

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 9 月 24 日 (24.09.2020)



(10) 国际公布号
WO 2020/186484 A1

- (51) 国际专利分类号:
G06N 3/04 (2006.01) *G06K 9/46* (2006.01)
G06N 3/08 (2006.01)
- (21) 国际申请号: PCT/CN2019/078915
- (22) 国际申请日: 2019 年 3 月 20 日 (20.03.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 深圳大学(SHENZHEN UNIVERSITY) [CN/CN]; 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。
- (72) 发明人: 王娜(WANG, Na); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。

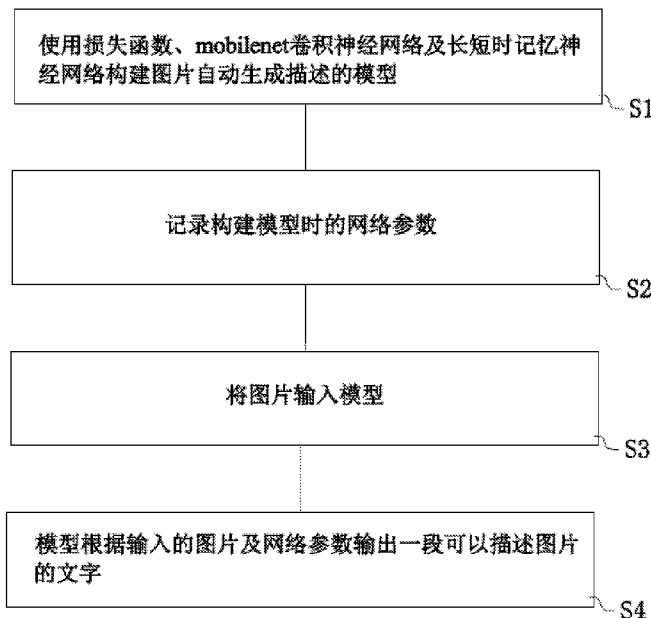
吕锦涛(LV, Jintao); 中国广东省深圳市南山区南海大道 3688 号, Guangdong 518060 (CN)。

(74) 代理人: 深圳市恒申知识产权事务所(普通合伙)(HENSEN INTELLECTUAL PROPERTY FIRM); 中国广东省深圳市福田区南园路 68 号上步大厦 10H, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** AUTOMATIC IMAGE DESCRIPTION GENERATION METHOD AND SYSTEM, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 图片自动生成描述的方法、系统、电子装置及存储介质



- S1 Use a loss function, a mobilenet convolutional neural network, and a long short-term memory network to construct an automatic image description generation model
- S2 Record a network parameter during the construction of the model
- S3 Input an image to model
- S4 The model outputs, according to the inputted image and the network parameter, a piece of text for describing the image

图 1

(57) **Abstract:** Disclosed are an automatic image description generation method and system, an electronic device, and a storage medium, for generating a text description for an image, and resolving the prior art issue in which a trained model does not meet an assessment criteria. The method comprises: using a loss function, a mobilenet convolutional neural network, and a long short-term memory network to construct an automatic image description generation model; recording a network parameter during construction of the model; inputting an image to the model; and the model outputting, according to the inputted image and the network parameter, a piece of text

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

for describing the image. The automatic image description generation method provided by the present application develops a novel loss function by means of reinforced learning, such that model training standards meet the assessment criteria.

(57) 摘要: 本发明公开了一种图片自动生成描述的方法、系统、电子装置及存储介质, 用于对图片生成文字描述, 解决了现有技术中训练的模型与评估标准不统一的问题, 包括: 使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型; 记录构建所述模型时的网络参数; 将图片输入所述模型; 所述模型根据所述输入的图片及所述网络参数输出一段可以描述图片的文字; 因此本申请提供的图片自动生成描述的方法, 通过强化学习设计新的损失函数, 使得模型的训练标准与评估标准能够统一。

图片自动生成描述的方法、系统、电子装置及存储介质技术领域

[0001] 本发明涉及图片处理技术领域，尤其涉及一种图片自动生成描述的方法、系统、电子装置及存储介质。

背景技术

[0002] 进入二十一世纪以来，互联网存储水平和计算机运算能力都经历了一个巨大的飞跃，智能手机用户的数量也在很大程度上得到了提升，用户通过手机终端、PC机等智能设备，每天都在互联网上共享大量的图片数据。这些庞大的图片数据资源可以更精准地反映客观世界，它们在视觉上产生的不同变化，蕴含了丰富的语义信息，为感知现实世界提供了充足的信息来源。

[0003] 随着神经网络和深度学习的发展，图片理解由原始的基于低层视觉特征的图片处理向高层次的基于图片语义信息和语义理解的方向发展。现有的神经网络模型一般是encoder-decoder的模型，即通过CNN(卷积神经网络)提取特征，将提取到的特征作为LSTM(长短时记忆神经网络)的初始状态，由LSTM生成一段可以描述该图片的文字。

发明概述

技术问题

[0004] 现有的模型在训练时，主要使用的是交叉熵损失函数，而在模型训练完成后，需要使用BLUE等评估指标对模型进行评估，但是使用交叉熵损失函数训练的模型，普遍存在训练标准与评估标准不统一的问题。

问题的解决方案

技术解决方案

[0005] 本发明第一方面提供一种图片自动生成描述的方法，包括：使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型；记录构建所述模型时的网络参数；将图片输入所述模型；所述模型根据所述输入的图片及所述网络参数输出一段可以描述图片的文字。

发明的有益效果

有益效果

- [0006] 使用mobilenet卷积神经网络能够提取图像特征，并且参数量和计算量都较少，能够有效提高效率，在使用长短时记忆神经网络后，能够生成图像特征的文本信息，并且整体使用损失函数进行反向传播网络参数来更新和完善模型，从而使得训练出的模型在生成图片描述的时候，数据能够更加准确，从而增加了训练的模型与评估标准的统一性。

对附图的简要说明

附图说明

- [0007] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域技术人员来讲，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。
- [0008] 图1为本发明实施例图片自动生成描述的方法的流程示意框图；
- [0009] 图2为本发明实施例图片自动生成描述的方法中Mobilenet模型的矩阵分解原理图；
- [0010] 图3为本发明实施例图片自动生成描述的系统的结构示意框图；
- [0011] 图4为本发明实施例电子装置的结构示意框图。

实施该发明的最佳实施例

本发明的最佳实施方式

- [0012] 本发明第一方面提供一种图片自动生成描述的方法，包括：使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型；记录构建所述模型时的网络参数；将图片输入所述模型；所述模型根据所述输入的图片及所述网络参数输出一段可以描述图片的文字。
- [0013] 进一步地，所述使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型包括：采用强化学习的思想设计并构建损失函数；使用预先训练的mobilenet卷积神经网络提取图片特征向量；使用矩阵变换加

入注意力机制，以将所述特征向量与从预先训练的长短时记忆神经网络中提取的原状态向量融合，得到新特征向量；将所述新特征向量输入长短时记忆神经网络，所述长短时记忆神经网络根据所述新特征向量生成具有图片信息的文本；对所述损失函数求导，并将所述文本输入至求导后的损失函数中；判断所述损失函数在接收所述文本后是否收敛；若所述损失函数收敛，则保存训练时的网络参数；若所述损失函数不收敛，则继续使用mobilenet卷积神经网络、矩阵变换及长短时记忆神经网络生成具有图片信息的文本，并将文本输入至求导后的损失函数中，直至将所述文本输入所述损失函数中后，所述损失函数收敛。

[0014] 进一步地，所述使用预先训练的mobilenet卷积神经网络提取图片特征向量包括：在将图片输入到预先训练的mobilenet卷积神经网络中后，保存mobilenet卷积神经网络的平均池化层输出的特征向量。

[0015] 本发明第二方面提供一种图片自动生成描述的系统，包括：模型训练模块，用于使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型；网络参数记录模块，用于记录构建所述模型训练模块训练模型时的网络参数；图片接收模块，用于接收输入所述模型训练模块训练的模型的图片；文字生成模块，用于使所述模型根据所述图片接收模块接收的图片及所述网络参数记录模块记录的网络参数输出一段可以描述图片的文字。

[0016] 进一步地，所述神经网络预训练模块包括：神经网络构建单元，用于构建mobilenet卷积神经网络；神经网络参数更新单元，用于根据现有的图片数据集对所述神经网络构建单元构建的mobilenet卷积神经网络进行预训练，更新mobilenet卷积神经网络的参数。

[0017] 进一步地，所述模型训练模块包括：损失函数构建单元，用于采用强化学习的思想设计并构建损失函数；特征向量提取单元，用于使用预先训练的mobilenet卷积神经网络提取样本图片的特征向量；注意力机制引入单元，用于使用矩阵变换引入注意力机制，以将所述特征向量与从预先训练的长短时记忆神经网络中提取的原状态向量融合，得到新特征向量；文本生成单元，用于将所述注意力机制引入单元得到的新特征向量输入长短时记忆神经网络，所述长短时记忆神经网络根据所述新特征向量生成具有图片信息的文本；损失函数求导单元，

用于对所述损失函数构建单元构建的损失函数求导，并将所述文本输入至求导后的损失函数中；损失函数收敛判断单元，用于判断所述损失函数求导单元求导后的损失函数是否收敛；网络参数保存单元，用于在所述损失函数收敛判断单元判断损失函数收敛后，保存训练时的网络参数；循环单元，用于在所述损失函数收敛判断单元判断损失函数不收敛后，继续使用所述特征向量提取单元、所述注意力机制引入单元及所述文本生成单元生成具有图片信息的文本，并将文本输入至所述损失函数求导单元得到的求导后的损失函数中，直至将所述文本输入所述损失函数中后，所述损失函数收敛。

[0018] 进一步地，所述特征向量提取单元包括：池化层输出保存子单元，用于在将图片输入到预先训练的mobilenet卷积神经网络中后，保存mobilenet卷积神经网络的平均池化层输出的特征向量。

[0019] 本发明第三方面提供一种电子装置，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时，实现上述中的任意一项所述方法。

[0020] 本发明第四方面提供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，实现上述中的任意一项所述方法。

发明实施例

本发明的实施方式

[0021] 为使得本发明的发明目的、特征、优点能够更加的明显和易懂，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而非全部实施例。基于本发明中的实施例，本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0022] 请参阅图1，为一种图片自动生成描述的方法，包括：S1、使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型；S2、记录构建模型时的网络参数；S3、将图片输入模型；S4、模型根据输入的图片及网络参数输出一段可以描述图片的文字。

[0023] 使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生

成描述的模型包括：采用强化学习的思想设计并构建损失函数；使用预先训练的mobilenet卷积神经网络提取图片特征向量；使用矩阵变换加入注意力机制，以将特征向量与从预先训练的长短时记忆神经网络中提取的原状态向量融合，得到新特征向量；将新特征向量输入长短时记忆神经网络，长短时记忆神经网络根据新特征向量生成具有图片信息的文本；对损失函数求导，并将文本输入至求导后的损失函数中；判断损失函数在接收文本后是否收敛；若损失函数收敛，则保存训练时的网络参数；若损失函数不收敛，则继续使用mobilenet卷积神经网络、矩阵变换及长短时记忆神经网络生成具有图片信息的文本，并将文本输入至求导后的损失函数中，直至将文本输入损失函数中后，损失函数收敛。

[0024] 使用预先训练的mobilenet卷积神经网络提取图片特征向量包括：在将图片输入到预先训练的mobilenet卷积神经网络中后，保存mobilenet卷积神经网络的平均池化层输出的特征向量。

[0025] Mobilenet重点在压缩模型，同时保证精度。其思想就是，分解一个标准的卷积为一个深度卷积和一个 1×1 的普通卷积(也叫pointwise卷积)。简单理解就是矩阵的因式分解，具体步骤如图2所示。假设，输入的图片特征向量大小为 $DF * DF$ ，维度为 M ，滤波器的大小为 $DK * DK$ ，维度为 N ，并假设步长为1。则原始的卷积操作，需要进行的矩阵运算次数为 $DK * DK * M * N * DF * DF$ ，卷积核参数为 $DK * DK * N$ 。Mobilenet中的卷积需要进行的矩阵运算次数为 $DK * DK * M * DF * DF + M * N * DF * DF$ ，卷积核参数为 $DK * DK * M + N$ 。由于卷积的过程，主要是一个空间维度减少，通道维度增加的过程，即 $N > M$ ，所以， $DK * DK * N > DK * DK * M + N$ 。因此，深度可分离卷积在模型大小上和模型计算量上都进行了大量的压缩，使得模型速度快，计算开销少，准确性好。

[0026] 将新状态向量作为长短时记忆神经网络下一时刻状态的输入，并使用强化学习及反向传播的方法重新训练预先训练的长短时记忆神经网络，得到新长短时记忆神经网络包括：使用预先训练的长短时记忆神经网络根据图片生成的句子作为baseline，将随机选择的单词组成的句子作为reward，使用baseline减去reward构建损失函数；使用损失函数及新状态向量，并通过反向传播再次对长短时记

忆神经网络进行训练，得到新长短时记忆神经网络，并更新新长短时记忆神经网络的训练参数。

[0027] 关于强化学习的方法：把序列问题看作是一个强化学习的问题：

[0028] Agent: LSTM;

[0029] 环境状态: 单词和图片特征;

[0030] 行动: 预测下一个单词;

[0031] 状态: LSTM的单元和隐含层的状态;

[0032] 奖励: CIDEr评分;

[0033] 训练目标是最小化负的期望 $L(\theta)$:

[0034] $L(\theta) = -E_{ws \sim p_\theta}[r(ws)]$ $L(\theta) = -E_{ws \sim p_\theta}[r(ws)] = -\sum r(ws)p_\theta(ws) = -\sum r(ws)p_\theta(ws)$;

[0035] $ws = (ws_1, \dots, ws_T)$ $ws = (w_1s, \dots, w_Ts)$ 是生成的句子。

[0036] 实际上， ws 可以依据 p_θ 的概率来进行随即选择（而不是选择概率最大的那一个）， $L(\theta)$ 可以近似为：

[0037] $L(\theta) \approx -r(ws), ws \sim p_\theta$;

[0038] L 关于 θ 的梯度为：

[0039] $\nabla_\theta L(\theta) = -E_{ws \sim p_\theta}[r(ws) \nabla_\theta \log p_\theta(ws)]$;

[0040] 再引入一个baseline来减少方差：

[0041] $\nabla_\theta L(\theta) = -E_{ws \sim p_\theta}[(r(ws) - b) \nabla_\theta \log p_\theta(ws)]$;

[0042] baseline可以是任意函数，只要它不依赖行动 ws ，引入它并不会改变梯度的值。

[0043] 实际上 $L(\theta)$ 可以被近似为：

[0044] $\nabla_\theta L(\theta) \approx -(r(ws) - b) \nabla_\theta \log p_\theta(ws)$;

[0045] 应用链式法则，梯度可以表示为：

[0046] $\partial L(\theta) / \partial st \approx (r(ws) - b)(p_\theta(w_t | h_t) - 1 w_{st})$;

[0047] 强化学习的思想就是用当前模型在测试阶段生成的词的reward作为baseline，梯度就变成了：

[0048] $\partial L(\theta) / \partial st \approx (r(ws) - r(w^\wedge))(p_\theta(w_t | h_t) - 1 w_{st})$;

[0049] 其中 $r(w^\wedge) = \arg\max_{wt} p(w_t | h_t)$ ，就是在测试阶段使用贪婪解码取概率最大的词来

生成句子；而 $r(ws)$ 是通过根据概率来随机选择的词，如果当前概率最大的词的概率为60%，那就有60%的概率选到它，而不是像贪婪解码一样100%选概率最大的。

[0050] 公式的意思就是：对于如果当前随机到的词比测试阶段生成的词好，那么在这次词的维度上，整个式子的值就是负的（因为后面那一项一定为负），这样梯度就会上升，从而提高这个词的分数；而对于其他词，后面那一项为正，梯度就会下降，从而降低其他词的分数。

[0051] 请参阅图3，为一种图片自动生成描述的系统，包括：模型训练模块1、网络参数记录模块2、图片接收模块3及文字生成模块4；模型训练模块1用于使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型；网络参数记录模块2用于记录构建模型训练模块1训练模型时的网络参数；图片接收模块3用于接收输入模型训练模块1训练的模型的图片；文字生成模块4用于使模型根据图片接收模块3接收的图片及网络参数记录模块2记录的网络参数输出一段可以描述图片的文字。

[0052] 模型训练模块1包括：损失函数构建单元、特征向量提取单元、注意力机制引入单元、文本生成单元、损失函数求导单元、损失函数收敛判断单元、网络参数保存单元及循环单元；损失函数构建单元用于采用强化学习的思想设计并构建损失函数；特征向量提取单元用于使用预先训练的mobilenet卷积神经网络提取样本图片的特征向量；注意力机制引入单元用于使用矩阵变换引入注意力机制，以将特征向量与从预先训练的长短时记忆神经网络中提取的原状态向量融合，得到新特征向量；文本生成单元用于将注意力机制引入单元得到的新特征向量输入长短时记忆神经网络，长短时记忆神经网络根据新特征向量生成具有图片信息的文本；损失函数求导单元用于对损失函数构建单元构建的损失函数求导，并将文本输入至求导后的损失函数中；损失函数收敛判断单元用于判断损失函数求导单元求导后的损失函数是否收敛；网络参数保存单元用于在损失函数收敛判断单元判断损失函数收敛后，保存训练时的网络参数；循环单元用于在损失函数收敛判断单元判断损失函数不收敛后，继续使用特征向量提取单元、注意力机制引入单元及文本生成单元生成具有图片信息的文本，并将文本

输入至损失函数求导单元得到的求导后的损失函数中，直至将文本输入损失函数中后，损失函数收敛。

[0053] 特征向量提取单元包括：池化层输出保存子单元，用于在将图片输入到预先训练的mobilenet卷积神经网络中后，保存mobilenet卷积神经网络的平均池化层输出的特征向量。

[0054] 本申请实施例提供一种电子装置，请参阅4，该电子装置包括：存储器601、处理器602及存储在存储器601上并可在处理器602上运行的计算机程序，处理器602执行该计算机程序时，实现前述中描述的图片自动生成描述的方法。

[0055] 进一步的，该电子装置还包括：至少一个输入设备603以及至少一个输出设备604。

[0056] 上述存储器601、处理器602、输入设备603以及输出设备604，通过总线605连接。

[0057] 其中，输入设备603具体可为摄像头、触控面板、物理按键或者鼠标等等。输出设备604具体可为显示屏。

[0058] 存储器601可以是高速随机存取记忆体（RAM，Random Access Memory）存储器，也可为非不稳定的存储器（non-volatile memory），例如磁盘存储器。存储器601用于存储一组可执行程序代码，处理器602与存储器601耦合。

[0059] 进一步的，本申请实施例还提供了一种计算机可读存储介质，该计算机可读存储介质可以是设置于上述各实施例中的电子装置中，该计算机可读存储介质可以是前述中的存储器601。该计算机可读存储介质上存储有计算机程序，该程序被处理器602执行时实现前述实施例中描述的图片自动生成描述的方法。

[0060] 进一步的，该计算机可存储介质还可以是U盘、移动硬盘、只读存储器601（ROM，Read-Only Memory）、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0061] 在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个模块或组件可以结合或者可以集成到另一个系统，或一些特征可

以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或模块的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0062] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，即可以位于一个地方，或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0063] 另外，在本发明各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。

[0064] 需要说明的是，对于前述的各方法实施例，为了简便描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本发明并不受所描述的动作顺序的限制，因为依据本发明，某些步骤可以采用其它顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定都是本发明所必须的。

[0065] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其它实施例的相关描述。

[0066] 以上为对本发明所提供的一种图片自动生成描述的方法、系统、电子装置及存储介质的描述，对于本领域的技术人员，依据本发明实施例的思想，在具体实施方式及应用范围上均会有改变之处，综上，本说明书内容不应理解为对本发明的限制。

工业实用性

[0067] 本发明提供了一种图片自动生成描述的方法、系统、电子装置及存储介质，解决了现有技术中训练的模型与评估标准不统一的技术问题。

权利要求书

- [权利要求 1] 一种图片自动生成描述的方法，其特征在于，包括：
- 使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型；
- 记录构建所述模型时的网络参数；
- 将图片输入所述模型；
- 所述模型根据所述输入的图片及所述网络参数输出一段可以描述图片的文字。
- [权利要求 2] 根据权利要求1所述的图片自动生成描述的方法，其特征在于，
- 所述使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型包括：
- 采用强化学习的思想设计并构建损失函数；
- 使用预先训练的mobilenet卷积神经网络提取图片特征向量；
- 使用矩阵变换加入注意力机制，以将所述特征向量与从预先训练的长短时记忆神经网络中提取的原状态向量融合，得到新特征向量；
- 将所述新特征向量输入长短时记忆神经网络，所述长短时记忆神经网络根据所述新特征向量生成具有图片信息的文本；
- 对所述损失函数求导，并将所述文本输入至求导后的损失函数中；
- 判断所述损失函数在接收所述文本后是否收敛；
- 若所述损失函数收敛，则保存训练时的网络参数；
- 若所述损失函数不收敛，则继续使用mobilenet卷积神经网络、矩阵变换及长短时记忆神经网络生成具有图片信息的文本，并将文本输入至求导后的损失函数中，直至将所述文本输入所述损失函数中后，所述损失函数收敛。
- [权利要求 3] 根据权利要求2所述的图片自动生成描述的方法，其特征在于，
- 所述使用预先训练的mobilenet卷积神经网络提取图片特征向量包括：
- 在将图片输入到预先训练的mobilenet卷积神经网络中后，保存mobilenet卷积神经网络的平均池化层输出的特征向量。

- [权利要求 4] 一种图片自动生成描述的系统，其特征在于，包括：
- 模型训练模块，用于使用损失函数、mobilenet卷积神经网络及长短时记忆神经网络构建图片自动生成描述的模型；
- 网络参数记录模块，用于记录构建所述模型训练模块训练模型时的网络参数；
- 图片接收模块，用于接收输入所述模型训练模块训练的模型的图片；
- 文字生成模块，用于使所述模型根据所述图片接收模块接收的图片及所述网络参数记录模块记录的网络参数输出一段可以描述图片的文字。
- [权利要求 5] 根据权利要求4所述的图片自动生成描述的系统，其特征在于，
- 所述模型训练模块包括：
- 损失函数构建单元，用于采用强化学习的思想设计并构建损失函数；
- 特征向量提取单元，用于使用预先训练的mobilenet卷积神经网络提取样本图片的特征向量；
- 注意力机制引入单元，用于使用矩阵变换引入注意力机制，以将所述特征向量与从预先训练的长短时记忆神经网络中提取的原状态向量融合，得到新特征向量；
- 文本生成单元，用于将所述注意力机制引入单元得到的新特征向量输入长短时记忆神经网络，所述长短时记忆神经网络根据所述新特征向量生成具有图片信息的文本；
- 损失函数求导单元，用于对所述损失函数构建单元构建的损失函数求导，并将所述文本输入至求导后的损失函数中；
- 损失函数收敛判断单元，用于判断所述损失函数求导单元求导后的损失函数是否收敛；
- 网络参数保存单元，用于在所述损失函数收敛判断单元判断损失函数收敛后，保存训练时的网络参数；
- 循环单元，用于在所述损失函数收敛判断单元判断损失函数不收敛后，继续使用所述特征向量提取单元、所述注意力机制引入单元及所述

文本生成单元生成具有图片信息的文本，并将文本输入至所述损失函数求导单元得到的求导后的损失函数中，直至将所述文本输入所述损失函数中后，所述损失函数收敛。

- [权利要求 6] 根据权利要求5所述的图片自动生成描述的系统，其特征在于，所述特征向量提取单元包括：
- 池化层输出保存子单元，用于在将图片输入到预先训练的mobilenet卷积神经网络中后，保存mobilenet卷积神经网络的平均池化层输出的特征向量。
- [权利要求 7] 一种电子装置，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，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3中的任意一项所述方法。
- [权利要求 8] 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3中的任意一项所述方法。

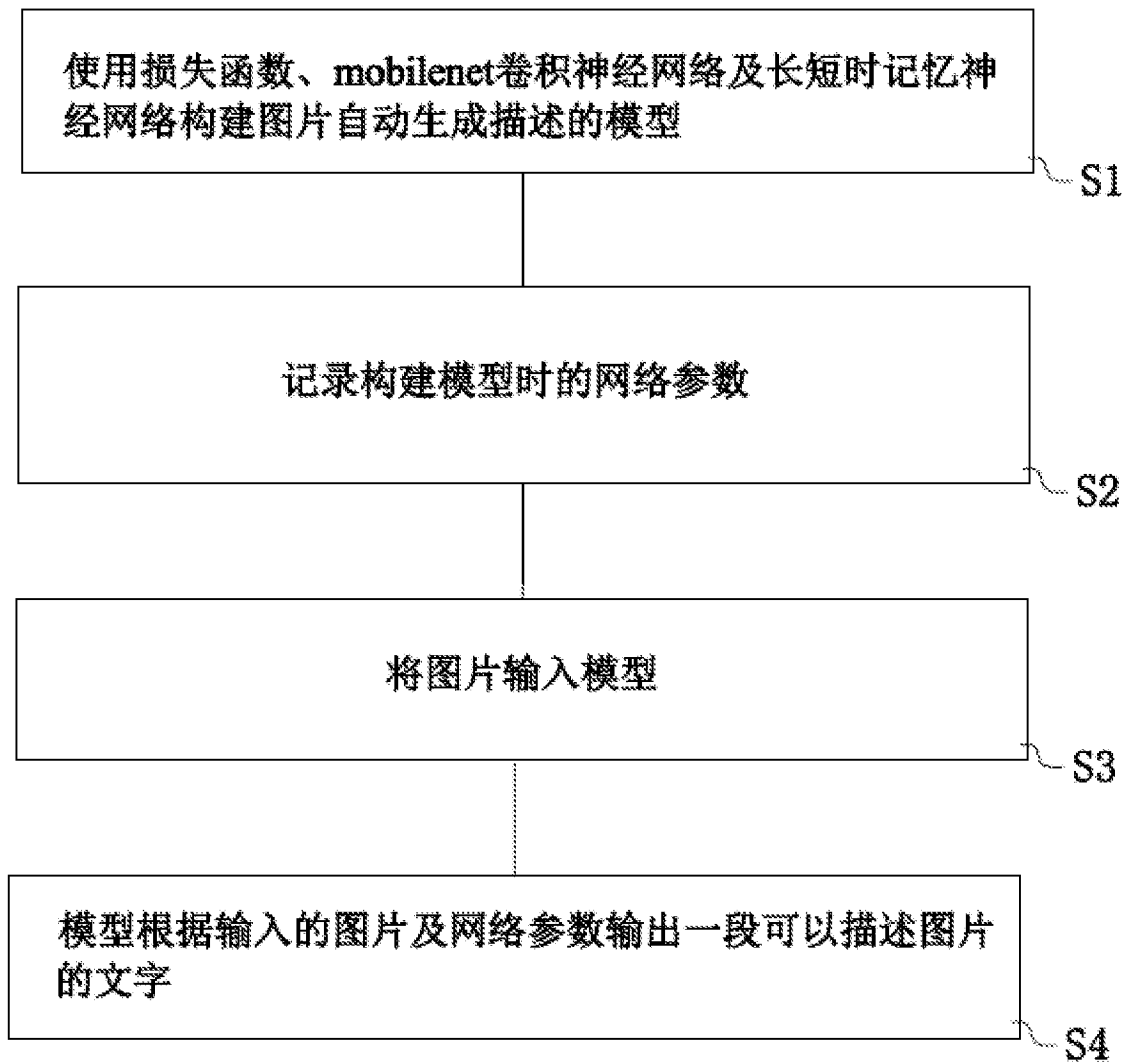


图 1

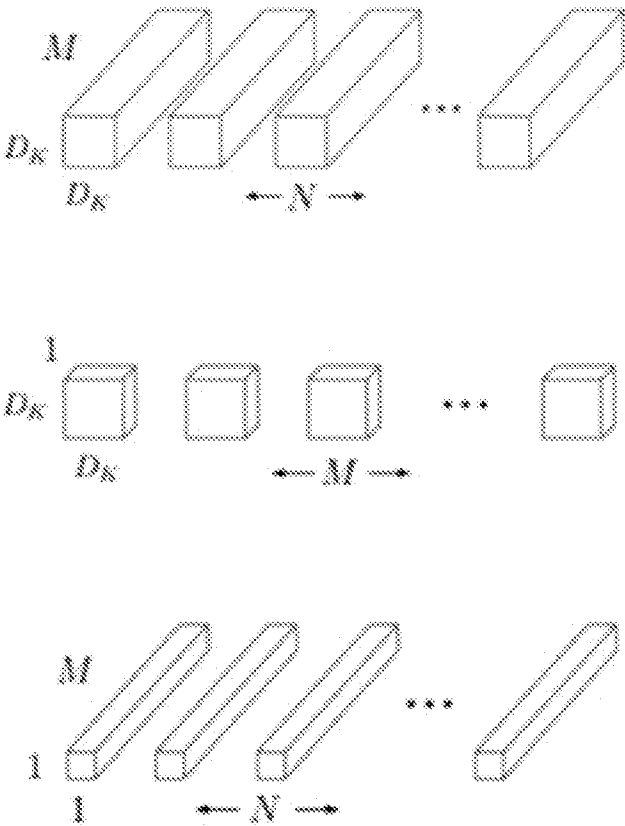


图 2

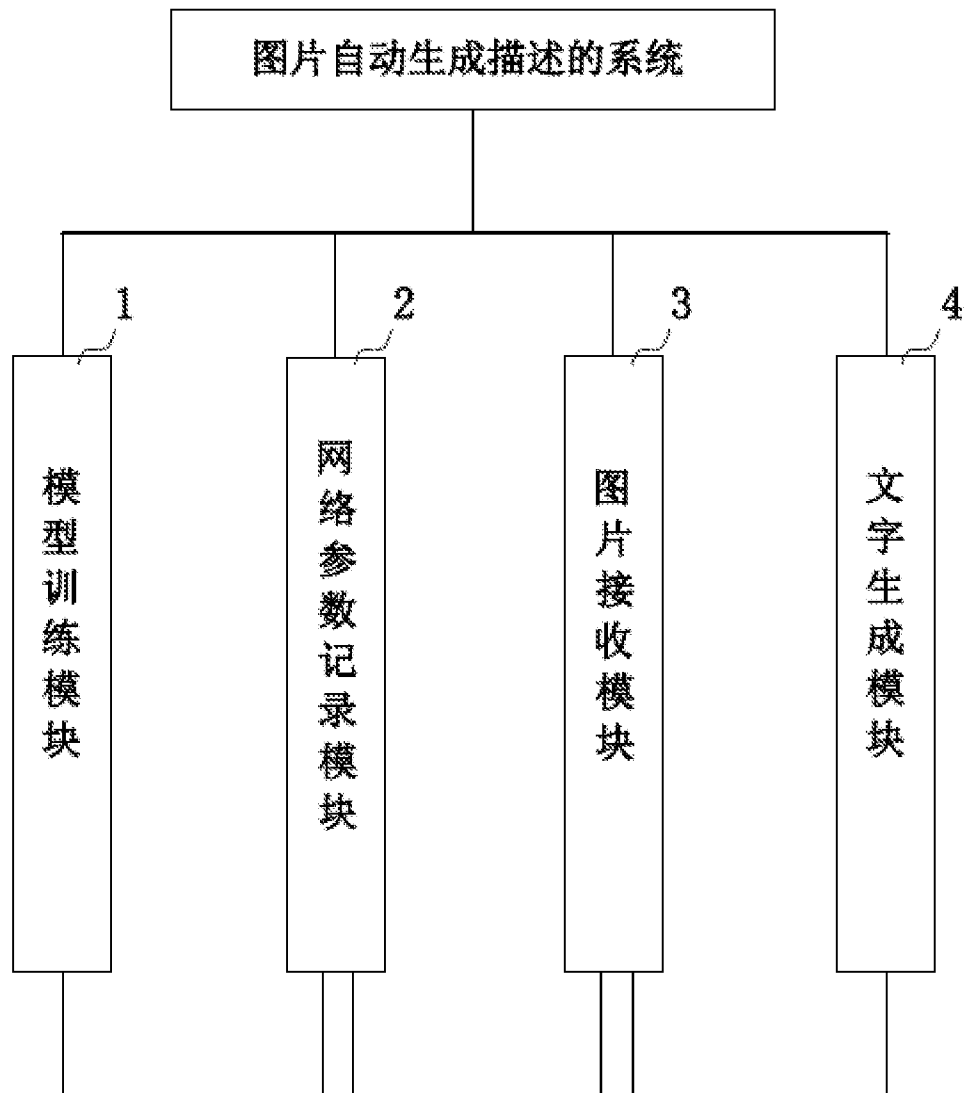


图 3

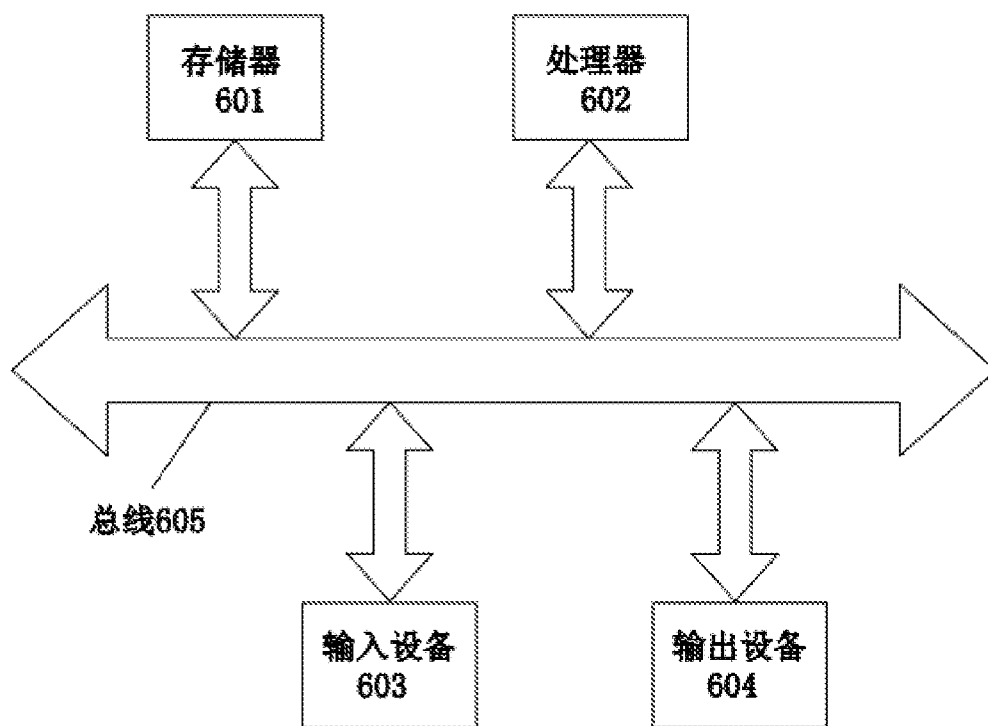


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/078915

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/04(2006.01)i; G06N 3/08(2006.01)i; G06K 9/46(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N; G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKI, VEN: 图片, 照片, 图像, 图象, 相片, 描述, 说明, 介绍, 摘要, 文字, 损失, 卷积, 神经网络, 长短期记忆, image, photo?, picture?, abstract, description, annotation, loss, convolutional, neural 1w network, LSTM, long 1w short 1w term 1w memory, mobilenet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108052512 A (TONGJI UNIVERSITY) 18 May 2018 (2018-05-18) description, paragraphs 6-96	1-8
X	CN 108898639 A (HUBEI UNIVERSITY OF TECHNOLOGY) 27 November 2018 (2018-11-27) claims 1-9, and description, paragraphs 5-52	1-8
X	CN 107665356 A (BEIJING INFORMATION SCIENCE & TECHNOLOGY UNIVERSITY) 06 February 2018 (2018-02-06) description, paragraphs 38-65	1-8
A	US 2018/0204112 A1 (GOOGLE LLC) 19 July 2018 (2018-07-19) entire document	1-8

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

13 December 2019

Date of mailing of the international search report

20 December 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/078915

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108052512	A	18 May 2018	None			
CN	108898639	A	27 November 2018	None			
CN	107665356	A	06 February 2018	None			
US	2018/0204112	A1	19 July 2018	US	10417557	B2	17 September 2019

国际检索报告

国际申请号

PCT/CN2019/078915

A. 主题的分类

G06N 3/04(2006.01)i; G06N 3/08(2006.01)i; G06K 9/46(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06N; G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS, CNTXT, CNKI, VEN:图片, 照片, 图像, 图象, 相片, 描述, 说明, 介绍, 摘要, 文字, 损失, 卷积, 神经网络, 长短时记忆, image, photo??. picture?, abstract, description, annotation, loss, convolutional, neural lw network, LSTM, long lw short lw term lw memory, mobilenet

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 108052512 A (同济大学) 2018年 5月 18日 (2018 - 05 - 18) 说明书第6-96段	1-8
X	CN 108898639 A (湖北工业大学) 2018年 11月 27日 (2018 - 11 - 27) 权利要求1-9, 说明书第5-52段	1-8
X	CN 107665356 A (北京信息科技大学) 2018年 2月 6日 (2018 - 02 - 06) 说明书第38-65段	1-8
A	US 2018/0204112 A1 (GOOGLE LLC) 2018年 7月 19日 (2018 - 07 - 19) 全文	1-8

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 12月 13日

国际检索报告邮寄日期

2019年 12月 20日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

孙泽站

传真号 (86-10)62019451

电话号码 62089929

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/078915

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	108052512	A	2018年 5月 18日	无			
CN	108898639	A	2018年 11月 27日	无			
CN	107665356	A	2018年 2月 6日	无			
US	2018/0204112	A1	2018年 7月 19日	US	10417557	B2	2019年 9月 17日