



(12) 发明专利

(10) 授权公告号 CN 110490213 B

(45) 授权公告日 2021.10.29

(21) 申请号 201910848729.7
 (22) 申请日 2017.09.11
 (65) 同一申请的已公布的文献号
 申请公布号 CN 110490213 A
 (43) 申请公布日 2019.11.22
 (62) 分案原申请数据
 201710814187.2 2017.09.11
 (73) 专利权人 腾讯科技(深圳)有限公司
 地址 518057 广东省深圳市南山区高新区
 科技中一路腾讯大厦35层
 (72) 发明人 姜文浩 马林 刘威
 (74) 专利代理机构 北京三高永信知识产权代理
 有限责任公司 11138
 代理人 祝亚男
 (51) Int. Cl.
 G06K 9/46 (2006.01)
 G06K 9/48 (2006.01)
 G06K 9/62 (2006.01)
 G06N 3/04 (2006.01)

(56) 对比文件
 CN 107038221 A, 2017.08.11
 CN 102982799 A, 2013.03.20
 CN 102577387 A, 2012.07.11
 US 2013173269 A1, 2013.07.04
 CN 106446782 A, 2017.02.22
 Zhilin Yang et al..Review Networks
 for Caption Generation.《arXiv:
 1605.07912v4 [cs.LG]》.2016,
 Tseng-Hung Chen et al..Show, Adapt
 and Tell: Adversarial Training of Cross-
 domain Image Captioner.《arXiv:
 1705.00930v2 [cs.CV]》.2017,
 Xu Jia et al..Guiding the Long-Short
 Term Memory model for Image Caption
 Generation.《2015 IEEE International
 Conference on Computer Vision (ICCV)》
 .2015,

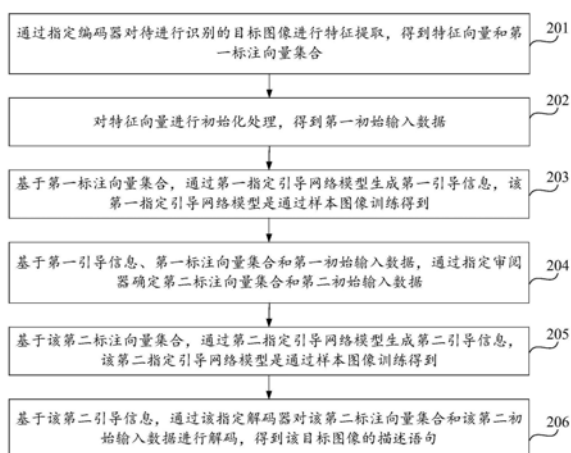
审查员 程呈

权利要求书3页 说明书21页 附图9页

(54) 发明名称
 图像识别方法、装置及存储介质

(57) 摘要
 本发明公开了一种图像识别方法、装置及存储介质,属于机器学习领域。所述方法包括:通过编码器对图像进行编码,得到第一标注向量集合和第一初始输入数据;基于第一标注向量集合,通过第一指定引导网络模型生成第一引导信息;基于第一引导信息、第一标注向量集合和第一初始输入数据,通过审阅器确定第二标注向量集合和第二初始输入数据;基于第二标注向量集合,通过第二指定引导网络模型生成第二引导信息;基于第二引导信息,通过解码器对第二标注向量集合和第二初始输入数据进行解码,得到描述语句。本发明提供了一种基于机器学习的图像识别方法,在编码器和解码器之间增加了两个引导网络模型和一个审阅器,提高了生成描述语句的质

量。



CN 110490213 B

1. 一种图像识别方法,其特征在于,所述方法包括:

通过指定编码器对待进行识别的目标图像进行特征提取,得到特征向量和第一标注向量集合;

对所述特征向量进行初始化处理,得到第一初始输入数据;

基于所述第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,所述第一指定引导网络模型是通过样本图像训练得到;

基于所述第一引导信息、所述第一标注向量集合和所述第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据;

基于所述第二标注向量集合,通过第二指定引导网络模型生成第二引导信息,所述第二指定引导网络模型是通过样本图像训练得到;

基于所述第二引导信息,通过指定解码器对所述第二标注向量集合和所述第二初始输入数据进行解码,得到所述目标图像的描述语句。

2. 如权利要求1所述的方法,其特征在于,所述第一指定引导网络模型为第一引导网络模型,所述第一引导网络模型是通过样本图像的标注向量集合训练得到;

所述基于所述第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,包括:

基于所述第一引导网络模型中的模型参数构成的第一矩阵对所述第一标注向量集合进行线性变换,得到第二矩阵;

基于所述第二矩阵中每一行的最大值确定所述第一引导信息。

3. 如权利要求1所述的方法,其特征在于,所述第一指定引导网络模型为第二引导网络模型,所述第二引导网络模型是通过样本图像的标注向量集合和属性信息训练得到,所述属性信息用于指示所述样本图像的描述语句中预测出现的词语的概率;

所述基于所述第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,包括:

将所述目标图像作为多示例模型的输入,通过所述多示例模型对所述目标图像进行处理,得到所述目标图像的属性信息;

基于所述第二引导网络模型中的模型参数构成的第二矩阵对所述第一标注向量集合进行线性变换,得到第三矩阵;

基于所述第三矩阵和所述目标图像的属性信息,生成第四矩阵;

基于所述第四矩阵中每一行的最大值确定所述第一引导信息。

4. 如权利要求1所述的方法,其特征在于,所述基于所述第一引导信息、所述第一标注向量集合和所述第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据,包括:

当所述指定审阅器采用第二RNN模型,且所述第二RNN模型用于执行N个第二时间步骤时,对于所述第二RNN模型执行的每个第二时间步骤,基于所述第一引导信息确定所述第二时间步骤的输入数据;

其中,所述N是指所述第二RNN模型循环处理输入数据的次数,且所述N为正整数,每个第二时间步骤为所述第二RNN模型对输入数据的处理步骤;

基于所述第二时间步骤的输入数据、所述第一标注向量集合和所述第二时间步骤的上

一个第二时间步骤的输出数据,确定所述第二时间步骤的输出数据;

其中,当所述第二时间步骤为所述N个第二时间步骤中的第一个第二时间步骤时,所述第二时间步骤的上一个第二时间步骤的输出数据是基于所述第一初始输入数据确定得到;

基于所述N个第二时间步骤中最后一个第二时间步骤的输出数据,确定所述第二初始输入数据;

基于所述N个第二时间步骤的所有输出数据,确定所述第二标注向量集合。

5.如权利要求4所述的方法,其特征在于,所述基于所述第一引导信息确定所述第二时间步骤的输入数据,包括:

基于所述第一引导信息,通过以下公式确定所述第二时间步骤的输入数据:

$$x_t' = Q'v$$

其中,t为所述第二时间步骤, x_t' 为所述第二时间步骤的输入数据, Q' 为所述第二RNN模型的模型参数构成的第六矩阵,v为所述第一引导信息。

6.如权利要求1所述的方法,其特征在于,所述基于所述第二引导信息,通过指定解码器对所述第二标注向量集合和所述第二初始输入数据进行解码,得到所述目标图像的描述语句,包括:

当所述指定解码器采用第一RNN模型,且所述第一RNN模型用于执行M个第一时间步骤时,对于所述第一RNN模型执行的每个第一时间步骤,基于所述第二引导信息确定所述第一时间步骤的输入数据;

其中,所述M是指所述第一RNN模型循环处理输入数据的次数,且所述M为正整数,每个第一时间步骤为所述第一RNN模型对输入数据的处理步骤;

基于所述第一时间步骤的输入数据、所述第二标注向量集合和所述第一时间步骤的上一个第一时间步骤的输出数据,确定所述第一时间步骤的输出数据;

其中,当所述第一时间步骤为所述M个第一时间步骤中的第一个第一时间步骤时,所述第一时间步骤的上一个第一时间步骤的输出数据是基于所述第二初始输入数据确定得到;

基于所述M个第一时间步骤的所有输出数据,确定所述目标图像的描述语句。

7.如权利要求6所述的方法,其特征在于,所述基于所述第二引导信息确定所述第一时间步骤的输入数据,包括:

基于所述第二引导信息,通过以下公式确定所述第一时间步骤的输入数据:

$$x_t = Ey_t + Qv'$$

其中,t为所述第一时间步骤, x_t 为所述第一时间步骤的输入数据,E为所述第一RNN模型的模型参数构成的词语嵌入矩阵, y_t 是所述第一时间步骤对应的词语的独热one-hot向量,所述第一时间步骤对应的词语是基于所述第一时间步骤的上一个第一时间步骤的输出数据确定得到,Q为所述第一RNN模型的模型参数构成的第五矩阵, v' 为所述第二引导信息。

8.如权利要求1-7任一所述的方法,其特征在于,所述通过指定编码器对目标图像进行特征提取,得到特征向量和第一标注向量集合之前,还包括:

将第二待训练编码器、第二待训练引导网络模型、待训练审阅器、第三待训练引导网络模型和第二待训练解码器进行组合,得到第二级联网络模型;

基于多个样本图像和所述多个样本图像的描述语句,采用梯度下降法对所述第二级联网络模型进行训练,得到所述指定编码器、所述第一指定引导网络模型、所述指定审阅器、

所述第二指定引导网络模型和所述指定解码器。

9. 一种图像识别装置,其特征在于,所述装置包括:

提取模块,用于通过指定编码器对待进行识别的目标图像进行特征提取,得到特征向量和第一标注向量集合;

处理模块,用于对所述特征向量进行初始化处理,得到第一初始输入数据;

生成模块,用于基于所述第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,所述第一指定引导网络模型是通过样本图像训练得到;

确定模块,用于基于所述第一引导信息、所述第一标注向量集合和所述第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据;基于所述第二标注向量集合,通过第二指定引导网络模型生成第二引导信息,所述第二指定引导网络模型是通过样本图像训练得到;基于所述第二引导信息,通过指定解码器对所述第二标注向量集合和所述第二初始输入数据进行解码,得到所述目标图像的描述语句。

10. 如权利要求9所述的装置,其特征在于,所述装置还包括:

第一组合模块,用于将第一待训练编码器、第一待训练引导网络模型和第一待训练解码器进行组合,得到第一级联网络模型;

第一训练模块,用于基于多个样本图像和所述多个样本图像的描述语句,采用梯度下降法对所述第一级联网络模型进行训练,得到所述指定编码器、所述第一指定引导网络模型和所述指定解码器。

11. 一种终端,其特征在于,所述终端包括处理器和存储器,所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集,所述指令、所述程序、所述代码集或所述指令集由所述处理器加载并执行以实现如权利要求1-8任一项所述的图像识别方法。

12. 一种计算机可读存储介质,其特征在于,所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,所述指令、所述程序、所述代码集或所述指令集由处理器加载并执行以实现如权利要求1-8任一项所述的图像识别方法。

图像识别方法、装置及存储介质

[0001] 本申请是申请日为2017年09月11日、申请号为201710814187.2、发明名称为“图像识别方法、装置及存储介质”的发明专利申请的分案申请。

技术领域

[0002] 本发明涉及机器学习领域,特别涉及一种图像识别方法、装置及存储介质。

背景技术

[0003] 随着科技的发展,以及人们对便捷的人机交互方式的需求,机器学习在图像识别领域得到了广泛应用。例如,在早期的儿童教育、图像检索和盲人导航等场景中,人们通常希望机器能够自动对图像进行识别,得到能够准确描述图像内容的描述语句,即将图像翻译成自然语言,以便通过自然语言快速理解图像或者对图像进行分类。

[0004] 目前,图像识别的系统框架通常包括编码器(Encoder)和解码器(Decoder),基于该系统框架,相关技术中提出了一种图像识别方法,包括:先通过编码器对图像进行特征提取,得到特征向量和标注向量(Annotation Vectors)集合,其中,特征向量是对图像进行全局特征提取得到,标注向量集合是对图像进行局部特征提取得到,然后对特征向量进行初始化处理,得到初始输入数据,该初始输入数据可以包括初始的隐含状态(Hidden State)信息和初始的记忆单元(Memory Cell)状态信息,之后从图像中提取人为设计的特定信息作为引导信息,并基于该引导信息,通过解码器对该标注向量集合和初始输入数据进行解码,得到描述语句。其中,该引导信息用于对编码器的编码过程进行引导,以提高生成描述语句的质量,使得所生成的描述语句能够较为准确地描述图像且符合语义。

[0005] 由于相关技术中所使用的引导信息是人为预先设计的特定信息,因此该引导信息可能不能对所有图像的编码过程进行准确引导,即通过该引导信息可能不能准确生成图像的描述语句,导致生成的描述语句的质量较低。

发明内容

[0006] 为了解决相关技术中存在的通过人为设计的特定引导信息不能准确生成图像的描述语句,导致生成的描述语句的质量较低的问题,本发明实施例提供了一种图像识别方法、装置及存储介质。所述技术方案如下:

[0007] 第一方面,提供了一种图像识别方法,所述方法包括:

[0008] 通过指定编码器对待进行识别的目标图像进行特征提取,得到特征向量和第一标注向量集合;

[0009] 对所述特征向量进行初始化处理,得到第一初始输入数据;

[0010] 基于所述第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,所述第一指定引导网络模型是通过样本图像训练得到;

[0011] 基于所述第一引导信息、所述第一标注向量集合和所述第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据;

[0012] 基于所述第二标注向量集合,通过第二指定引导网络模型生成第二引导信息,所述第二指定引导网络模型是通过样本图像训练得到;

[0013] 基于所述第二引导信息,通过指定解码器对所述第二标注向量集合和所述第二初始输入数据进行解码,得到所述目标图像的描述语句。

[0014] 第二方面,提供了一种图像识别装置,所述装置包括:

[0015] 提取模块,用于通过指定编码器对待进行识别的目标图像进行特征提取,得到特征向量和第一标注向量集合;

[0016] 处理模块,用于对所述特征向量进行初始化处理,得到第一初始输入数据;

[0017] 生成模块,用于基于所述第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,所述第一指定引导网络模型是通过样本图像训练得到;

[0018] 确定模块,用于基于所述第一引导信息、所述第一标注向量集合和所述第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据;基于所述第二标注向量集合,通过第二指定引导网络模型生成第二引导信息,所述第二指定引导网络模型是通过样本图像训练得到;基于所述第二引导信息,通过指定解码器对所述第二标注向量集合和所述第二初始输入数据进行解码,得到所述目标图像的描述语句。

[0019] 第三方面,提供了一种终端,所述终端包括处理器和存储器,所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集,所述指令、所述程序、所述代码集或所述指令集由所述处理器加载并执行以实现如第一方面所述的图像识别方法。

[0020] 第四方面,提供了一种计算机可读存储介质,其特征在于,所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,所述指令、所述程序、所述代码集或所述指令集由处理器加载并执行以实现如第一方面所述的图像识别方法。

[0021] 本发明实施例提供的技术方案带来的有益效果是:

[0022] 本发明实施例中,在编码器和解码器之间增加了引导网络模型,从图像中提取标注向量集合之后,可以基于该标注向量集合通过该引导网络模型生成引导信息,由于该引导网络模型是通过样本图像训练得到,可以在训练过程中自适应地学习引导信息,因此通过该引导网络模型所生成的引导信息准确度较高,能够对图像的编码过程进行准确引导,从而提高了生成描述语句的质量。另外,通过在编码器和解码器之间增加审阅器,可以通过审阅器进一步挖掘目标图像的局部特征和全局特征的交互关系,使得生成的第二标注向量集合和第二初始输入数据能够更准确地指示目标图像的特征,进一步提高了图像识别系统的系统性能和生成描述语句的质量。

附图说明

[0023] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例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发明的一些实施例,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。

[0024] 图1A是本发明实施例提供的一种RNN模型的逻辑结构示意图;

[0025] 图1B是本发明实施例提供的一种LSTM模型的逻辑结构示意图;

[0026] 图1C是本发明实施例提供的一种图像识别系统的结构示意图;

- [0027] 图1D是本发明实施例提供的另一种图像识别系统的结构示意图；
- [0028] 图1E是本发明实施例提供的又一种图像识别系统的结构示意图；
- [0029] 图1F是本发明实施例提供的又一种图像识别系统的结构示意图；
- [0030] 图1G是本发明实施例提供的一种图像识别方法流程图；
- [0031] 图2是本发明实施例提供的另一种图像识别方法流程图；
- [0032] 图3A是本发明实施例提供的一种图像识别装置的结构示意图；
- [0033] 图3B是本发明实施例提供的一种生成模块303的结构示意图；
- [0034] 图3C是本发明实施例提供的另一种生成模块303的结构示意图；
- [0035] 图3D是本发明实施例提供的一种确定模块304的结构示意图；
- [0036] 图3E是本发明实施例提供的另一种图像识别装置的结构示意图；
- [0037] 图3F是本发明实施例提供的另一种确定模块304的结构示意图；
- [0038] 图3G是本发明实施例提供的又一种图像识别装置的结构示意图；
- [0039] 图4是本发明实施例提供的一种终端400的结构示意图。

具体实施方式

[0040] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

[0041] 在对本发明实施例进行详细地解释说明之前，先对本发明实施例涉及的名词进行解释说明。

[0042] 编码器

[0043] 编码器用于对图像进行编码生成向量，编码器通常采用CNN (Convolutional Neural Networks, 卷积神经网络) 模型。

[0044] 解码器

[0045] 解码器用于对编码器生成的向量进行解码，即将编码器生成的向量翻译成图像的描述语句，解码器通常采用RNN (Recurrent Neural Network, 循环神经网络) 模型。

[0046] 引导信息

[0047] 引导信息是对图像进行处理得到的信息，通常表示为向量，能够作为解码器输入的一部分来对解码过程进行引导。在解码器中引入引导信息可以提高解码器的性能，保证解码器能够生成更好的描述语句，提高生成描述语句的质量。

[0048] CNN模型

[0049] CNN模型是指在传统的多层神经网络的基础上发展起来的一种针对图像分类和识别的神经网络模型，CNN模型通常包括多个卷积层和至少一个全连接层，能够对图像进行特征提取。

[0050] RNN模型

[0051] 由于传统的神经网络没有记忆功能，也即，对于传统的神经网络而言，其输入为独立的没有上下文关联的数据。但是实际应用中，输入通常为一些有明显上下文特征的序列化输入，比如需要预测描述语句中的下一个词语，此时神经网络的输出必须依赖上一次的输入。也即，要求神经网络应具有记忆功能，而RNN模型即为一种节点定向连接成环且具有记忆功能的神经网络，可以利用内部的记忆功能循环处理输入数据。

[0052] 图1A是本发明实施例提供的一种RNN模型的逻辑结构示意图,如图1A左侧所示,该RNN模型包括输入层、隐含层和输出层三层结构,且隐含层为环形结构。其中,输入层和隐含层相连,隐含层和输出层相连。

[0053] 为了便于说明该RNN模型的功能,将图1A左侧所示的RNN模型的结构按照时间顺序进行展开,可以得到如图1A右侧所示的结构。由于RNN模型的输入层接收到的输入数据为按照一定时间序列排序的数据,也即输入层接收到的输入数据为序列数据,为了便于说明,将该序列数据标记为 $x_1, x_2, \dots, x_i, \dots, x_n$,该序列数据中的各个数据分别对应的时刻为 $t_1, t_2, \dots, t_i, \dots, t_n$,将对 $x_1, x_2, \dots, x_i, \dots, x_n$ 分别进行处理得到的输出数据标记为 $f_1, f_2, \dots, f_i, \dots, f_n$,而RNN模型按照时间顺序对各个输入数据依次进行处理的步骤可以称为时间步骤。其中, n 为RNN模型循环处理输入数据的次数。

[0054] 如图1A右侧所示,在展开之后的RNN模型中, t_1 时刻输入层接收到的输入数据为 x_1 ,并将 x_1 传输至隐含层,隐含层对 x_1 进行处理,并将处理后的数据传输至输出层,得到 t_1 时刻的输出数据 f_1 。 t_2 时刻输入层接收到的输入数据为 x_2 ,并将 x_2 传输至隐含层,此时隐含层根据 t_1 时刻的输出数据 f_1 对 x_2 进行处理,并将处理后的数据传输至输出层,得到 t_2 时刻的输出数据 f_2 。也即,在任意时刻 t_i ,隐含层除了接收到 t_i 时刻输入层传输的输入数据 x_i ,还接收到 t_{i-1} 时刻的输出数据 f_{i-1} ,并根据 f_{i-1} 对 x_i 进行处理,得到 t_i 时刻的输出数据 f_i 。

[0055] LSTM (Long Short-Term Memory, 长短期记忆) 网络模型

[0056] LSTM网络模型是一种特殊的RNN模型,能够处理和预测时间序列中间隔和延迟相对较长的重要事件。LSTM网络模型包括LSTM单元,LSTM单元设置有输入门、遗忘门和输出门,在每个时间步骤可以基于设置的输入门、遗忘门和输出门对输入数据进行处理。

[0057] 图1B是本发明实施例提供的一种LSTM网络模型的逻辑结构示意图,如图1B左侧所示,该LSTM网络模型包括LSTM单元,且LSTM单元为环形结构,对于LSTM单元执行的任一时间步骤 t 来说,该LSTM单元可以对时间步骤 t 的输入数据 x_t 和上一个时间步骤 $t-1$ 的输出数据 f_{t-1} 进行处理,得到时间步骤 t 的输出数据 f_t 。

[0058] 如图1B右侧所示,在按照时间顺序展开之后的LSTM网络模型中,LSTM单元接收到时间步骤 t_1 的输入数据 x_1 之后,可以对 x_1 进行处理得到时间步骤 t_1 的输出数据 f_1 ,然后将 f_1 再输入LSTM单元,LSTM单元接收到时间步骤 t_2 的输入数据 x_2 之后,可以对 f_1 和 x_2 进行处理,得到时间步骤 t_2 的输出数据 f_2 ,直至基于时间步骤 t_n 的输入数据 x_n 和时间步骤 t_{n-1} 的输出数据 f_{n-1} 得到时间步骤 t_n 的输出数据 f_n 为止。其中, n 为LSTM网络模型循环处理输入数据的次数。

[0059] 审阅网络 (Review-net)

[0060] 审阅网络是一种基于编码器-解码器框架的图像识别算法,包括审阅器 (reviewer) 和解码器。审阅器和解码器通常都采用CNN模型。审阅器可以进一步挖掘编码器从图像中提取的全局特征和局部特征之间的交互关系,并基于全局特征和局部特征之间的交互关系为解码器生成初始输入数据,以提高解码器的性能。

[0061] 人工智能 (Artificial Intelligence, AI)

[0062] 人工智能是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能

以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。

[0063] 人工智能技术是一门综合学科,涉及领域广泛,既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0064] 计算机视觉技术(Computer Vision,CV)

[0065] 计算机视觉是一门研究如何使机器“看”的科学,更进一步的说,就是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等机器视觉,并进一步做图形处理,使电脑处理成为更适合人眼观察或传送给仪器检测的图像。作为一个科学学科,计算机视觉研究相关的理论和技术,试图建立能够从图像或者多维数据中获取信息的人工智能系统。计算机视觉技术通常包括图像处理、图像识别、图像语义理解、图像检索、OCR、视频处理、视频语义理解、视频内容/行为识别、三维物体重建、3D技术、虚拟现实、增强现实、同步定位与地图构建等技术,还包括常见的人脸识别、指纹识别等生物特征识别技术。

[0066] 语音技术(Speech Technology)

[0067] 语音技术的关键技术有自动语音识别技术(ASR)和语音合成技术(TTS)以及声纹识别技术。让计算机能听、能看、能说、能感觉,是未来人机交互的发展方向,其中语音成为未来最被看好的人机交互方式之一。

[0068] 自然语言处理(Nature Language processing,NLP)

[0069] 自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此,这一领域的研究将涉及自然语言,即人们日常使用的语言,所以它与语言学的研究有着密切的联系。自然语言处理技术通常包括文本处理、语义理解、机器翻译、机器人问答、知识图谱等技术。

[0070] 机器学习(Machine Learning,ML)

[0071] 机器学习是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0072] 自动驾驶技术通常包括高精地图、环境感知、行为决策、路径规划、运动控制等技术,自动驾驶技术有着广泛的应用前景,

[0073] 随着人工智能技术研究和进步,人工智能技术在多个领域展开研究和应用,例如常见的智能家居、智能穿戴设备、虚拟助理、智能音箱、智能营销、无人驾驶、自动驾驶、无人机、机器人、智能医疗、智能客服等,相信随着技术的发展,人工智能技术将在更多的领域得到应用,并发挥越来越重要的价值。

[0074] 本申请实施例提供的方案涉及人工智能的计算机视觉技术、机器学习技术和自然语言处理等技术,具体通过如下图1G和图2实施例进行说明。

[0075] 接下来对本发明实施例的应用场景予以说明。

[0076] 本发明实施例可以应用于早期的儿童教育、图像检索、聊天系统和盲人导航等场景中,在这些场景中通常需要机器能够自动将图像翻译成自然语言。

[0077] 例如,为了提高幼龄儿童的看图识物能力,可以通过机器将幼龄儿童看到的图像翻译成自然语句,然后将自然语句转换成语音播放出来,以便幼龄儿童能够结合图像和语音学习图像内容。

[0078] 再例如,对于数据库中存储的大量图像,可以将图像翻译成对应的描述语句,以便根据图像的描述语句对图像进行准确分类,或者根据图像的描述语句对图像进行准确检索。

[0079] 接下来对本发明实施例涉及的系统架构进行介绍。

[0080] 图1C是本发明实施例提供的一种图像识别系统的结构示意图,如图1C所示,该图像识别系统包括指定编码器10、第一指定引导网络模型20和指定解码器30。

[0081] 其中,指定编码器10用于对待进行识别的目标图像进行编码,即对目标图像进行特征提取,得到特征向量和第一标注向量集合。特征向量用于指示目标图像的全局特征,第一标注向量集合用于指示目标图像的局部特征。

[0082] 对于第一标注向量集合,指定编码器10可以将其分别输出给指定解码器30和第一指定引导网络模型20。对于特征向量,指定编码器10可以对其进行初始化处理得到第一初始输入数据,然后将第一初始输入数据输出给指定解码器30;或者,指定编码器10也可以输出特征向量,然后通过其他模型对指定编码器10输出的特征向量进行初始化处理得到第一初始输入数据,并将第一初始输入数据输出给指定解码器30。

[0083] 其中,第一指定引导网络模型20用于基于指定编码器10输出的第一标注向量集合生成第一引导信息,然后将第一引导信息输出给指定解码器30,且该第一指定引导网络模型是通过样本图像训练得到。

[0084] 其中,指定解码器30用于基于第一引导信息、第一标注向量集合和第一初始输入数据确定该目标图像的描述语句。

[0085] 由上可知,图1C所示的图像识别系统与相关技术相比,在编码器和解码器之间增加了预先通过样本图像训练得到的引导网络模型,由于该引导网络模型可以在训练的过程中自适应地学习引导信息,因此与人为设计的引导信息相比,通过该引导网络模型所生成的引导信息准确度较高,能够对图像的编码过程进行准确引导,从而提高了生成描述语句的质量。

[0086] 图1D是本发明实施例提供的另一种图像识别系统的结构示意图,如图1D所示,该图像识别系统包括指定编码器10、第一指定引导网络模型20、指定解码器30和多示例模型40。

[0087] 其中,图1D与图1C中的指定编码器10和指定解码器30的作用相同,具体描述可以参考图1C,在此不再详细赘述。

[0088] 其中,多示例模型40用于对待识别的目标图像进行处理,得到目标图像的属性信息,该属性信息用于指示该目标图像的描述语句中预测出现的词语的概率,并将目标图像的属性信息输出给第一指定引导网络模型20。

[0089] 其中,第一指定引导网络模型20用于基于指定编码器10输出的第一标注向量集合和多示例模型40输出的目标图像的属性信息生成第一引导信息。

[0090] 图1D中,通过在第一指定引导网络模型20之前增加多示例模型40,使得第一指定引导网络模型20可以根据目标图像的第一标注向量集合和属性信息综合确定第一引导信息,进一步提高了所生成的第一引导信息的准确性。

[0091] 图1E是本发明实施例提供的又一种图像识别系统的结构示意图,如图1E所示,该图像识别系统包括指定编码器10、第一指定引导网络模型20、指定审阅器50、第二指定引导网络模型60和指定解码器30。

[0092] 其中,图1E与图1C中指定编码器10的作用相同,具体描述可以参考图1C,在此不再详细赘述。

[0093] 其中,第一指定引导网络模型20用于基于指定编码器10输入的第一标注向量集合生成第一引导信息,并将第一引导信息输出给指定审阅器50。

[0094] 其中,指定审阅器50用于基于第一初始输入数据、第一标注向量集合和第一引导信息确定第二标注向量集合和第二初始输入数据,并将第二标注向量集合和第二初始输入数据输出给指定解码器30,以及将第二标注向量集合输出给第二指定引导网络模型60。第二初始输入数据包括初始的隐含状态信息和初始的记忆单元状态信息。

[0095] 其中,第二指定引导网络模型60用于基于第二标注向量集合生成第二引导信息,并将第二引导信息输出给指定解码器30,且该第二指定引导网络模型也是通过样本图像训练得到。

[0096] 其中,指定解码器30用于基于第二引导信息,对第二标注向量集合和第二初始输入数据进行解码,得到该目标图像的描述语句。

[0097] 图1E中,通过在编码器和解码器之间增加审阅器,可以通过审阅器进一步挖掘目标图像的局部特征和全局特征的交互关系,使得生成的第二标注向量集合和第二初始输入数据能够更准确地指示目标图像的特征,进一步提高了图像识别系统的系统性能,进而提高了生成描述语句的质量。

[0098] 图1F是本发明实施例提供的又一种图像识别系统的结构示意图,如图1F所示,该图像识别系统包括指定编码器10、第一指定引导网络模型20、指定审阅器50、第二指定引导网络模型60、指定解码器30和多示例模型40。

[0099] 其中,图1F与图1E中编码器10、指定审阅器50和指定解码器30的作用相同,具体描述可以参考图1E,在此不再赘述。

[0100] 其中,多示例模型40用于对待识别的目标图像进行处理,得到目标图像的属性信息,并将目标图像的属性信息分别输出给第一指定引导网络模型20和第二指定引导网络模型60。

[0101] 其中,第一指定引导网络模型20用于基于指定编码器10输出的第一标注向量集合和多示例模型40输出的目标图像的属性信息生成第一引导信息,并将第一引导信息输出给指定审阅器50。

[0102] 其中,第二指定引导网络模型60用于基于指定审阅器50输出的第二标注向量集合和多示例模型40输出的目标图像的属性信息生成第二引导信息,并将第二引导信息输出给指定解码器30,以便指定编码器30基于第二引导信息,对第二标注向量集合和第二初始输入数据进行编码,得到目标图像的描述语句。

[0103] 图1F中,通过在第一指定引导网络模型20和第二引导网络模型60之前增加多示例

模型40,使得第一指定引导网络模型20和第二引导网络模型60均可以根据目标图像的属性信息和标注向量集合综合确定引导信息,进一步提高了所生成的引导信息的准确性。

[0104] 需要说明的是,上述图1C-图1F所示的图像识别系统均可以基于多个样本图像和多个样本图像的描述语句训练得到,也即是,可以通过训练得到上述指定编码器、第一指定引导网络模型、指定审阅器、第二指定引导网络模型和指定解码器,使得第一指定引导网络模型和第二指定引导网络模型可以在训练的过程中自适应的学习引导信息,从而提高生成引导信息的准确性。

[0105] 接下来将结合上述图1C-图1F所示图像识别系统的结构示意图,对本发明实施例提供的图像识别方法进行详细介绍。图1G是本发明实施例提供的一种图像识别方法流程图,该方法应用于终端中,该终端可以包括上述图像识别系统,例如可以通过安装的软件承载上述图像识别系统。参见图1G,该方法包括:

[0106] 步骤101:通过指定编码器对待进行识别的目标图像进行特征提取,得到特征向量和第一标注向量集合。

[0107] 在对待识别的目标图像进行识别时,可以先将目标图像输入该指定编码器,该指定编码器用于对目标图像进行特征提取,分别得到特征向量和第一标注向量集合。

[0108] 具体地,可以通过指定编码器对目标图像进行全局特征提取,得到特征向量,通过指定编码器对目标图像进行局部特征提取,得到标注向量集合。其中,特征向量用于指示目标图像的全局特征,第一标注向量集合中的标注向量用于指示目标图像的局部特征。

[0109] 实际应用中,该指定编码器通常采用CNN模型,当该指定编码器采用CNN模型对目标图像进行特征提取时,该特征向量可以通过CNN模型的最后一个全连接层提取得到,该第一标注向量集合可以通过CNN模型的最后一个卷积层提取得到。

[0110] 步骤102:对特征向量进行初始化处理,得到第一初始输入数据。

[0111] 其中,第一初始输入数据是指待输入给指定编码器的下一个处理模型的输入数据,用于指示下一个处理模型的初始状态,该下一个处理模型可以为指定解码器或者指定审阅器。

[0112] 其中,第一初始输入数据包括第一初始隐含状态信息和第一初始记忆单元状态信息,第一初始隐含状态信息用于指示下一个处理模型的隐含层的初始状态,第一初始记忆单元状态信息用于指示下一个处理模型的记忆单元的初始状态。

[0113] 具体地,可以对特征向量进行线性变换等初始化处理,得到第一初始输入数据。而且,实际应用中,可以通过指定编码器对该特征向量进行初始化处理,得到第一初始输入数据,也可以通过其他模型对指定编码器输出的特征向量进行初始化处理,得到第一初始输入数据,本发明实施例对此不做限定。

[0114] 例如,该指定编码器可以包括RNN模型和初始化模型,RNN模型用于对目标图像进行特征提取,初始化模型用于对特征向量进行初始化处理,该指定编码器通过RNN模型对图像进行特征提取得到特征向量之后,可以再通过初始化模型对特征向量进行初始化处理,得到第一初始输入数据。

[0115] 或者,指定编码器也可以仅用于对目标图像进行特征提取,并在指定编码器之后增加初始化模型,该初始化模型用于对特征向量进行初始化处理,通过指定编码器对目标图像进行特征提取得到特征向量之后,可以将特征向量输出给该初始化模型,然后通过该

初始化模型对该特征向量进行初始化处理,得到第一初始输入数据。

[0116] 步骤103:基于第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,该第一指定引导网络模型是通过样本图像训练得到。

[0117] 具体地,基于第一标注向量集合,通过第一指定引导网络模型生成第一引导信息可以包括以下两种方式实现:

[0118] 第一种实现方式:该第一指定引导网络模型为第一引导网络模型,第一引导网络模型是通过样本图像的标注向量集合训练得到;基于第一引导网络模型中的模型参数构成的第一矩阵对第一标注向量集合进行线性变换,得到第二矩阵;基于第二矩阵中每一行的最大值确定该第一引导信息。

[0119] 其中,第一引导网络模型是训练得到的能够基于图像的标注向量集合生成第一引导信息的引导网络模型。例如,可以将图1C中的各个模型变换为待训练的模型,然后基于多个样本图像和多个样本图像的描述语句对变换后的图像识别系统进行训练,则在训练的过程中,待训练编码器即可分别从多个样本图像中提取标注向量,并输出给待训练引导网络模型进行训练,如此,对整个图像识别系统训练完成之后,即可将待训练引导网络模型训练为第一引导网络模型。

[0120] 其中,待训练编码器可以为未训练过的编码器,也可以为预训练好的编码器,本发明实施例对此不做限定。通过使用预训练好的编码器对待训练引导网络模型进行训练,可以提高整个图像识别系统的训练效率,进而提高其中的待训练引导网络模型的训练效率。

[0121] 其中,第一标注向量集合也是矩阵形式,第一矩阵为第一引导网络模型的模型参数构成的且用于对第一标注向量集合进行线性变换的矩阵。具体地,可以将第一标注向量集合与第一矩阵进行相乘,以对第一标注向量集合进行线性变换,得到第二矩阵。

[0122] 其中,基于第二矩阵中每一行的最大值确定该第一引导信息包括:选取第二矩阵中每一行的最大值,然后将选取的最大值按照行数不变的原则组成列数为1的矩阵,并将组成的矩阵确定为该第一引导信息。

[0123] 具体地,假设第一标注向量集合为 $\mathcal{A} = [a_1, a_2, \dots, a_k]$, a_1 - a_k 为从目标图像中提取的各个标注向量,第一矩阵为 P_1 ,第一引导信息为 v ,则可以采用如下公式(1)确定第一引导信息:

$$[0124] \quad v = \max([P_1 a_1, P_1 a_2, \dots, P_1 a_k]) \quad (1)$$

[0125] 其中, \max 函数是指对待处理的矩阵的每一行取最大值,并组成行数不变且列数为1的矩阵。

[0126] 第二种实现方式:该第一指定引导网络模型为第二引导网络模型,第二引导网络模型是通过样本图像的标注向量集合和属性信息训练得到,该属性信息用于指示该样本图像的描述语句中预测出现的词语的概率;将该目标图像作为多示例模型的输入,通过该多示例模型对该目标图像进行处理,得到该目标图像的属性信息;基于该第二引导网络模型中的模型参数构成的第二矩阵对该第一标注向量集合进行线性变换,得到第三矩阵;基于该第三矩阵和该目标图像的属性信息,生成第四矩阵;基于该第四矩阵中每一行的最大值确定该第一引导信息。

[0127] 其中,该第二引导网络模型是训练得到的能够基于图像的标注向量集合和属性信息共同生成第一引导信息的引导网络模型。

[0128] 其中,该多示例模型是通过多个样本图像和该多个样本图像的描述语句训练得到的,且能够输出样本图像的属性信息的模型,也即是,该多示例模型能够对图像的描述语句中可能出现的词语的概率进行预测。示例的,该属性信息可以为MIL (Multi-instance learning,多示例学习) 信息等。

[0129] 例如,可以将图1D的各个模型变换为待训练的模型,然后基于多个样本图像和多个样本图像的描述语句对变换后的图像识别系统进行训练,则在训练的过程中,待训练编码器可以从样本图像中提取标注向量并输出给待训练引导网络模型,且待训练多示例模型可以对图像进行处理得到属性信息,并将属性信息输出给待训练引导网络模型,待训练的引导网络模型即可基于样本图像的标注向量和属性信息进行训练,如此,对整个图像识别系统训练完成之后,即可将待训练引导网络模型训练为第二引导网络模型。

[0130] 其中,待训练编码器可以为未训练过的编码器,也可以为预训练好的编码器;待训练多示例模型可以为未训练过的多示例模型,也可以为预训练好的多示例模型,本发明实施例对此不做限定。通过使用预训练好的编码器和/或预训练好的多示例模型来对待训练引导网络模型进行训练,可以提高整个图像识别系统的训练效率,进而提高其中的待训练引导网络模型的训练效率。

[0131] 其中,第一标注向量集合也是矩阵形式,第二矩阵为第二引导网络模型的模型参数构成的且用于对第一标注向量集合进行线性变换的矩阵。具体地,可以将第一标注向量集合与第二矩阵进行相乘,以对第一标注向量集合进行线性变换,得到第三矩阵,然后基于第三矩阵和目标图像的属性信息,生成第四矩阵。

[0132] 其中,基于第四矩阵中每一行的最大值确定第二引导信息包括:选取第四矩阵中每一行的最大值,然后将选取的最大值按照行数不变的原则组成列数为1的矩阵,并将组成的矩阵确定为该第二引导信息。

[0133] 具体地,假设第一标注向量集合为 $\mathcal{A} = [a_1, a_2, \dots, a_k]$, a_1 - a_k 为从目标图像中提取的各个标注向量,第二矩阵为 P_2 ,目标图像的属性信息为 e ,第二引导信息为 v' ,则可以采用如下公式(2)确定第二引导信息 v' :

$$[0134] \quad v' = \max([e, P_2 a_1, P_2 a_2, \dots, P_2 a_k]) \quad (2)$$

[0135] 其中,max函数是指对待处理的矩阵的每一行取最大值,并组成行数不变且列数为1的矩阵。

[0136] 由上可知,第一指定引导网络模型可以通过学习得到,也即是,可以通过多个样本图像和多个样本图像的描述语句训练得到,且在训练的过程中可以自动学习引导信息,因此,通过该第一指定引导网络模型生成第一引导信息的准确度较高,所生成的第一引导信息能够对指定编码的编码过程进行准确引导,进而可以提高生成目标图像的描述语句的质量。

[0137] 步骤104:基于第一引导信息、第一标注向量集合和第一初始输入数据,通过指定解码器确定该目标图像的描述语句。

[0138] 本发明实施例中,基于第一引导信息、第一标注向量集合和第一初始输入数据,通过指定解码器确定该目标图像的描述语句可以包括以下两种实现方式:

[0139] 第一种实现方式:基于第一引导信息,通过指定解码器对第一标注向量集合和第一初始输入数据进行解码,得到该目标图像的描述语句。

[0140] 实际应用中,指定解码器通常采用RNN模型,比如可以采用LSTM网络模型。

[0141] 具体地,基于第一引导信息,通过指定解码器对第一标注向量集合和第一初始输入数据进行解码,得到该目标图像的描述语句可以包括以下步骤1)-3):

[0142] 1) 当该指定解码器采用第一RNN模型,且该第一RNN模型用于执行M个第一时间步骤时,对于该第一RNN模型执行的每个第一时间步骤,基于该第一指定引导信息确定该第一时间步骤的输入数据。

[0143] 其中,所述M是指该第一RNN模型循环处理输入数据的次数,且该M为正整数,每个第一时间步骤为该第一RNN模型对输入数据的处理步骤。

[0144] 其中,基于第一引导信息确定该第一时间步骤的输入数据可以包括基于该第一引导信息,通过以下公式(3)确定该第一时间步骤的输入数据:

$$[0145] \quad x_t = Ey_t + Qv \quad (3)$$

[0146] 其中,t为该第一时间步骤, x_t 为该第一时间步骤的输入数据,E为该第一RNN模型的模型参数构成的词语嵌入矩阵, y_t 是该第一时间步骤对应的词语的独热one-hot向量,该第一时间步骤对应的词语是基于该第一时间步骤的上一个第一时间步骤的输出数据确定得到,Q为该第一RNN模型的模型参数构成的第五矩阵,v为该第一引导信息。

[0147] 2) 基于该第一时间步骤的输入数据、该第一标注向量集合和该第一时间步骤的上一个第一时间步骤的输出数据,确定该第一时间步骤的输出数据。

[0148] 本发明实施例中,通过该第一RNN模型,对该第一时间步骤的输入数据、该第一标注向量集合和该第一时间步骤的上一个第一时间步骤的输出数据进行处理,即可得到该第一时间步骤的输出数据。

[0149] 其中,该第一时间步骤的输出数据可以包括隐含状态信息和记忆单元状态信息。而且,当该第一时间步骤为该M个第一时间步骤中的第一个第一时间步骤时,该第一时间步骤的上一个第一时间步骤的输出数据是基于该第一初始输入数据确定得到。例如,当该第一初始输入数据包括第一初始隐含状态信息 h_0 和第一初始记忆单元状态信息 c_0 ,且该第一时间步骤为第一个第一时间步骤时,则该第一时间步骤的上一个第一时间步骤的输出数据即为 h_0 和 c_0 。

[0150] 本发明实施例中,为了提高所生成的描述语句的质量,所使用的第一RNN模型可以为LSTM网络模型。以LSTM网络模型为例,基于该第一时间步骤的输入数据、该第一标注向量集合和该第一时间步骤的上一个第一时间步骤的输出数据,确定该第一时间步骤的输出数据可以抽象表示为如下公式(4):

$$[0151] \quad h_t = LSTM(x_t, h_{t-1}, \mathcal{A}) \quad (4)$$

[0152] 其中,t为该第一时间步骤, x_t 为该第一时间步骤的输入数据, h_{t-1} 为该第一时间步骤的上一个时间步骤的隐含状态信息, \mathcal{A} 为第一标注向量集合, h_t 为该第一时间步骤的隐含状态信息,LSTM表示LSTM网络模型的处理过程。

[0153] 具体地,LSTM网络模型的处理过程可以采用如下公式表示:

$$[0154] \quad \begin{cases} \begin{pmatrix} i_t \\ f_t \\ o_t \\ g_t \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} T \begin{pmatrix} x_t \\ h_{t-1} \\ d_t \end{pmatrix} \\ c_t = f_t \odot c_{t-1} + i_t \odot g_t \\ h_t = o_t \odot \tanh(c_t) \end{cases} \quad (5)$$

[0155] 其中, i_t 、 f_t 、 c_t 和 o_t 分别为该第一时间步骤在输入门、遗忘门、记忆门和输出门的输出数据, σ 是LSTM网络模型的激活函数,如sigmoid函数, $\tanh()$ 是双曲正切函数, T 是用于线性变换的矩阵, x_t 为该第一时间步骤的输入数据, h_{t-1} 为该第一时间步骤的上一个时间步骤的隐含状态信息, d_t 为基于第一标注向量集合确定得到的指定数据, c_t 为该第一时间步骤的记忆单元状态信息, c_{t-1} 为该第一时间步骤的上一个第一时间步骤的记忆单元状态信息, h_t 为该第一时间步骤的隐含状态信息。

[0156] 其中, 指定数据 d_t 可以为第一标注向量集合,也可以为基于第一标注向量集合和该第一时间步骤的上一个时间步骤的隐含状态信息,通过注意力模型确定得到的上下文向量(Context Vector)。

[0157] 注意力模型可以用来确定上一个第一时间步骤注意的是目标图像的哪个区域,也即是可以为 \mathcal{A} 中的每个标注向量计算一个权重值,标注向量的权重越高表示该标注向量越被注意。

[0158] 在一种可能的实现方式中,该LSTM网络模型可以为设置有注意力模型的LSTM网络模型,在得到第一标注向量集合和该第一时间步骤的上一个时间步骤的隐含状态信息之后,可以基于该第一标注向量集合和该第一时间步骤的上一个时间步骤的隐含状态信息,通过注意力模型确定上下文向量,并将该上下文向量作为该指定数据。

[0159] 具体地,该注意力模型可以计算 \mathcal{A} 中任一个标注向量 a_i 和 h_{t-1} 的相似度 e_i ,然后计算 a_i 的注意力的权重 $w_i = \frac{\exp(a_i)}{\sum_j \exp(a_j)}$,之后使用每个标注向量的权重即可生成上下文向量

$z_t = \sum w_i a_i$ 。

[0160] 3) 基于该M个第一时间步骤的所有输出数据,确定该目标图像的描述语句。

[0161] 具体地,可以对该M个第一时间步骤中所有第一时间步骤的输出数据进行组合处理,得到该目标图像的描述语句。实际应用中,每个第一时间步骤的输出数据通常是一个词语,然后将该M个第一时间步骤输出的M个词语进行组合,即可得到该目标图像的描述语句。

[0162] 以图1C中所示的目标图像为例,该M个第一时间步骤的所有输出数据可能分别为男孩、给、女孩、送、花,则该目标图像的描述语句即为“男孩给女孩送花”。

[0163] 进一步地,为了得到上述能够基于目标图像的标注向量集合准确生成引导信息的第一指定引导网络模型,在通过指定编码器对目标图像进行特征提取,得到特征向量和第一标注向量集合之前,还可以将第一待训练编码器、第一待训练引导网络模型和第一待训练解码器进行组合,得到第一级联网络模型,然后基于多个样本图像和该多个样本图像的描述语句,采用梯度下降法对该第一级联网络模型进行训练,得到该指定编码器、该第一指定引导网络模型和该指定解码器。

[0164] 也即是,可以先将第一待训练编码器、第一待训练引导网络模型和第一待训练解

码器按照图1C或图1D的连接方式构建成能够对图像进行处理,得到图像的描述语句的图像识别系统,然后基于多个样本图像和该多个样本图像的描述语句对该图像识别系统进行训练,在对图像识别系统进行训练的过程中,即可对其中的第一待训练引导网络模型进行训练,使得第一待训练引导网络模型能够在训练的过程中自适应地学习引导信息,保证生成的引导信息能够越来越准确。

[0165] 其中,在训练第一待训练引导网络模型的过程中,可以使用Multi-label margin loss (基于间隔的多标记损失函数)作为该第一待训练引导网络模型的损失函数,并基于该损失函数采用随机梯度下降法对该第一待训练引导网络模型的模型参数进行调整,以得到该第一指定引导网络模型。

[0166] 实际训练中,可以使用已标注的训练集进行训练,该训练集是<样本图像,描述语句>对的集合,比如MSCOCO数据集(一种常用数据集)等。

[0167] 其中,第一待训练编码器可以为未训练过的编码器,也可以为预训练好的编码器,本发明实施例对此不做限定。例如,该第一待训练编码器可以采用在ImageNet(一个计算机视觉系统识别项目名称,是目前世界上图像识别最大的数据库)上预训练好的CNN模型,该CNN模型可以为inception V3模型(一种CNN模型)、Resnet模型(一种CNN模型)或者VGG模型(一种CNN模型)等。

[0168] 通过使用预训练好的编码器作为第一待训练编码器来训练第一指定引导网络模型,可以提高整个第一级联网络模型的训练效率,进而提高其中的第一指定引导网络模型的训练效率。

[0169] 需要说明的是,本发明实施例中,对目标图像进行识别,得到目标图像的描述语句的过程和对引导网络模型进行训练的过程可以在相同的终端上执行,也可以在不同的终端上执行,本发明实施例对此不做限定。

[0170] 第二种实现方式:基于第一引导信息、第一标注向量集合和第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据;基于该第二标注向量集合,通过第二指定引导网络模型生成第二引导信息;基于该第二引导信息,通过该指定编码器对该第二标注向量集合和该第二初始输入数据进行编码,得到该目标图像的描述语句。

[0171] 需要说明的是,该第二种实现方式将在下述图2实施例中进行详细说明,本发明实施例在此不做详细赘述。

[0172] 本发明实施例中,在编码器和解码器之间增加了引导网络模型,从图像中提取标注向量集合之后,可以基于该标注向量集合通过该引导网络模型生成引导信息,由于该引导网络模型是通过样本图像训练得到,可以在训练过程中自适应地学习引导信息,因此通过该引导网络模型所生成的引导信息准确度较高,能够对图像的编码过程进行准确引导,从而提高了生成描述语句的质量。

[0173] 接下来将结合上述图1E和图1F所示的图像识别系统的结构示意图,对本发明实施例提供的图像识别方法进行详细介绍。图2是本发明实施例提供的另一种图像识别方法流程图,该方法应用于终端中。参见图2,该方法包括:

[0174] 步骤201:通过指定编码器对待进行识别的目标图像进行特征提取,得到特征向量和第一标注向量集合。

[0175] 步骤202:对特征向量进行初始化处理,得到第一初始输入数据。

[0176] 步骤203:基于第一标注向量集合,通过第一指定引导网络模型生成第一引导信息。

[0177] 其中,步骤201-步骤203的具体实现方式可以参考上述步骤101-步骤103的相关描述,本发明实施例在此不再赘述。

[0178] 步骤204:基于第一引导信息、第一标注向量集合和第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据。

[0179] 实际应用中,指定解码器和指定审阅器通常均采用RNN模型,当然也可以采用其他模型,本发明实施例对此不做限定。

[0180] 其中,指定审阅器用于进一步挖掘指定编码器从图像中提取的全局特征和局部特征之间的交互关系,并基于全局特征和局部特征之间的交互关系为指定解码器生成初始输入数据,即第二初始输入数据,以提高解码器的性能,进而提高生成描述语句的质量。

[0181] 其中,第一初始输入数据是指待输入给指定审阅器的输入数据,用于指示该指定审阅器的初始状态,具体可以包括第一初始隐含状态信息和第一初始记忆单元状态信息,第一初始隐含状态信息用于指示指定审阅器的隐含层的初始状态,第一初始记忆单元状态信息用于指示指定审阅器的记忆单元的初始状态。

[0182] 其中,第二初始输入数据是指待输入给指定解码器的输入数据,用于指示该指定解码器的初始状态,具体可以包括第二初始隐含状态信息和第二初始记忆单元状态信息,第二初始隐含状态信息用于指示指定解码器的隐含层的初始状态,第二初始记忆单元状态信息用于指示指定解码器的记忆单元的初始状态。

[0183] 具体地,基于该第一引导信息、该第一标注向量集合和该第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据可以包括如下步骤1)-3):

[0184] 1) 当该第一审阅器采用第二RNN模型,且该第二RNN模型用于执行N个第二时间步骤时,对于该第二RNN模型执行的每个第二时间步骤,基于该第一指定引导信息确定该第二时间步骤的输入数据。

[0185] 其中,该N是指该第二RNN模型循环处理输入数据的次数,且该N为正整数,每个第二时间步骤为该第二RNN模型对输入数据的处理步骤。

[0186] 具体地,可以基于该第二引导信息,通过以下公式(6)确定该第二时间步骤的输入数据:

$$[0187] \quad x'_t = Q'v' \quad (6)$$

[0188] 其中,t为该第一时间步骤, x'_t 为该第二时间步骤的输入数据, Q' 为该第二RNN模型的模型参数构成的第六矩阵, v' 为该第二引导信息。

[0189] 2) 基于该第二时间步骤的输入数据、该第一标注向量集合和该第二时间步骤的上一个第二时间步骤的输出数据,确定该第二时间步骤的输出数据。

[0190] 其中,该第二时间步骤的输出数据可以包括隐含状态信息和记忆单元状态信息,当该第二时间步骤为该N个第二时间步骤中的第一个第二时间步骤时,该第二时间步骤的上一个第二时间步骤的输出数据是基于该第一初始输入数据确定得到。

[0191] 本发明实施例中,通过该第二RNN模型,对该第二时间步骤的输入数据、该第二标注向量集合和该第二时间步骤的上一个第二时间步骤的输出数据进行处理,即可得到该第二时间步骤的输出数据。

[0192] 具体地,可以按照上述基于该第一时间步骤的输入数据、该第一标注向量集合和该第一时间步骤的上一个第一时间步骤的输出数据,确定该第一时间步骤的输出数据的方法,基于该第二时间步骤的输入数据、该第一标注向量集合和该第二时间步骤的上一个第二时间步骤的输出数据,确定该第二时间步骤的输出数据,具体实现方式可以参考上述相关描述,在此不再详细赘述。

[0193] 3) 基于该N个第二时间步骤中最后一个第二时间步骤的输出数据,确定该第二初始输入数据。

[0194] 具体地,可以将最后一个第二时间步骤的输出数据确定为该第二初始输入数据,例如,可以将最后一个第二时间步骤的隐含状态信息和记忆单元状态信息确定为该第二初始输入数据,即确定为该指定编码器的初始隐含状态信息和初始记忆单元状态信息。

[0195] 4) 基于该N个第二时间步骤的所有输出数据,确定该第二标注向量集合。

[0196] 具体地,可以将该N个第二时间步骤中所有时间步骤的隐含状态信息的集合确定为该第二标注向量集合。

[0197] 步骤205:基于该第二标注向量集合,通过第二指定引导网络模型生成第二引导信息,该第二指定引导网络模型是通过样本图像训练得到。

[0198] 具体地,可以按照上述图1G实施例中步骤103所述的基于第一标注向量集合,通过第一指定引导网络模型生成第一引导信息的方法,基于第二标注向量集合,通过第二指定引导网络模型生成第二引导信息。具体实现方式可以参数上述步骤103的相关描述,此处不再详细赘述。

[0199] 其中,第二指定引导网络模型可以与第一指定引导网络模型一起通过样本图像进行训练得到,且在训练的过程中可以自动学习引导信息,因此,通过该第一指定引导网络模型和第二指定引导网络模型生成的引导信息的准确度都较高,所生成的引导信息能够对指定编码的编码过程进行准确引导,进而可以提高生成目标图像的描述语句的质量。

[0200] 步骤206:基于该第二引导信息,通过该指定编码器对该第二标注向量集合和该第二初始输入数据进行编码,得到该目标图像的描述语句。

[0201] 具体地,可以按照上述图1G实施例中步骤104所述的基于第一引导信息,通过指定解码器对第一标注向量集合和第一初始输入数据进行解码,得到该目标图像的描述语句的方法,基于该第二引导信息,通过该指定编码器对该第二标注向量集合和该第二初始输入数据进行编码,得到该目标图像的描述语句。具体实现方式可以参考上述步骤104中第一种实现方式的相关描述,此处不再详细赘述。

[0202] 进一步地,为了得到上述能够基于目标图像的第一标注向量集合准确生成第一引导信息的第一指定引导网络模型,以及基于第二标注向量集合准确生成第二引导信息的第二指定引导网络模型,在通过指定编码器对目标图像进行特征提取,得到特征向量和第一标注向量集合之前还可以将第二待训练编码器、第二待训练引导网络模型、待训练审阅器、第三待训练引导网络模型和第二待训练解码器进行组合,得到第二级联网络模型,然后基于多个样本图像和该多个样本图像的描述语句,采用梯度下降法对该第二级联网络模型进行训练,得到该指定编码器、该第一指定引导网络模型、该指定审阅器、该第二指定引导网络模型和该指定解码器。

[0203] 也即是,可以先将第二待训练编码器、第二待训练引导网络模型、待训练审阅器、

第三待训练引导网络模型和第二待训练解码器按照图1E或图1E的连接的方式构建成能够对图像进行处理,得到图像的描述语句的图像识别系统,然后基于多个样本图像和该多个样本图像的描述语句对该图像识别系统进行训练,在对图像识别系统进行训练的过程中,即可对其中的第二待训练引导网络模型和第三待训练引导网络模型进行训练,使得第二待训练引导网络模型和第三待训练引导网络模型能够在训练的过程中自适应地学习引导信息,保证生成的引导信息能够越来越准确。

[0204] 其中,第二待训练编码器可以为未训练过的编码器,也可以为预训练好的编码器,训练审阅器可以为未训练过的审阅器,也可以为预训练好的审阅器,本发明实施例对此不做限定。

[0205] 需要说明的是,通过使用预训练好的编码器作为第二待训练编码器,或者使用预训练好的审阅器最为待训练审阅器来训练第一指定引导网络模型和第二指定引导网络模型,可以提高整个第二级联网络模型的训练效率,进而提高其中的第一指定引导网络模型和第二指定引导网络模型的训练效率。

[0206] 还需要说明的是,本发明实施例中,对目标图像进行识别,得到目标图像的描述语句的过程和对引导网络模型进行训练的过程可以在相同的终端上执行,也可以在不同的终端上执行,本发明实施例对此不做限定。

[0207] 本发明实施例中,在编码器和解码器之间增加了引导网络模型,从图像中提取标注向量集合之后,可以基于该标注向量集合通过该引导网络模型生成引导信息,由于该引导网络模型是通过样本图像训练得到,可以在训练过程中自适应地学习引导信息,因此通过该引导网络模型所生成的引导信息准确度较高,能够对图像的编码过程进行准确引导,从而提高了生成描述语句的质量。

[0208] 进一步地,通过在编码器和解码器之间增加审阅器,可以通过审阅器进一步挖掘目标图像的局部特征和全局特征的交互关系,使得生成的第二标注向量集合和第二初始输入数据能够更准确地指示目标图像的特征,进一步提高了图像识别系统的系统性能,进而提高了生成描述语句的质量。

[0209] 图3A是本发明实施例提供的一种图像识别装置的结构示意图,该装置可以为终端。参见图3A,该装置包括:

[0210] 提取模块301,用于通过指定编码器对待进行识别的目标图像进行特征提取,得到特征向量和第一标注向量集合;

[0211] 处理模块302,用于对该特征向量进行初始化处理,得到第一初始输入数据;

[0212] 生成模块303,用于基于该第一标注向量集合,通过第一指定引导网络模型生成第一引导信息,该第一指定引导网络模型是通过样本图像训练得到;

[0213] 确定模块304,用于基于该第一引导信息、该第一标注向量集合和该第一初始输入数据,通过指定解码器确定该目标图像的描述语句。

[0214] 可选地,参见图3B,该第一指定引导网络模型为第一引导网络模型,该第一引导网络模型是通过样本图像的标注向量集合训练得到;

[0215] 该生成模块303包括:

[0216] 第一线性变换单元3031,用于基于该第一引导网络模型中的模型参数构成的第一矩阵对该第一标注向量集合进行线性变换,得到第二矩阵;

[0217] 第一确定单元3032,用于基于该第二矩阵中每一行的最大值确定该第一引导信息。

[0218] 可选地,参见图3C,该第一指定引导网络模型为第二引导网络模型,该第二引导网络模型是通过样本图像的标注向量集合和属性信息训练得到,该属性信息用于指示该样本图像的描述语句中预测出现的词语的概率;

[0219] 该生成模块303包括:

[0220] 处理单元3033,用于将该目标图像作为多示例模型的输入,通过该多示例模型对该目标图像进行处理,得到该目标图像的属性信息;

[0221] 第二线性变换单元3034,用于基于该第二引导网络模型中的模型参数构成的第二矩阵对该第一标注向量集合进行线性变换,得到第三矩阵;

[0222] 第一生成单元3035,用于基于该第三矩阵和该目标图像的属性信息,生成第四矩阵;

[0223] 第二确定单元3036,用于基于该第四矩阵中每一行的最大值确定该第一引导信息。

[0224] 可选地,该确定模型304用于:

[0225] 基于该第一引导信息,通过该指定解码器对该第一标注向量集合和该第一初始输入数据进行解码,得到该目标图像的描述语句。

[0226] 可选地,参见图3D,该确定模型304包括:

[0227] 第三确定单元3041,用于当该指定解码器采用第一循环神经网络RNN模型,且该第一RNN模型用于执行M个第一时间步骤时,对于该第一RNN模型执行的每个第一时间步骤,基于该第一指定引导信息确定该第一时间步骤的输入数据;

[0228] 其中,该M是指该第一RNN模型循环处理输入数据的次数,且该M为正整数,每个第一时间步骤为该第一RNN模型对输入数据的处理步骤;

[0229] 第四确定单元3042,用于基于该第一时间步骤的输入数据、该第一标注向量集合和该第一时间步骤的上一个第一时间步骤的输出数据,确定该第一时间步骤的输出数据;

[0230] 其中,当该第一时间步骤为该M个第一时间步骤中的第一个第一时间步骤时,该第一时间步骤的上一个第一时间步骤的输出数据是基于该第一初始输入数据确定得到;

[0231] 第五确定单元3043,用于基于该M个第一时间步骤的所有输出数据,确定该目标图像的描述语句。

[0232] 可选地,该第三确定单元3041用于:

[0233] 基于该第一引导信息,通过以下公式确定该第一时间步骤的输入数据:

$$[0234] \quad x_t = Ey_t + Qv$$

[0235] 其中,t为该第一时间步骤, x_t 为该第一时间步骤的输入数据,E为词语嵌入矩阵且为该第一RNN模型的模型参数, y_t 是该第一时间步骤对应的词语的独热one-hot向量,该第一时间步骤对应的词语是基于该第一时间步骤的上一个第一时间步骤的输出数据确定得到,Q为第五矩阵且为该第一RNN模型的模型参数,v为该第一引导信息。

[0236] 可选地,参见图3E,该装置还包括:

[0237] 第一组合模块305,用于将第一待训练编码器、第一待训练引导网络模型和第一待训练解码器进行组合,得到第一级联网络模型;

[0238] 第一训练模块306,基于多个样本图像和该多个样本图像的描述语句,采用梯度下降法对该第一级联网络模型进行训练,得到该指定编码器、该第一指定引导网络模型和该指定解码器。

[0239] 可选地,参见图3F,该确定模型304包括:

[0240] 第六确定单元3044,用于基于该第一引导信息、该第一标注向量集合和该第一初始输入数据,通过指定审阅器确定第二标注向量集合和第二初始输入数据;

[0241] 第二生成单元3045,用于基于该第二标注向量集合,通过第二指定引导网络模型生成第二引导信息,该第二指定引导网络模型是通过样本图像训练得到;

[0242] 编码单元3046,用于基于该第二引导信息,通过该指定编码器对该第二标注向量集合和该第二初始输入数据进行编码,得到该目标图像的描述语句。

[0243] 可选地,该第六确定单元3044用于:

[0244] 当该第一审阅器采用第二RNN模型,且该第二RNN模型用于执行N个第二时间步骤时,对于该第二RNN模型执行的每个第二时间步骤,基于该第一指定引导信息确定该第二时间步骤的输入数据;

[0245] 其中,该N是指该第二RNN模型循环处理输入数据的次数,且该N为正整数,每个第二时间步骤为该第二RNN模型对输入数据的处理步骤;

[0246] 基于该第二时间步骤的输入数据、该第一标注向量集合和该第二时间步骤的上一个第二时间步骤的输出数据,确定该第二时间步骤的输出数据;

[0247] 其中,当该第二时间步骤为该N个第二时间步骤中的第一个第二时间步骤时,该第二时间步骤的上一个第二时间步骤的输出数据是基于该第一初始输入数据确定得到;

[0248] 基于该N个第二时间步骤中最后一个第二时间步骤的输出数据,确定该第二初始输入数据;

[0249] 基于该N个第二时间步骤的所有输出数据,确定该第二标注向量集合。

[0250] 可选地,参见图3G,该装置还包括:

[0251] 第二组合模块307,用于将第二待训练编码器、第二待训练引导网络模型、待训练审阅器、第三待训练引导网络模型和第二待训练解码器进行组合,得到第二级联网络模型;

[0252] 第二训练模块308,用于基于多个样本图像和该多个样本图像的描述语句,采用梯度下降法对该第二级联网络模型进行训练,得到该指定编码器、该第一指定引导网络模型、该指定审阅器、该第二指定引导网络模型和该指定解码器。

[0253] 本发明实施例中,在编码器和解码器之间增加了引导网络模型,从图像中提取标注向量集合之后,可以基于该标注向量集合通过该引导网络模型生成引导信息,由于该引导网络模型是通过样本图像训练得到,可以在训练过程中自适应地学习引导信息,因此通过该引导网络模型所生成的引导信息准确度较高,能够对图像的编码过程进行准确引导,从而提高了生成描述语句的质量。

[0254] 需要说明的是:上述实施例提供的图像识别装置在进行图像识别时,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,即将装置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,上述实施例提供的图像识别装置与图像识别方法实施例属于同一构思,其具体实现过程详见方法实施例,这里不再赘述。

[0255] 图4是本发明实施例提供的一种终端400的结构示意图。参见图4,终端400可以包括通信单元410、包括有一个或一个以上计算机可读存储介质的存储器420、输入单元430、显示单元440、传感器450、音频电路460、WIFI (Wireless Fidelity,无线保真) 模块470、包括有一个或者一个以上处理核心的处理器480、以及电源490等部件。本领域技术人员可以理解,图4中示出的终端结构并不构成对终端的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。其中:

[0256] 通信单元410可用于收发信息或通话过程中,信号的接收和发送,该通信单元410可以为RF (Radio Frequency,射频) 电路、路由器、调制解调器、等网络通信设备。特别地,当通信单元410为RF电路时,将基站的下行信息接收后,交由一个或者一个以上处理器480处理;另外,将涉及上行的数据发送给基站。通常,作为通信单元的RF电路包括但不限于天线、至少一个放大器、调谐器、一个或多个振荡器、用户身份模块 (SIM) 卡、收发信机、耦合器、LNA (Low Noise Amplifier,低噪声放大器)、双工器等。此外,通信单元410还可以通过无线通信与网络和其他设备通信。所述无线通信可以使用任一通信标准或协议,包括但不限于 GSM (Global System of Mobile communication,全球移动通讯系统)、GPRS (General Packet Radio Service,通用分组无线服务)、CDMA (Code Division Multiple Access,码分多址)、WCDMA (Wideband Code Division Multiple Access,宽带码分多址)、LTE (Long Term Evolution,长期演进)、电子邮件、SMS (Short Messaging Service,短消息服务) 等。存储器420可用于存储软件程序以及模块,处理器480通过运行存储在存储器420的软件程序以及模块,从而执行各种功能应用以及数据处理。存储器420可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序 (比如声音播放功能、图像播放功能等) 等;存储数据区可存储根据终端400的使用所创建的数据 (比如音频数据、电话本等) 等。此外,存储器420可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。相应地,存储器420还可以包括存储器控制器,以提供处理器480和输入单元430对存储器420的访问。

[0257] 输入单元430可用于接收输入的数字或字符信息,以及产生与用户设置以及功能控制有关的键盘、鼠标、操作杆、光学或者轨迹球信号输入。优选地,输入单元430可包括触敏表面431以及其他输入设备432。触敏表面431,也称为触摸显示屏或者触控板,可收集用户在其上或附近的触摸操作 (比如用户使用手指、触笔等任何适合的物体或附件在触敏表面431上或在触敏表面431附近的操作),并根据预先设定的程式驱动相应的连接装置。可选的,触敏表面431可包括触摸检测装置和触摸控制器两个部分。其中,触摸检测装置检测用户的触摸方位,并检测触摸操作带来的信号,将信号传送给触摸控制器;触摸控制器从触摸检测装置上接收触摸信息,并将它转换成触点坐标,再送给处理器480,并能接收处理器480发来的命令并加以执行。此外,可以采用电阻式、电容式、红外线以及表面声波等多种类型实现触敏表面431。除了触敏表面431,输入单元430还可以包括其他输入设备432。优选地,其他输入设备432可以包括但不限于物理键盘、功能键 (比如音量控制按键、开关按键等)、轨迹球、鼠标、操作杆等中的一种或多种。

[0258] 显示单元440可用于显示由用户输入的信息或提供给用户的信息以及终端400的各种图形用户接口,这些图形用户接口可以由图形、文本、图标、视频和其任意组合来构成。显示单元440可包括显示面板441,可选的,可以采用LCD (Liquid Crystal Display,液晶显

示器)、OLED (Organic Light-Emitting Diode, 有机发光二极管) 等形式来配置显示面板441。进一步的, 触敏表面431可覆盖显示面板441, 当触敏表面431检测到在其上或附近的触摸操作后, 传送给处理器480以确定触摸事件的类型, 随后处理器480根据触摸事件的类型在显示面板441上提供相应的视觉输出。虽然在图4中, 触敏表面431与显示面板441是作为两个独立的部件来实现输入和输入功能, 但是在某些实施例中, 可以将触敏表面431与显示面板441集成而实现输入和输出功能。

[0259] 终端400还可包括至少一种传感器450, 比如光传感器、运动传感器以及其他传感器。光传感器可包括环境光传感器及接近传感器, 其中, 环境光传感器可根据环境光线的明暗来调节显示面板441的亮度, 接近传感器可在终端400移动到耳边时, 关闭显示面板441和/或背光。作为运动传感器的一种, 重力加速度传感器可检测各个方向上(一般为三轴)加速度的大小, 静止时可检测出重力的大小及方向, 可用于识别手机姿态的应用(比如横竖屏切换、相关游戏、磁力计姿态校准)、振动识别相关功能(比如计步器、敲击)等; 至于终端400还可配置的陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器, 在此不再赘述。

[0260] 音频电路460、扬声器461, 传声器462可提供用户与终端400之间的音频接口。音频电路460可将接收到的音频数据转换后的电信号, 传输到扬声器461, 由扬声器461转换为声音信号输出; 另一方面, 传声器462将收集的声音信号转换为电信号, 由音频电路460接收后转换为音频数据, 再将音频数据输出处理器480处理后, 经通信单元410以发送给比如另一终端, 或者将音频数据输出至存储器420以便进一步处理。音频电路460还可能包括耳塞插孔, 以提供外设耳机与终端400的通信。

[0261] 为了实现无线通信, 该终端上可以配置有无线通信单元470, 该无线通信单元470可以为WIFI模块。WIFI属于短距离无线传输技术, 终端400通过无线通信单元470可以帮助用户收发电子邮件、浏览网页和访问流式媒体等, 它为用户提供了无线的宽带互联网访问。虽然图中示出了无线通信单元470, 但是可以理解的是, 其并不属于终端400的必须构成, 完全可以根据需要在不改变发明的本质的范围内而省略。

[0262] 处理器480是终端400的控制中心, 利用各种接口和线路连接整个手机的各个部分, 通过运行或执行存储在存储器420内的软件程序和/或模块, 以及调用存储在存储器420内的数据, 执行终端400的各种功能和处理数据, 从而对手机进行整体监控。可选的, 处理器480可包括一个或多个处理核心; 优选的, 处理器480可集成应用处理器和调制解调处理器, 其中, 应用处理器主要处理操作系统、用户界面和应用程序等, 调制解调处理器主要处理无线通信。可以理解的是, 上述调制解调处理器也可以不集成到处理器480中。

[0263] 终端400还包括给各个部件供电的电源490(比如电池), 优选的, 电源可以通过电源管理系统与处理器480逻辑相连, 从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。电源490还可以包括一个或一个以上的直流或交流电源、再充电系统、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。

[0264] 尽管未示出, 终端400还可以包括摄像头、蓝牙模块等, 在此不再赘述。

[0265] 在本实施例中, 终端包括处理器和存储器, 存储器中还存储有至少一条指令、至少一段程序、代码集或指令集, 所述指令、所述程序、所述代码集或所述指令集由所述处理器加载并执行以实现上述图1G或图2实施例所述的图像识别方法。

[0266] 在另一实施例中, 还提供了一种计算机可读存储介质, 所述存储介质中存储有至

少一条指令、至少一段程序、代码集或指令集,所述指令、所述程序、所述代码集或所述指令集由处理器加载并执行以实现上述图1G或图2实施例所述的图像识别方法。

[0267] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0268] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

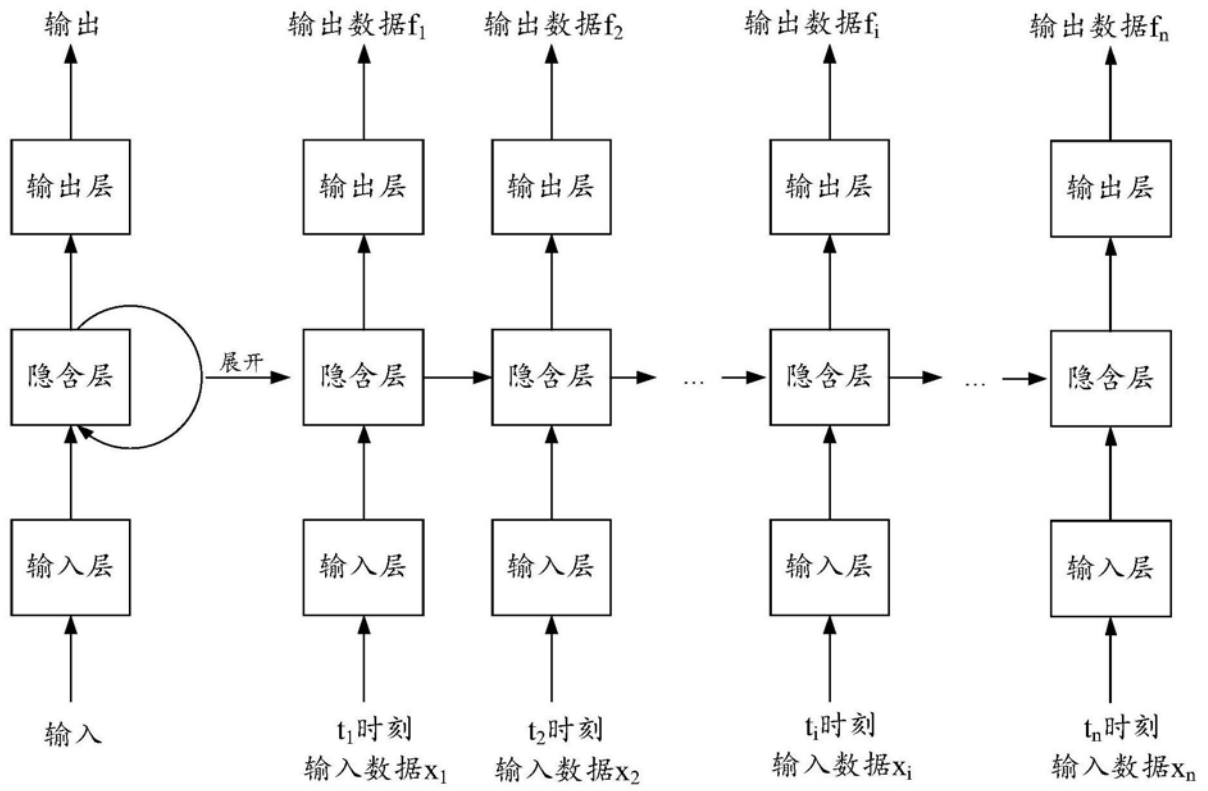


图1A

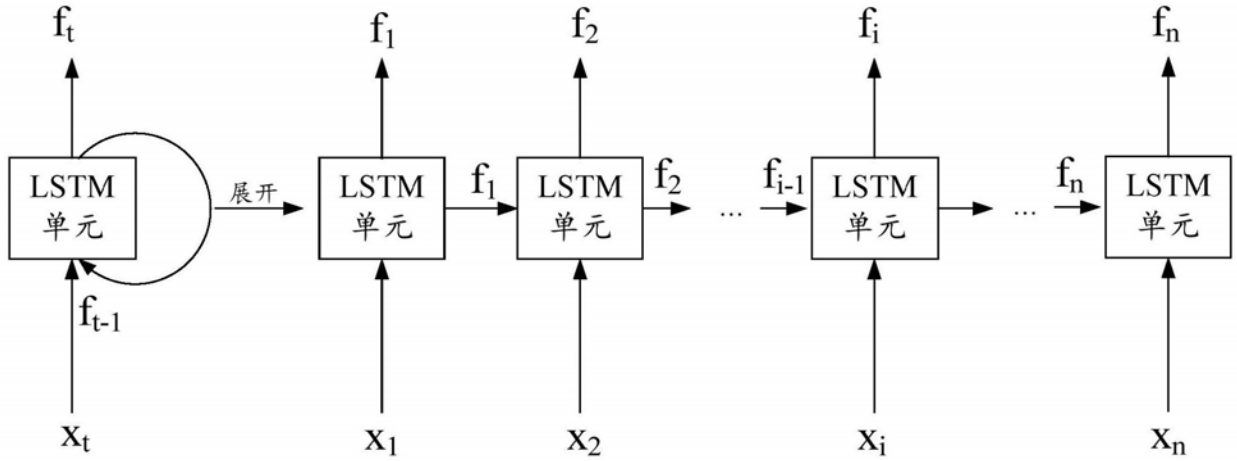


图1B

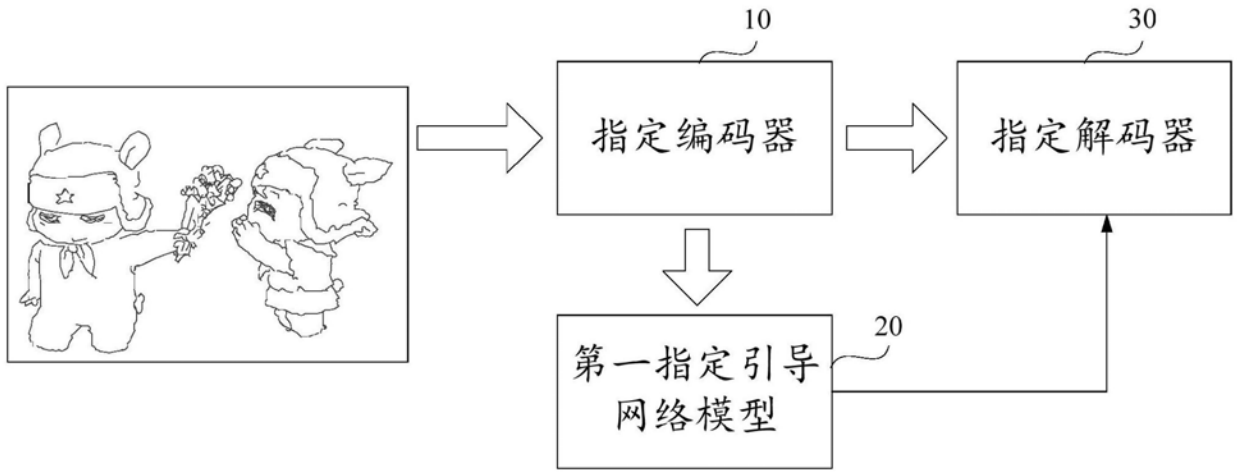


图1C

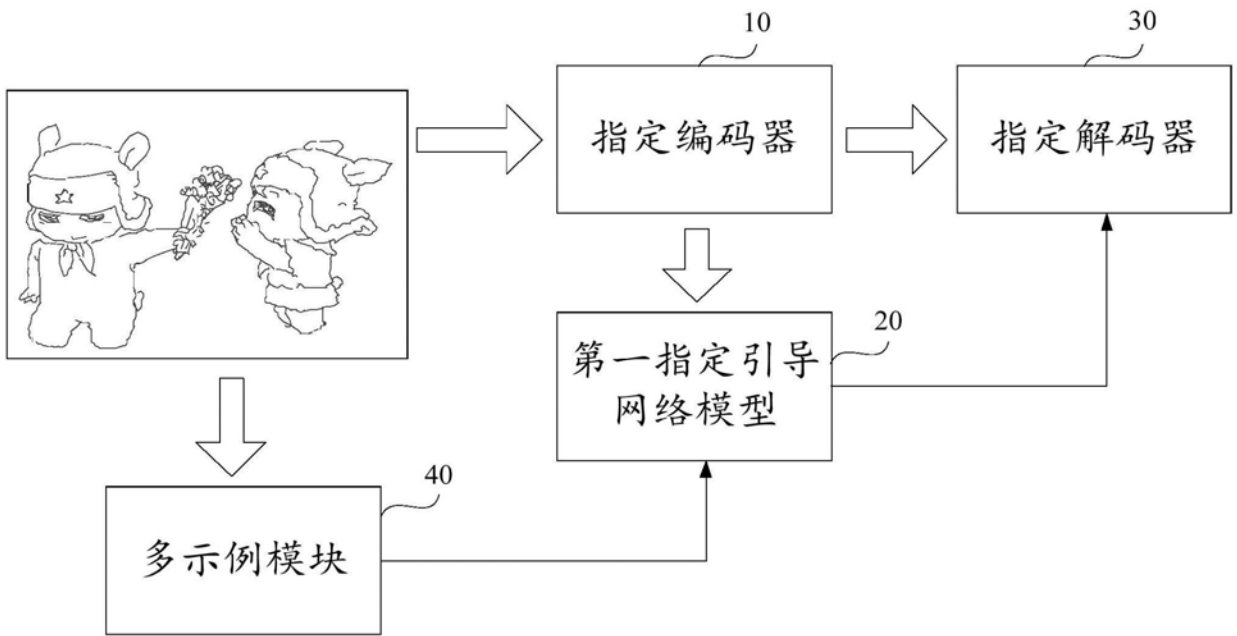


图1D

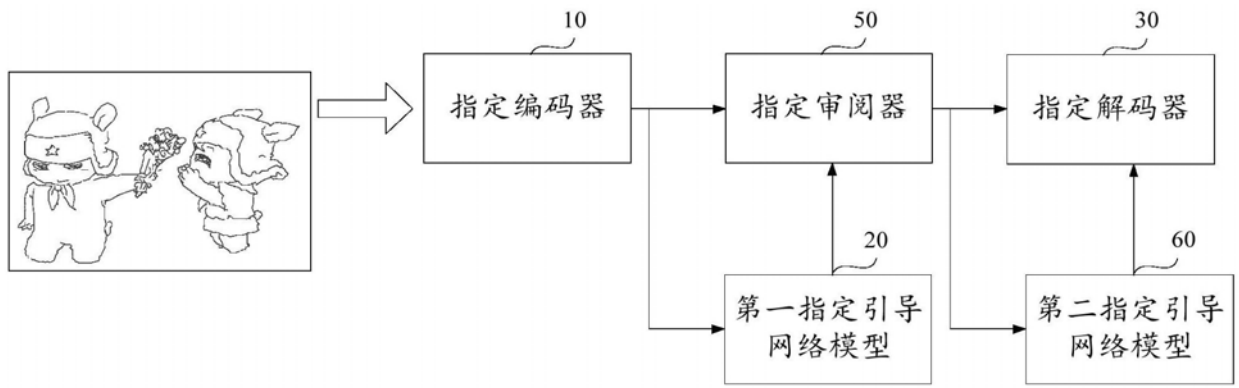


图1E

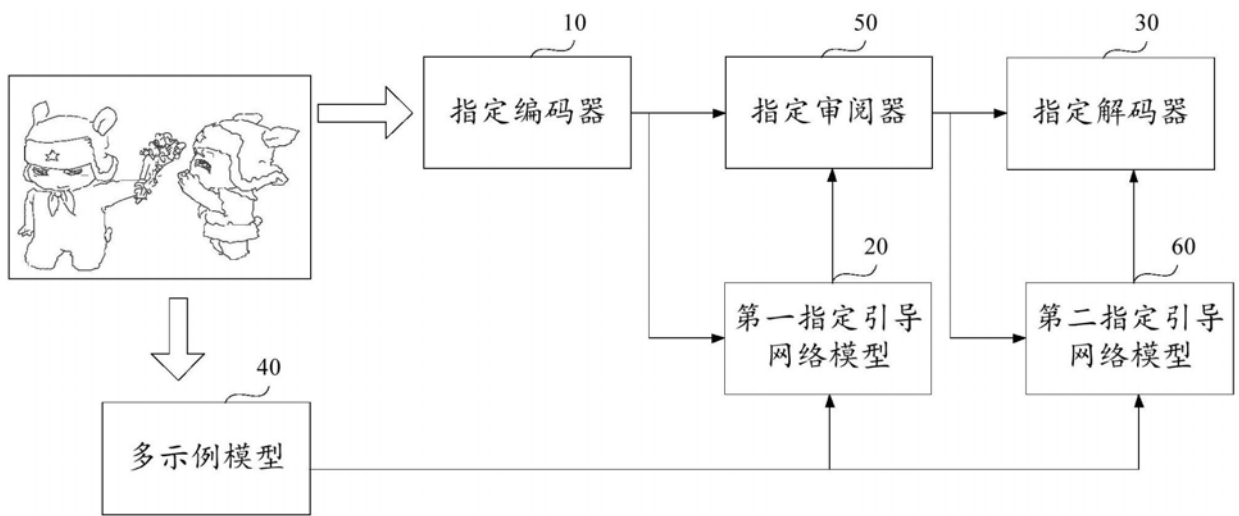


图1F

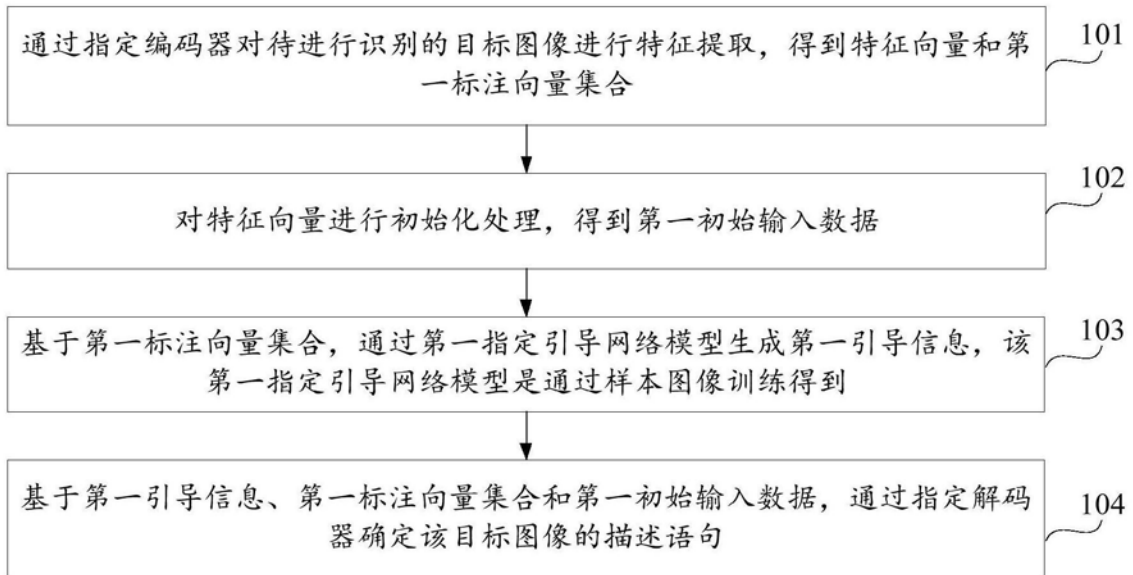


图1G

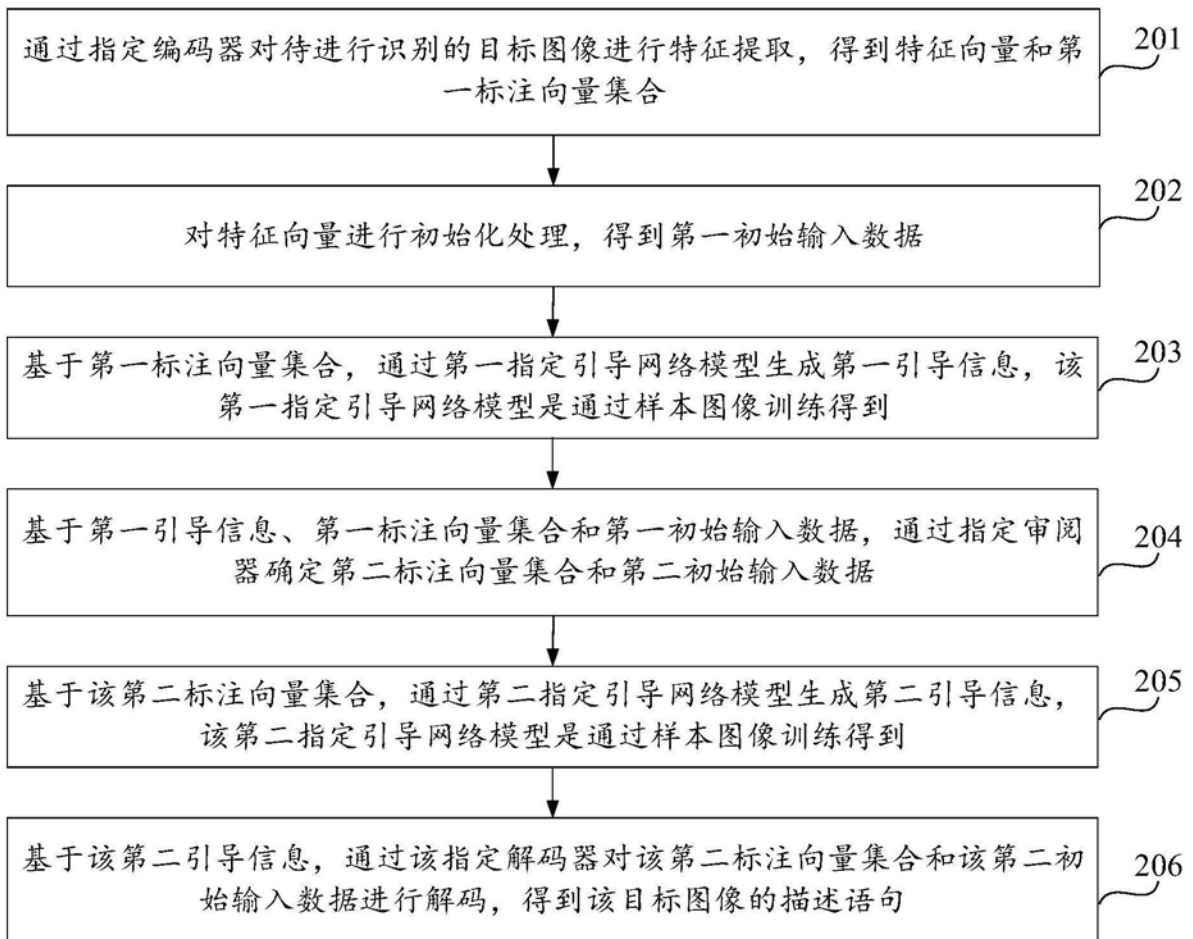


图2

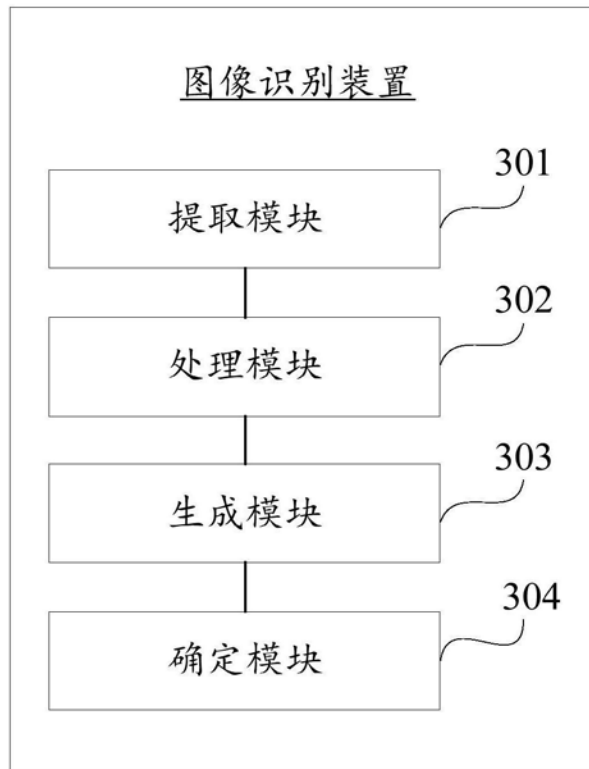


图3A

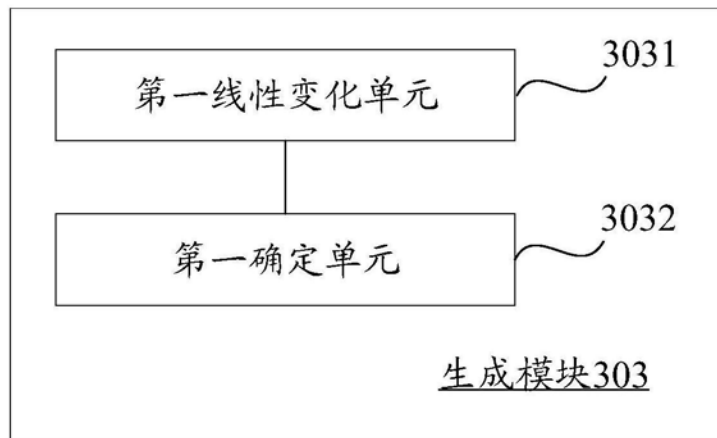


图3B

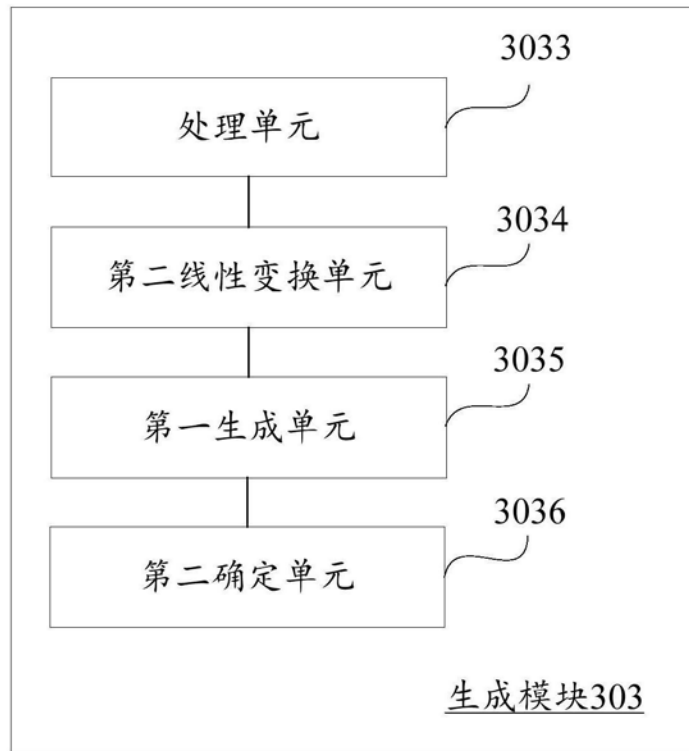


图3C

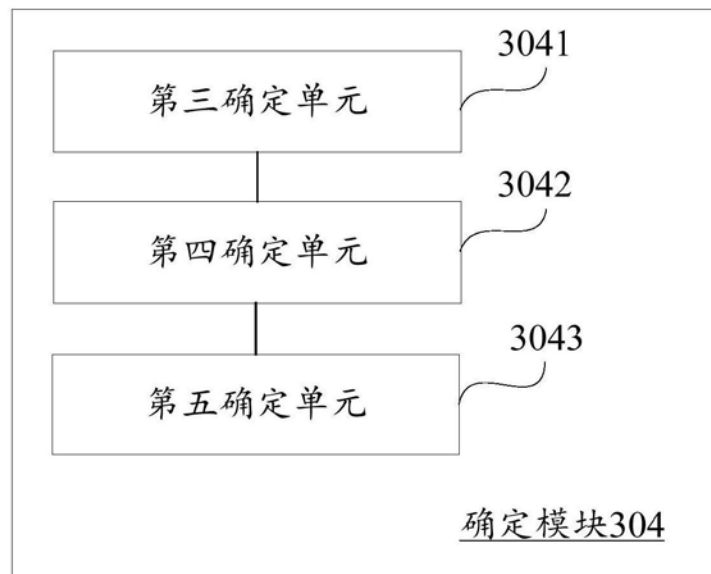


图3D

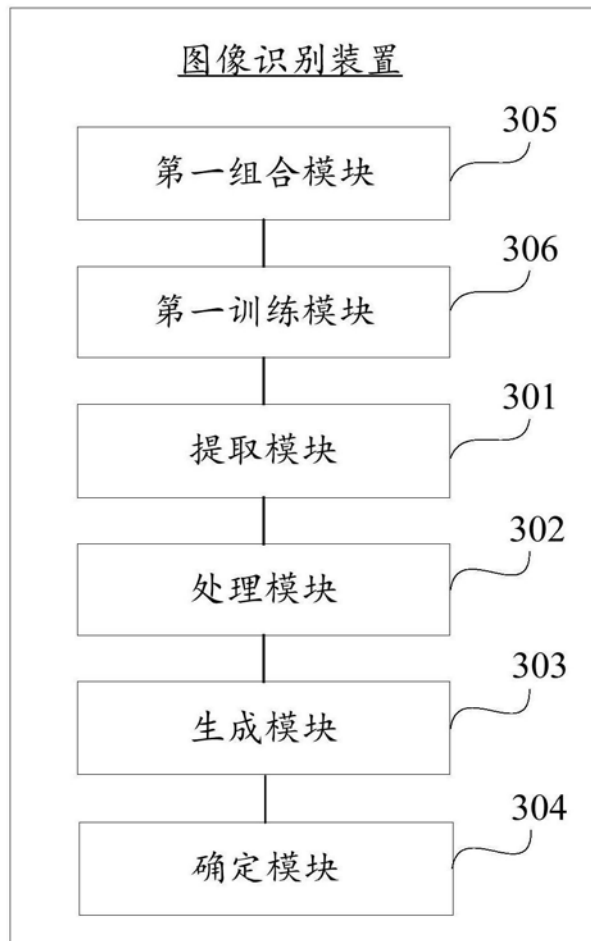


图3E

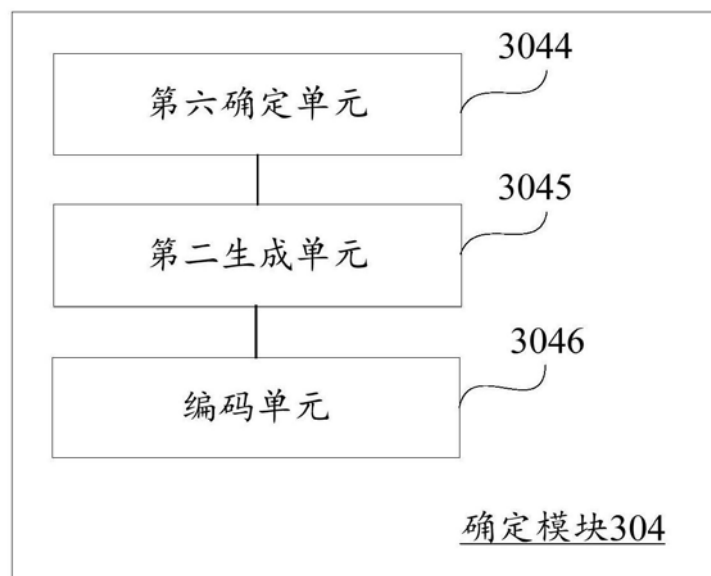


图3F

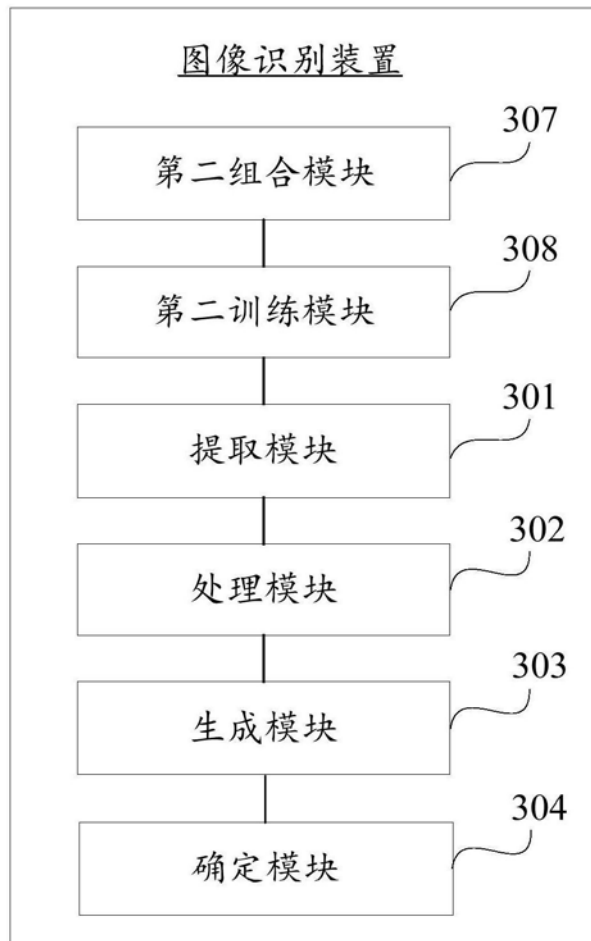


图3G

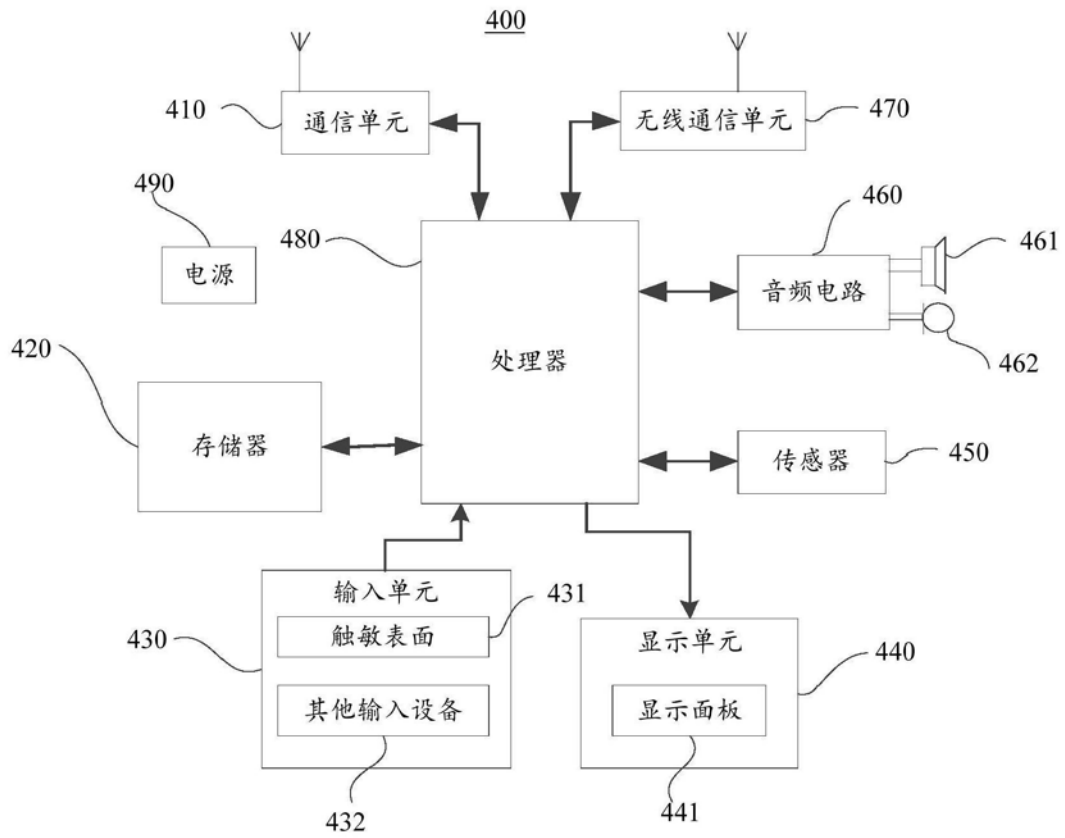


图4