布内嵌入具有举例说明语义概念的图像内容的图像区域来训练联合-图像文本嵌入空间402。

[0112] 将语义概念建模为分布的嵌入空间被用来对图像作注解(块704)。举例而言,MIE模块114使用联合-图像文本嵌入空间402以用描述图像202的内容的文本标签对图像202作注解。例如,MIE模块114使用联合-图像文本嵌入空间402以发现文本标签204、206、208、210、212,并且然后用发现的标签对图像202作注解。

[0113] 已经描述了根据一个或者多个实现方式的示例过程,现在考虑可以用来实施这里描述的各种技术的示例系统和设备。

[0114] 示例系统和设备

[0115] 图8大体地在800处图示了示例系统,该系统包括代表实施这里描述的各种技术的一个或者多个计算系统和/或设备的示例计算设备802。通过包括MIE模块114来图示这一点,MIE模块114如以上描述的那样操作。计算设备802可以例如是服务器提供方的服务器、与客户端关联的设备(例如,客户端设备)、片上系统和/或任何其它适当计算设备或者计算系统。

[0116] 示例计算设备802包括相互通信地耦合的处理系统804、一个或者多个计算机可读介质806和一个或者多个I/O接口808。虽然未示出,但是计算设备802还可以包括相互耦合各种部件的系统总线或者其它数据和命令传送系统。系统总线可以包括不同总线结构(比如存储器总线或者存储器控制器、外围总线、通用串行总线和/或利用多种总线架构中的任何总线架构的处理器或者本地总线)中的任何总线结构或者组合。也设想了多种其它示例,比如控制和数据线。

[0117] 处理系统804代表用于使用硬件来执行一个或者多个操作的功能。因而,处理系统804被图示为包括可以被配置为处理器、功能块等的硬件单元810。这包括在硬件中被实施为专用集成电路或者使用一个或者多个半导体形成的其它逻辑器件。硬件单元810不受形成它们的材料或者其中运用的处理机制所限制。例如,处理器可以由半导体和/或晶体管(例如,电子集成电路(IC))组成。在这样的上下文中,处理器可执行指令可以是电子地可执行的指令。

[0118] 计算机可读存储介质806被图示为包括存储器/存储装置812。存储器/存储装置812代表与一个或者多个计算机可读介质关联的存储器/存储容量。存储器/存储装置812可以包括易失性介质(比如随机存取存储器(RAM))和/或非易失性介质(比如只读存储器(ROM)、闪存、光盘、磁盘等)。存储器/存储装置812可以包括固定介质(例如,RAM、ROM、固定硬驱动等)以及可拆卸介质(例如,闪存、可拆卸硬驱动、光盘等)。可以用如以下进一步描述的多种其它方式配置计算机可读介质806。

[0119] 输入/输出接口808代表用于允许用户向计算设备802录入命令和信息并且也允许向用户呈现信息的功能和/或使用各种输入/输出设备的其它部件或者设备。输入设备的示例包括键盘、光标控制设备(例如,鼠标)、麦克风、扫描仪、触摸功能(例如,被配置为检测物理触摸的电容或者其它传感器)、相机(例如,运用可见光不可见光波长(比如红外线频率)以将移动识别为不涉及触摸的手势的相机)等。输出设备的示例包括显示设备(例如,监视器或者投影仪)、扬声器、打印机、网卡、触觉响应设备等等。因此,可以用如以下进一步描述的多种方式配置计算设备802以支持用户交互。

[0120] 这里在软件、硬件单元或者程序模块的一般上下文中描述了各种技术。一般而言,这样的模块包括执行特定任务或者实施特定抽象数据类型的例程、程序、对象、单元、部件、数据结构等。如这里所用的术语“模块”、“功能”和“部件”一般地代表软件、固件、硬件或者它们的组合。这里描述的技术的特征独立于平台,这意味着可以在具有多种处理器的多种商用计算平台上实施技术。

[0121] 可以在某个形式的计算机可读介质上存储或者跨越该计算机可读介质传输描述

的模块和技术的实现方式。计算机可读介质可以包括计算设备802可以访问的多种介质。举例而言而非限制,计算机可读介质包括“计算机可读存储介质”和“计算机可读信号介质”。

[0122] “计算机可读存储介质”是指与仅信号传输、载波或者信号本身对照而言实现信息的持久和/或非瞬态存储的介质和/或设备。因此,计算机可读存储介质没有包括信号本身或者信号承载介质。计算机可读存储介质包括硬件(比如在适合用于存储信息、比如计算机可读指令、数据结构、程序模块、逻辑元件/电路或者其它数据)的方法或者技术中实施的硬件,比如易失性和非易失性、可拆卸和非可拆卸介质和/或存储设备。计算机可读存储介质的示例可以包括但不限于RAM、ROM、EEPROM、闪存或者其它存储器技术、CD-ROM、数字万用盘(DVD)或者其它光存储装置、硬盘、磁盒、磁带、磁盘存储装置或者其它磁存储设备或者适合用来存储希望的信息并且可以由计算机访问的其它存储设备、有形介质或者制造品。

[0123] “计算机可读信号介质”是指被配置为比如经由网络向计算设备802的硬件传输指令的信号承载介质。信号介质通常地在调制的数据信号(比如载波、数据信号)或者其它传送机制中体现计算机可读指令、数据结构、程序模块或者其它数据。信号介质也包括任何信息递送介质。术语“调制的数据信号”意味着如下信号,该信号让它的特性中的一个或者多个特性以对信号中的信息编码这样的方式被设置或者改变。举例而言而非限制,通信介质包括有线介质(比如有线网络或者直接有线连接)和无线介质(比如声学、RF、红外线和其它无线介质)。

[0124] 如先前描述的那样,硬件单元810和计算机可读介质806代表以如下硬件形式实施的模块、可编程逻辑器件和/或固定器件逻辑,可以在一些实现方式中运用该硬件形式以实施这里描述的技术的至少一些方面,比如执行一个或者多个指令。硬件可以包括集成电路或者片上系统的部件、专用集成电路(ASIC)、现场可编程门阵列(FPGA)、复杂可编程逻辑器件(CPLD)和在硅或者其它硬件中的其它实现方式。在本文中,硬件作为处理设备操作,该处理设备执行由指令定义的程序任务和/或由硬件体现的逻辑以及用来存储用于执行的指令的硬件,例如,先前描述的计算机可读存储介质。

[0125] 也可以运用前述各项的组合以实施这里描述的各种技术。因而,软件、硬件或者可执行模块被实施为在某个形式的计算机可读存储介质上和/或由一个或者多个硬件单元810体现的一个或者多个指令和/或逻辑。计算设备802可以被配置为实施与软件和/或硬件模块对应的特定指令和/或功能。因而,可以至少部分地在硬件中(例如,通过使用计算机可读存储介质和/或处理系统804的硬件单元810)来实现作为软件可由计算设备802执行的模块的实现方式。指令和/或功能由一个或者多个制造品(例如,一个或者多个计算设备802和/或处理系统804)可执行/可操作以实施这里描述的技术、模块和示例。

[0126] 这里描述的技术由计算设备802的各种配置支持而限于这里描述的技术的具体示例。也可以全部或者部分通过使用分布式系统(比如如以下描述的那样经由平台816通过“云”814)来实施这一功能。

[0127] 云814包括和/或代表用于资源818的平台816。平台816将云814的硬件(例如,服务器)和软件资源的下层功能抽象化。资源818可以包括可以在从计算设备802远离的服务器上执行计算机处理之时被利用的应用和/或数据。资源818也可以包括通过因特网和/或通过订户网络(比如蜂窝或者Wi-Fi网络)提供的服务。

[0128] 平台816将用于连接计算设备802与其它计算设备的资源和功能抽象化。平台816

也服务于将资源的伸缩抽象化以提供与对于经由平台816被实施的资源818的所需需求对应的规模级别。因而，在一个互连设备实施例中，贯穿系统800分布这里描述的功能的实现方式。例如，部分在计算设备802上以及经由将云814的功能抽象化的平台816实施功能。

[0129] 结论

[0130] 虽然已经用结构特征和/或方法动作特有的言语描述了本发明，但是将理解，在所附权利要求中定义的本发明未必限于描述的具体特征或者动作。实际上，具体特征和动作被公开作为实施要求包含的本发明的示例形式。

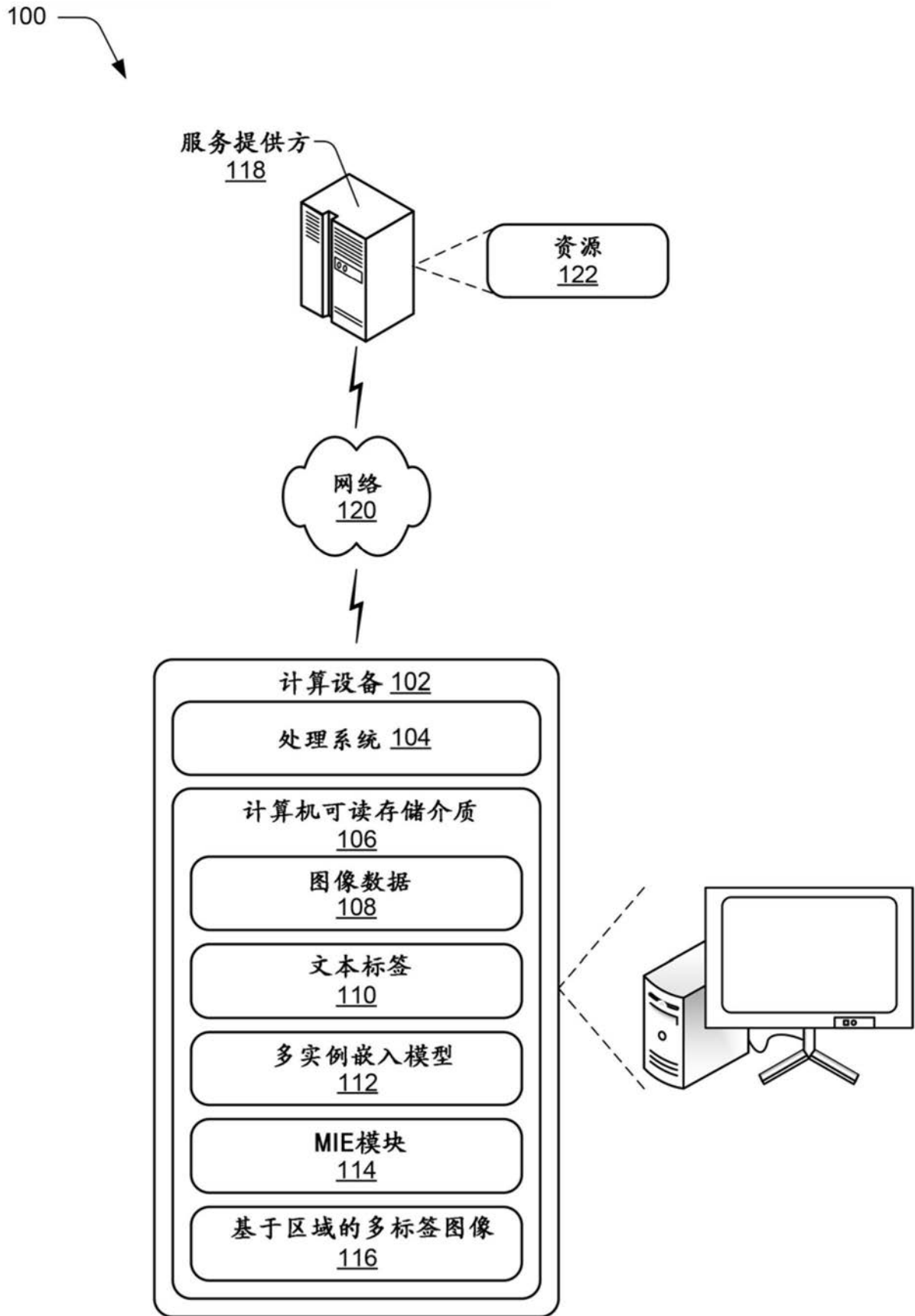


图1

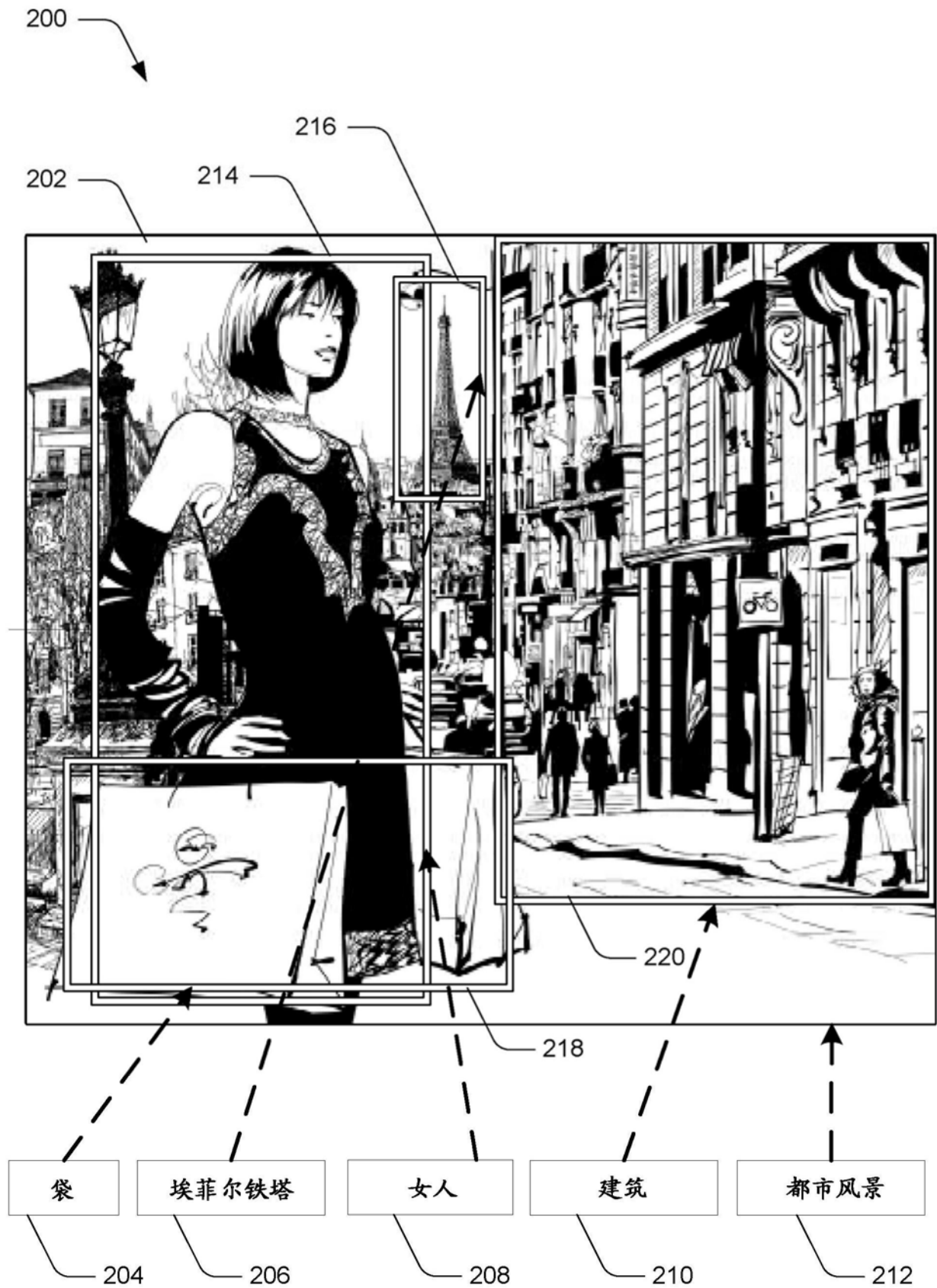


图2

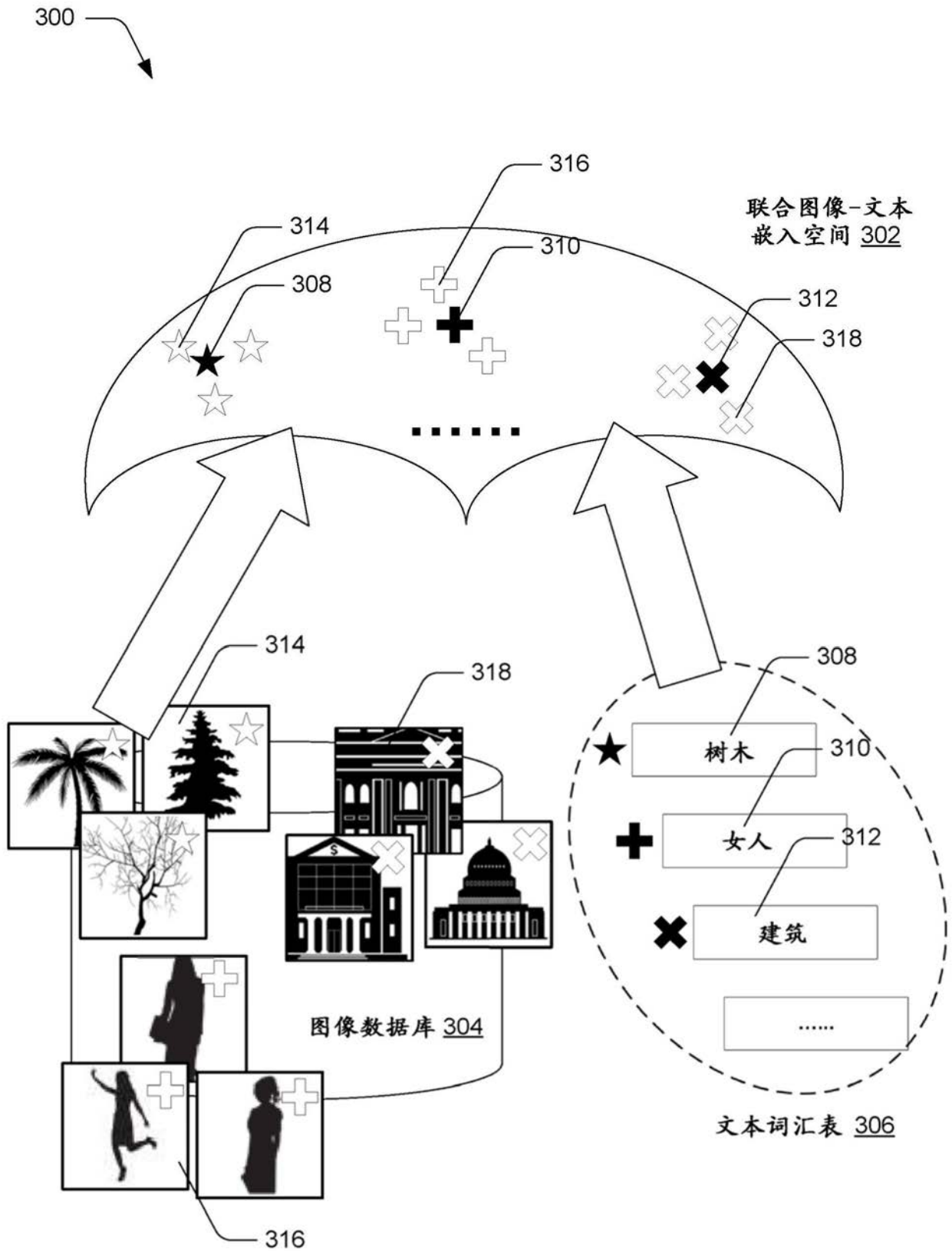


图3

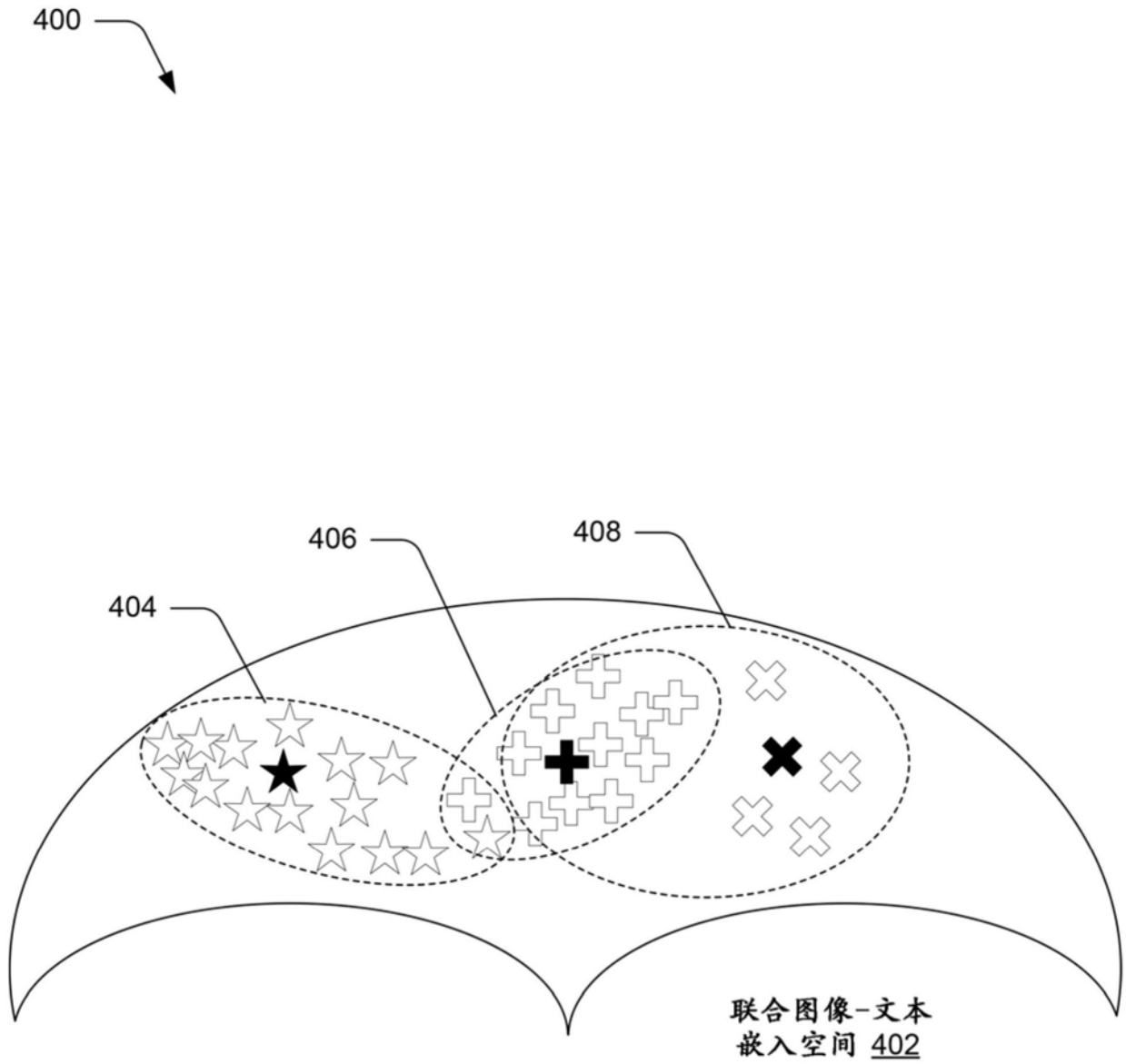


图4

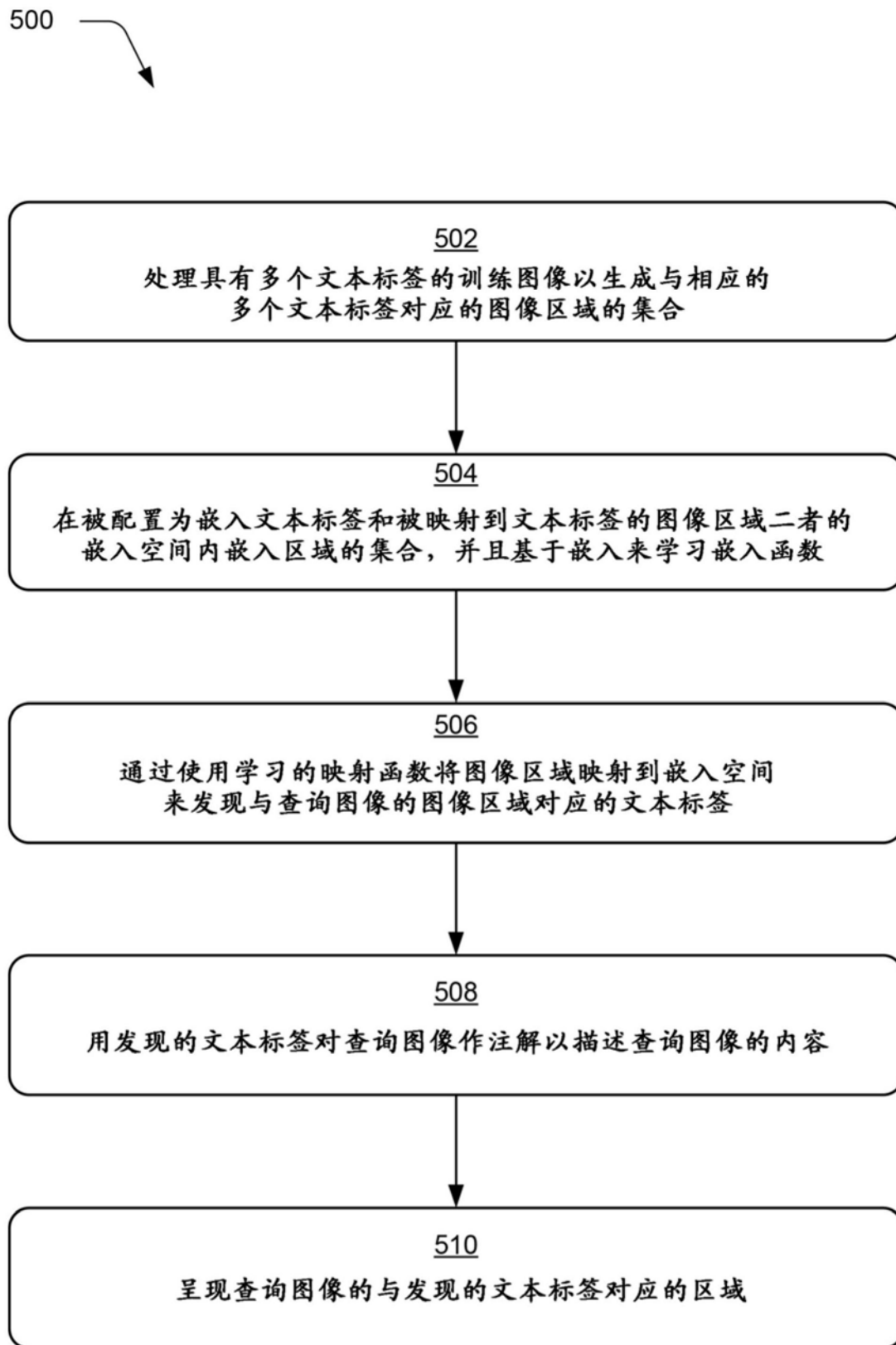


图5

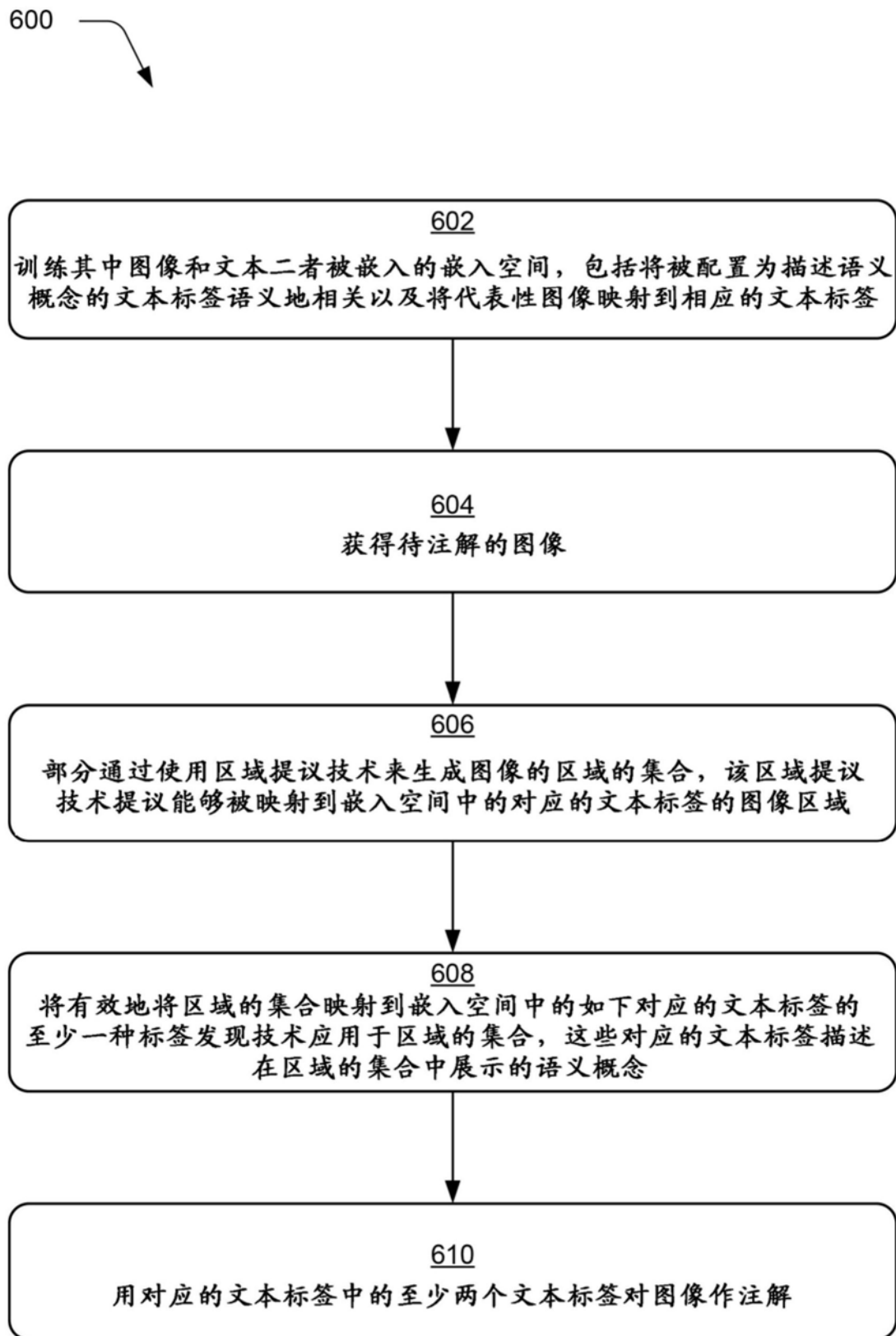


图6

700

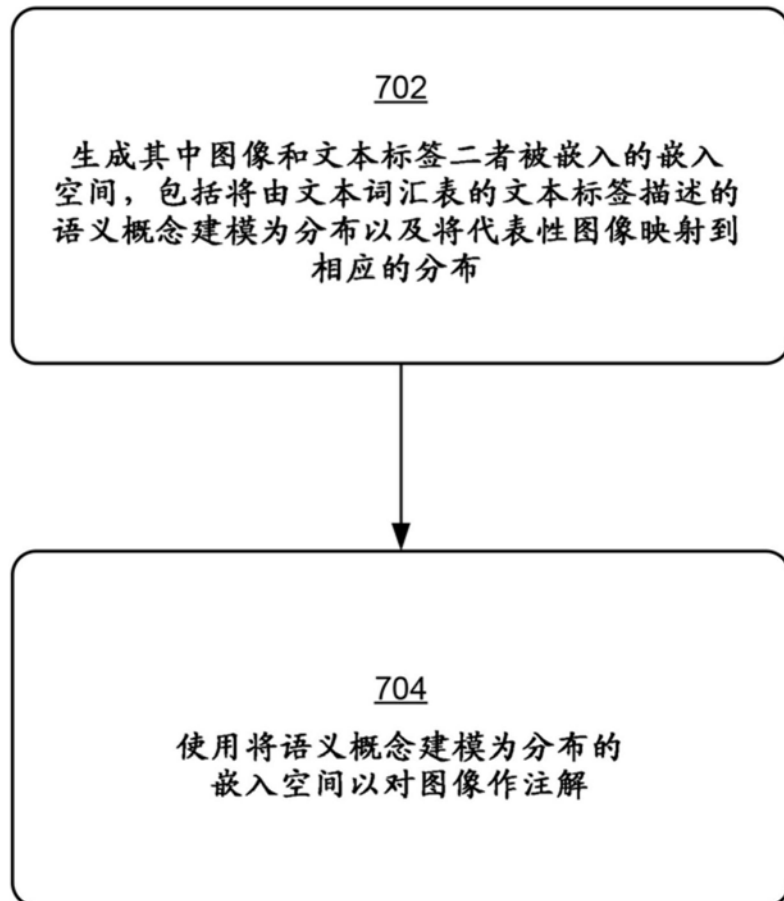


图7

800

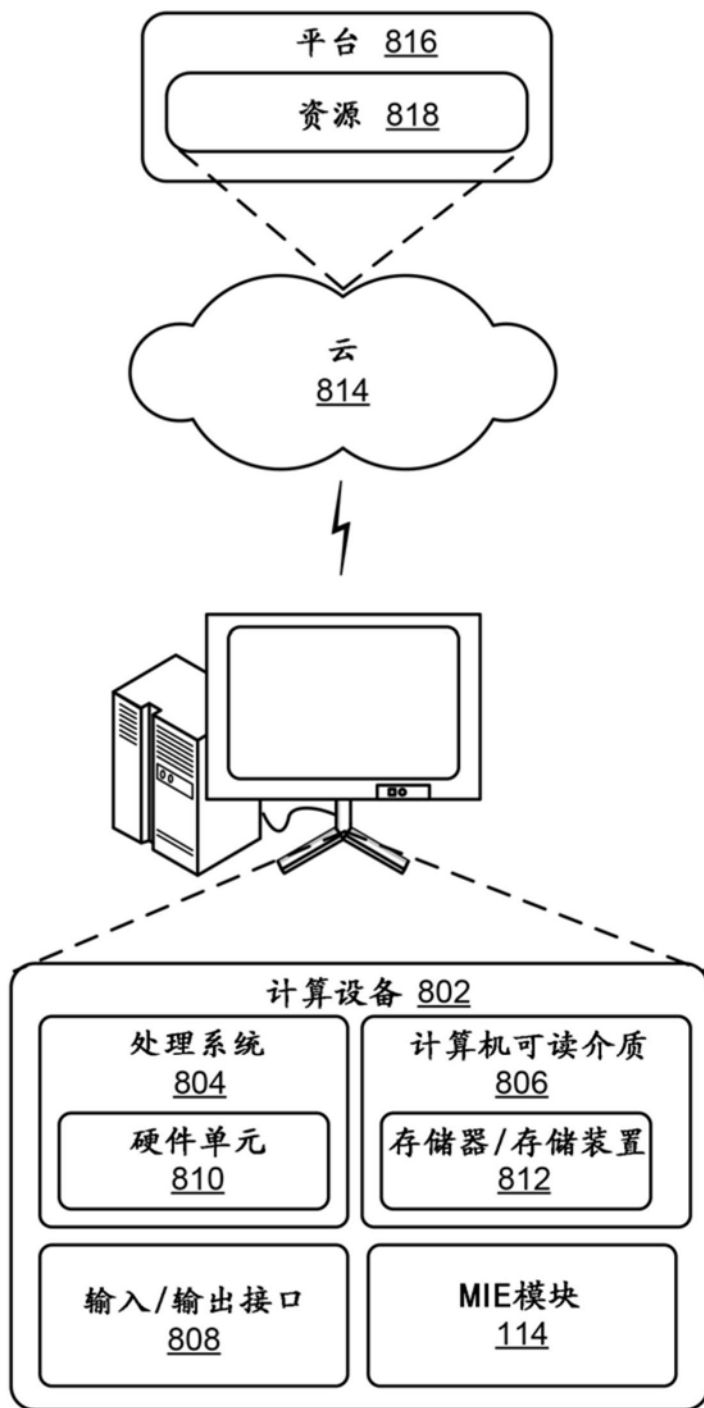


图8