

(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
2025 年 1 月 16 日 (16.01.2025)



(10) 国际公布号
WO 2025/010945 A1

(51) 国际专利分类号:
G06F 16/31 (2019.01) **G06V 30/413** (2022.01)

(21) 国际申请号: PCT/CN2023/138801

(22) 国际申请日: 2023 年 12 月 14 日 (14.12.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310833232.4 2023年7月7日 (07.07.2023) CN

(71) 申请人: 中国电信股份有限公司技术创新中心 (**CHINA TELECOM CORPORATION LIMITED TECHNOLOGY INNOVATION CENTER**) [CN/CN]; 中国北京市昌平区北七家镇未来科技城南区中国电信北京信息科技创新园 11 层 1118 室、1116 室, Beijing 102209 (CN)。中国电信

股份有限公司(**CHINA TELECOM CORPORATION LIMITED**) [CN/CN]; 中国北京市西城区金融大街 31 号, Beijing 100033 (CN)。

(72) 发明人: 王昊 (**WANG, Hao**); 中国北京市西城区金融大街 31 号, Beijing 100033 (CN)。杨明川 (**YANG, Mingchuan**); 中国北京市西城区金融大街 31 号, Beijing 100033 (CN)。刘振华 (**LIU, Zhenhua**); 中国北京市西城区金融大街 31 号, Beijing 100033 (CN)。李伟 (**LI, Wei**); 中国北京市西城区金融大街 31 号, Beijing 100033 (CN)。

(74) 代理人: 北京律智知识产权代理有限公司 (**BEIJING INTELLEGAL INTELLECTUAL PROPERTY AGENT LTD.**); 中国北京市朝阳区慧忠路 5 号 B1605、B1606、B1607, Beijing 100101 (CN)。

(54) **Title:** VISUAL QUESTION ANSWERING MODEL TRAINING METHOD AND APPARATUS, AND VISUAL QUESTION ANSWERING TASK PROCESSING METHOD AND APPARATUS

(54) 发明名称: 视觉问答模型的训练、视觉问答任务处理方法及装置

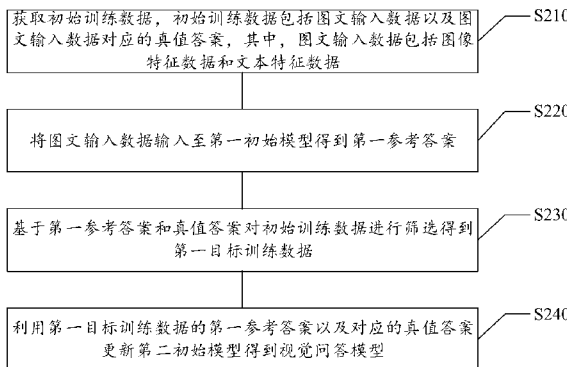


图 2

- S210 Acquire initial training data, the initial training data comprising image-text input data and a true value answer corresponding thereto, and the image-text input data comprising image feature data and text feature data
- S220 Input the image-text input data into a first initial model to obtain a first reference answer
- S230 On the basis of the first reference answer and the true value answer, screen the initial training data to obtain first target training data
- S240 Use the first reference answer of the first target training data and the corresponding true value answer to update a second initial model, so as to obtain a visual question answering model

(57) **Abstract:** The present invention relates to the technical field of machine learning, and relates to a visual question answering model training method and apparatus, a visual question answering task processing method and apparatus, a computer-readable storage medium and an electronic device. The visual question answering model training method comprises: acquiring initial training data, the initial training data comprising image-text input data and a true value answer corresponding thereto, and the image-text input data comprising

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))。

image feature data and text feature data; inputting the image-text input data into a first initial model to obtain a first reference answer; on the basis of the first reference answer and the true value answer, screening the initial training data to obtain first target training data; and using the first reference answer of the first target training data and the corresponding true value answer to update a second initial model, so as to obtain a visual question answering model, the model architecture of the first initial model being consistent with that of the second initial model. The technical solution of the present invention improves the processing precision of visual question answering models.

(57) 摘要: 本发明涉及机器学习技术领域, 涉及一种视觉问答模型的训练方法及装置、视觉问答任务处理方法及装置、计算机可读存储介质及电子设备, 视觉问答模型的训练方法包括: 获取初始训练数据, 初始训练数据包括图文输入数据以及图文输入数据对应的真值答案, 其中, 图文输入数据包括图像特征数据和文本特征数据; 将图文输入数据输入至第一初始模型得到第一参考答案; 基于第一参考答案和真值答案对初始训练数据进行筛选得到第一目标训练数据; 利用第一目标训练数据的第一参考答案以及对应的真值答案更新第二初始模型得到视觉问答模型; 其中, 第一初始模型和第二初始模型的模型架构一致。本发明的技术方案提高了得到视觉问答模型的处理精度。

视觉问答模型的训练、视觉问答任务处理方法及装置

相关申请的交叉引用

本申请要求于 2023 年 07 月 07 日提交的，申请号为 202310833232.4，名称为“视觉问答模型的训练、视觉问答任务处理方法及装置”的中国专利申请的优先权，该中国专利申请的全部内容通过引用结合在本文中。

技术领域

本公开涉及机器学习技术领域，具体而言，涉及一种视觉问答模型的训练方法及装置、视觉问答任务处理方法及装置、计算机可读存储介质及电子设备。

背景技术

视觉问答模型在多模态学习领域有了长足发展，但目前的视觉问答方法都需要准确的数据标签才能构建完善的数据集。

然而，构建视觉问答模型标注数据集需要极大的人力和时间成本，且标注的数据也不可避免地存在着噪声，导致训练得到的视觉问答模型的处理精度较低。

发明内容

本公开的目的在于提供一种视觉问答模型的训练方法、视觉问答模型的训练装置、视觉问答任务处理方法、视觉问答任务处理装置、计算机可读介质和电子设备，进而至少在一定程度上提高了得到视觉问答模型的处理精度。

根据本公开的第一方面，提供一种视觉问答模型的训练方法，包括：获取初始训练数据，初始训练数据包括图文输入数据以及图文输入数据对应的真值答案，其中，图文输入数据包括图像特征数据和文本特征数据；将图文输入数据输入至第一初始模型得到第一参考答案；基于第一参考答案和真值答案对初始训练数据进行筛选得到第一目标训练数据；利用第一目标训练数据的第一参考答案以及对应的真值答案更新第二初始模型得到视觉问答模型；其中，第一初始模型和第二初始模型的模型架构一致。

根据本公开的第二方面，提供一种视觉问答模型的训练装置，包括：数据获取模块，用于获取初始训练数据，初始训练数据包括图文输入数据以及图文输入数据对应的真值答案，其中，图文输入数据包括图像特征数据和文本特征数据；数据处理模块，用于将图文输入数据输入至第一初始模型得到第一参考答案；数据筛选模块，用于基于第一参考答案和真值答案对初始训练数据进行筛选得到第一目标训练数据；模型更新模块，利用第一目标训练数据的第一参考答案以及对应的真值答案更新第二初始模型得到视觉问答模型；其中，第一初始模型和第二初始模型的模型架构一致。

根据本公开的第三方面，提供一种视觉问答任务处理方法，包括：获取待提问图像对应的参考图像特征以及对应的参考文本特征；将参考图像特征和参考文本特征输入至视觉问答模型得到目标答案；其中，视觉问答模型可以根据视觉问答模型的训练方法得到。

根据本公开的第四方面，提供一种视觉问答任务处理装置，包括：特征获取模块，用于获取待提问图像对应的参考图像特征以及对应的参考文本特征；任务处理模块，用于将参考图像特征和参考文本特征输入至视觉问答模型得到目标答案；其中，视觉问答模型是根据视觉问答模型的训练方法得到。

根据本公开的第五方面，提供一种计算机可读介质，其上存储有计算机程序，计算机程序被处理器执行时实现上述的方法。

根据本公开的第六方面，提供一种电子设备，其特征在于，包括：一个或多个处理器；以及存储器，用于存储一个或多个程序，当一个或多个程序被一个或多个处理器执行时，使得一个或多个处理器实现上述的方法。

本公开的一种实施例所提供的视觉问答模型的训练方法，一方面，通过第一初始模型对初始训练数据进行筛选得到第一目标训练数据，提升了视觉问答模型的训练数据的精度，使得得到的视觉问答模型的精度更高；另一方面，通过第一初始模型的输出的第一参考答案来更新第二初始模型的参数得到视觉问答模型，进一步提升了得到的视觉问答模型的抗噪声能力，提升了视觉问答模型的精度。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本公开。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本公开的实施例，并与说明书一起用于解释本公开的原理。显而易见地，下面描述中的附图仅仅是本公开的一些实施例，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。在附图中：

图 1 示出了可以应用本公开实施例的一种示例性系统架构的示意图；

图 2 示意性示出本公开示例性实施例中一种视觉问答模型的训练方法的流程图；

图 3 示意性示出本公开示例性实施例中一种视觉问答模型架构图；

图 4 示意性示出本公开示例性实施例中一种初始子模型的划分图；

图 5 示意性示出本公开示例性实施例中另一种视觉问答模型的训练方法的流程图；

图 6 示意性示出本公开示例性实施例中一种视觉问答模型的训练方法的数据流向图；

图 7 示意性示出本公开示例性实施例中一种视觉问答任务处理方法的流程图；

图 8 示意性示出本公开示例性实施例中一种视觉问答任务处理的数据流向图；

图 9 示意性示出本公开示例性实施例中视觉问答任务处理装置的组成示意图；

图 10 示意性示出本公开示例性实施例中视觉问答模型的训练装置的组成示意图；

图 11 示出了可以应用本公开实施例的一种电子设备的示意图。

具体实施方式

现在将参考附图更全面地描述示例实施方式。然而，示例实施方式能够以多种形式实施，且不应被理解为限于在此阐述的范例；相反，提供这些实施方式使得本公开将更加全面和完整，并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。

此外，附图仅为本公开的示意性图解，并非一定是按比例绘制。图中相同的附图标记表示相同或类似的部分，因而将省略对它们的重复描述。附图中所示的一些方框图是功能实体，不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体，或在一个或多个硬件模块或集成电路中实现这些功能实体，或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

视觉问答是一个实用的机器学习任务，该任务需要一个 AI 模型来输出一个关于图像的视觉问题的答案。该任务的挑战性在于其多任务性和开放性，它涉及同时解决计算机视觉和自然语言理解中的多个技术研究问题。这项任务的进展将使多模态机器学习进行广泛的应用，从帮助盲人和视障人士，或与机器人交流，以增强用户的视觉体验与外部知识。

现有的视觉问答系统存在以下问题：如果没有高质量的视觉图片、问题和回答且多样化的三元组大规模初始训练数据，就较难训练出有效且表现稳定的视觉问答模型。

但是，制备完善的数据集既耗时又繁重。目前视觉问答研究领域主要关注复杂模型的设计，而非真实场景中所需要的训练数据构建。因此，如何使模型在较低质量的标注下进行鲁棒的训练是目前亟需解决的问题。

基于上述缺点，本公开提供一种视觉问答模型的训练方法，图 1 示出了可以实现上述视觉问答模型的训练方法的系统架构的示意图，该系统架构 100 可以包括终端 110 与服务

器 120。其中，终端 110 可以是智能手机、平板电脑、台式电脑、笔记本电脑等终端设备，服务器 120 泛指提供本示例性实施方式中视觉问答相关服务的后台系统，可以是一台服务器或多台服务器形成的集群。终端 110 与服务器 120 之间可以通过有线或无线的通信链路形成连接，以进行数据交互。

5 在一种实施方式中，可以由终端 110 执行上述视觉问答模型的训练方法。例如，用户使用终端 110 获取初始训练数据，初始训练数据包括图文输入数据以及图文输入数据对应的真值答案，其中，图文输入数据包括图像特征数据和文本特征数据，由终端 110 首先将图文输入数据输入至第一初始模型得到第一参考答案，然后基于第一参考答案和真值答案对初始训练数据进行筛选得到第一目标训练数据，最后利用第一目标训练数据的第一参考
10 答案以及对应的真值答案更新第二初始模型得到视觉问答模型；其中，第一初始模型和第二初始模型的模型架构一致。

在一种实施方式中，可以由服务器 120 可以执行上述视觉问答模型的训练方法。例如，用户使用终端 110 获取初始训练数据，初始训练数据包括图文输入数据以及图文输入数据对应的真值答案，其中，图文输入数据包括图像特征数据和文本特征数据后，终端 110 将
15 该初始训练数据上传至服务器 120，由服务器 120 对首先将图文输入数据输入至第一初始模型得到第一参考答案，然后基于第一参考答案和真值答案对初始训练数据进行筛选得到第一目标训练数据，最后利用第一目标训练数据的第一参考答案以及对应的真值答案更新第二初始模型得到视觉问答模型，然后向终端 110 返回视觉问答模型。

由上可知，本示例性实施方式中的视觉问答模型的训练方法的执行主体可以是上述终端
20 110 或服务器 120，本公开对此不做限定。

下面结合图 2 对本示例性实施方式中的视觉问答模型的训练方法进行说明，图 2 示出了该视觉问答模型的训练方法的示例性流程，可以包括步骤 S210 至步骤 S240。

参考图 2，在步骤 S210 中，获取初始训练数据，初始训练数据包括图文输入数据以及
25 图文输入数据对应的真值答案，其中，图文输入数据包括图像特征数据和文本特征数据。

在本公开的一种示例实施方式中，处理器可以首先获取初始训练数据，其中，上述初
30 始训练数据中包括图文输入数据和图文输入数据对应的真值答案，其中，上述图文输入数据可以包括图像特征数据和文本特征数据。

具体的，处理器可以首先获取初始图像、初始文本问题以及初始文本问题对应的真值
35 答案，然后可以对上述初始图像和初始文本问题进行特征提取得到图像特征数据和文本特征数据。

在一种示例实施方式中，可以使用 Faster-RCNN (Regions with CNN features, 区域生成网络) 来对初始图像进行特征提取得到图像特征数据。在本示例实施方式中，使用
40 Faster-RCNN 实现自下向上的关注模型，通过设定的阈值允许兴趣框的重叠能够更有效的理解图像内容，得到更加精确的图像特征数据。

在一种示例实施方式中，可以采用文本特征提取模型来实现对初始问题文本的特征提
35 取，其中，文本特征提取模型可以包括两个 LSTM (Long Short-Term Memory, 长短期记忆网络)，其中包括一个自顶向下注意力机制的 LSTM，其可以帮助图像特征进行重加权。提取特征后要使用 GRU (Gated Recurrent Unit, 门控循环单元) 模块做序列编码处理。

在步骤 S220 中，将图文输入数据输入至第一初始模型得到第一参考答案。

在本公开的一种示例实施方式中，在得到上述图文输入数据之后，可以将上述图文输
40 入数据输入至上述第一初始模型中得到上述第一参考答案。

第一初始模型和第二初始模型的框架结构相同，第一初始模型和第二初始模型中的模型参数可以相同也可以不同，优选的，上述第一初始模型与第二初始模型中的参数不相同，在训练时能够使得第二初始模型吸收第一初始模型的抗噪声能力，使得训练得到的视觉问
45 答模型的精度更高。

在本示例实施方式中，上述第一初始模型和第二初始模型可以是 UpDn 和 GRU 的模型结构，初始训练数据为 $D = \{I_i, Q_i, a_i\}^N$ ，其包含 N 个图像 I_i ，问题 Q_i ，回答 a_i 对。

具体的，参照图 3 所示，上述第一初始模型和第二初始模型可以包括 top-down attention（自顶向下的注意）LSTM 模块 310，attend（加权）模块 320、language（语言）LSTM 模块 330 以及损失函数层 340。

具体的，top-down attention（自顶向下的注意）LSTM 模块的第 t 个时间步长的输入为：

$$x_t^1 = [h_{t-1}^2, \bar{v}, W_e \Pi_t]$$

其中， h_{t-1}^2 为 language（语言）LSTM 模块上一轮的输出， $\bar{v} = \frac{1}{K} \sum_i v_i$ 为图像特征数据的平均值， W_e 是问答文本的 embedding 矩阵， Π_t 是当前轮输入的 one-hot 编码的词，分别为 top-down attention（自顶向下的注意）LSTM 模块提供了当前语言模型的文本信息、图像相信和概要和一头的描述信息。

进一步的可以对每个特征数据赋予权重

$$a_{i,t} = w_a^T \tanh(W_{va} v_i + W_{ha} h_t^1)$$

并通过一个 softmax 函数，得到： $a_t = \text{softmax}(a_t)$

最终获得图像加权特征：

$$\hat{v} = \sum_{i=1}^K a_{i,t} v_i$$

另一方面，通过自底向上的 LSTM 网络来计算概率分布，具体的，将图像加权特征和 h_{t-1}^2 组成 language（语言）LSTM 模块的输入：

$$x_t^2 = [\hat{v}_t, h_t^1]$$

这时， t 时刻的预测词（即参考答案）的概率分布为：

$$p(y_t | y_{1:t-1}) = \text{softmax}(W_p h_t^2 + b_p)$$

其中上述第一初始模型和第二初始模型的损失函数为：

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)]$$

其中， y_i 表示样本 i 的标签，正类为 1，负类为 0。 p_i 表示样本 i 预测为正类的概率。

需要说明的是，上述关于第一初始模型和第二初始模型的具体框架的说明是示例性的，在本公开中不对第一初始模型和第二初始模型的具体架构做详细限定。

在步骤 S230 中，基于第一参考答案和真值答案对初始训练数据进行筛选得到第一目标训练数据。

在本示例实施方式中，上述方法可以包括步骤 S310 和步骤 S330。

在步骤 S310 中，计算第一参考答案与真值答案的损失值。

在本示例实施方式中，可以采用上述确定的损失函数开确定上述第一参考答案与真值答案的损失值，也可以首先确定上述第一参考答案与真值答案之间的相似度，利用 1 减去上述相似度得到上述损失值，损失值的计算还可以根据用户需求进行自定义，在本示例实施方式中不做具体限定。

在步骤 S320 中，将损失值小于预设阈值的第一参考答案对应的初始训练数据作为第一目标训练数据。

在得到各第一参考对应损失值之后，可以确定预设阈值，预设阈值可以是 0.2、0.3 等，还可以根据用户需求进行自定义，然后将上述损失值分别与预设阈值进行比较，将损失值小于预设阈值的第一参考答案对应的初始训练数据作为上述第一目标训练数据。

利用损失值对初始训练数据及进行筛选得到第一目标数据，能够降低训练时的噪声，提升了第一目标训练数据的精度，能够使得训练得到的视觉问答模型的处理精度更高。

在本公开的一种示例实施方式中，获取第一目标训练数据时，还可以包括步骤 S410

至步骤 S430。

在步骤 S410 中，将初始训练数据划分为多组初始子数据；

在本示例实施方式中，参照图 4 所示，可以首先将上述初始训练数据划分为多组初始子数据，其中，每组初始子数据中的图像特征数据与文本特征数据的数量可以一个，也可以是多个，如、3 个、5 个等，各初始子数据中的初始训练数据的数量可以相同也可以不同，在本示例实施方式中不做具体限定。

在步骤 S420 中，基于第一参考答案和真值答案对各初始子数据进行筛选得到目标子数据；

在完成分组之后，利用上述第一参考答案和上述真值答案分别对各初始子数据进行筛选得到多个目标子数据，具体的筛选过程可以参考步骤 S310 至步骤 S320，在本示例实施方式中不做具体限定。

在步骤 S430 中，将多组目标子数据作为第一目标训练数据。

在得到多个目标子数据之后，可以将多个目标子数据作为上述第一目标训练数据，将初始训练数据划分为多组进行筛选，可以多组同时进行筛选，能够增减筛选速率，同时，可以在处理器繁忙时，依次对各初始子数据进行筛选能够降低处理器的处理压力，可以在窄带宽的情况下完成对上述初始训练数据进行筛选。

在步骤 S240 中，利用第一目标训练数据的第一参考答案以及对应的真值答案更新第二初始模型得到视觉问答模型。

在本示例实施方式中，在得到上述第一目标训练数据之后，可以利用目标初始训练数据中第一参考答案和第一参考答案对应的真值答案更新第二初始模型得到上述视觉问答模型。

具体的，可以基于上述第一目标训练数据中的第一参考答案和第一参考答案对应的真值答案确定反向传播的第一梯度值，然后利用计算得到的第一梯度值更新第二初始模型中的参数得到视觉问答模型。

在本公开的一种示例实施方式中，参照图 5 所示，上述视觉问答模型的训练方法还可以包括步骤 S250 至步骤 S270。

在步骤 S250 中，将图文输入数据输入至第二初始模型得到第二参考答案。

在本公开的一种示例实施方式中，在得到上述图文输入数据之后，可以将上述图文输入数据输入至上述第二初始模型中得到上述第二参考答案。

第一初始模型和第二初始模型的框架结构相同，第一初始模型和第二初始模型中的模型参数可以相同也可以不同，优选的，上述第一初始模型与第二初始模型中的参数不相同，在训练时能够使得第二初始模型吸收第一初始模型的抗噪声能力，使得训练得到的视觉问答模型的精度更高。

在步骤 S260 中，基于第二参考答案和真值答案对初始训练数据进行筛选得到第二目标训练数据。

在本示例实施方式中，上述方法可以包括步骤 S510 和步骤 S530。

在步骤 S510 中，计算第二参考答案与真值答案的损失值。

在本示例实施方式中，可以采用上述确定的损失函数开确定上述第二参考答案与真值答案的损失值，也可以首先确定上述第二参考答案与真值答案之间的相似度，利用 1 减去上述相似度得到上述损失值，损失值的计算还可以根据用户需求进行自定义，在本示例实施方式中不做具体限定。

在步骤 S520 中，将损失值小于预设阈值的第二参考答案对应的初始训练数据作为第二目标训练数据。

在得到各第二参考对应损失值之后，可以确定预设阈值，预设阈值可以是 0.2、0.3 等，还可以根据用户需求进行自定义，然后将上述损失值分别与预设阈值进行比较，将损失值

小于预设阈值的第二参考答案对应的初始训练数据作为上述第二目标训练数据。

利用损失值对初始训练数据及进行筛选得到第二目标数据，能够降低训练时的噪声，提升了第二目标训练数据的精度，能够使得训练得到的视觉问答模型的处理精度更高。

在本公开的一种示例实施方式中，获取第二目标训练数据时，还可以包括步骤 S610 至步骤 S630。

在步骤 S610 中，将初始训练数据划分为多组初始子数据；

在本示例实施方式中，参照图 5 所示，可以首先将上述初始训练数据划分为多组初始子数据，其中，每组初始子数据中的图像特征数据与文本特征数据的数量可以一个，也可以是多个，如、3 个、5 个等，各初始子数据中的初始训练数据的数量可以相同也可以不同，在本示例实施方式中不做具体限定。

在步骤 S620 中，基于第二参考答案和真值答案对各初始子数据进行筛选得到目标子数据；

在完成分组之后，利用上述第二参考答案和上述真值答案分别对各初始子数据进行筛选得到多个目标子数据，具体的筛选过程可以参考步骤 S510 至步骤 S520，在本示例实施方式中不做具体限定。

在步骤 S630 中，将多组目标子数据作为第二目标训练数据。

在得到多个目标子数据之后，可以将多个目标子数据作为上述第二目标训练数据，将初始训练数据划分为多组进行筛选，可以多组同时进行筛选，能够增减筛选速率，同时，可以在处理器繁忙时，依次对各初始子数据进行筛选能够降低处理器的处理压力，可以在窄带宽的情况下完成对上述初始训练数据进行筛选。

在步骤 S270 中，利用第二目标训练数据的第二参考答案以及对应的真值答案更新第一初始模型得到视觉问答模型。

在本示例实施方式中，在得到上述第二目标训练数据之后，可以利用第二目标训练数据中第二参考答案和第二参考答案对应的真值答案更新第二初始模型得到上述视觉问答模型。

具体的，可以基于上述第二目标训练数据中的第二参考答案和第二参考答案对应的真值答案确定反向传播的第二梯度值，然后利用计算得到的第二梯度值更新第一初始模型中的参数得到视觉问答模型。

在本示例实施方式中，参照图 6 所示，可以利用第一初始模型的输出和真值答案来对第二初始模型进行更新，利用第二初始模型的输出和真值答案对第一初始模型进行更新得到两个视觉问答模型，任意一视觉问答模型都可以完成视觉问答任务。

举例而言，初始子数据包括三个初始训练数据，可以将初始子数据送入第一初始模型。分别计算三个初始训练数据的损失值选出两个第一目标子数据，计算第一目标子数据对应的反向传播的梯度值，并利用该梯度值训练第二初始模型，同理，可以将初始子数据送入第二初始模型。分别计算三个初始训练数据的损失值选出两个第二目标子数据，计算第二目标子数据对应的反向传播的梯度值，并利用该梯度值训练第一初始模型。在将所有初始子数据完成训练之后得到两个视觉问答模型，两个视觉问答模型均可以较为精确地完成视觉问答任务。

综上，本示例性实施方式中，一方面，通过第一初始模型对初始训练数据进行筛选得到第一目标训练数据，提升了初始训练数据的精度，使得得到的视觉问答模型的精度更高；另一方面，通过第一初始模型的输出的第一参考答案来更新第二初始模型的参数得到视觉问答模型，进一步提升了得到的视觉问答模型的抗噪声能力，提升了视觉问答模型的精度。再一方面，利用损失值对初始训练数据及进行筛选得到第一目标数据，能够降低训练时的噪声，提升了第一目标训练数据的精度，能够使得训练得到的视觉问答模型的处理精度更高，又一方面，将初始训练数据划分为多组进行筛选，可以多组同时进行筛选，能够增减

筛选速率，同时，可以在处理器繁忙时，依次对各初始子数据进行筛选能够降低处理器的处理压力，可以在窄带宽的情况下完成对上述初始训练数据进行筛选。

进一步的，参照图 7 所示，本公开还提供一种视觉问答任务处理方法，该方法可以由图 1 中的上述终端执行，也可以由图 1 中的服务器执行，例如，用户使用终端 110 获取待提问图像对应的参考图像特征以及对应的参考文本特征，同时加载视觉问答模型，其中视觉问答模型是由上述视觉模型的训练方法训练得到的。然后将上述参考图像特征和参考文本特征输入至视觉问答模型得到目标答案，以完成对上述视觉问答任务的处理。

再例如，用户使用终端 110 获取待提问图像对应的参考图像特征以及对应的参考文本特征之后，将待提问图像对应的参考图像特征以及对应的参考文本特征上传至服务器 120，由服务器 120 对加载视觉问答模型，其中视觉问答模型是由上述视觉模型的训练方法训练得到的，然后利用视觉问答模型对上述待提问图像对应的参考图像特征以及对应的参考文本特征进行处理得到目标答案。

视觉问答任务处理方法具体可以包括步骤 S710 至步骤 S720。

在步骤 S710 中，获取待提问图像对应的参考图像特征以及对应的参考文本特征。

在本示例实施方式中，参照图 8 所示，在得到待提问图像和对应的问题之后，可以将待提问图像输入至图像特征提取模块得到参考图像特征，将问题输入至文本特征提取模块得到参考文本特征。

图像特征提取以及文本特征提取的具体流程可以参照视觉问答模型的训练方法，在此不再赘述。

在步骤 S720 中，将参考图像特征和参考文本特征输入至视觉问答模型得到目标答案。

在得到上述参考图像特征和参考文本特征之后，可以将参考图像特征和参考文本特征输入至上述视觉问答模型中得到目标答案。

视觉问答模型的具体结构可以参考视觉问答模型的训练方法在此不再赘述。

需要注意的是，上述附图仅是根据本公开示例性实施例的方法所包括的处理的示意性说明，而不是限制目的。易于理解，上述附图所示的处理并不表明或限制这些处理的时间顺序。另外，也易于理解，这些处理可以是例如在多个模块中同步或异步执行的。

进一步的，参考图 9 所示，本示例的实施方式中还提供一种视觉问答模型的训练装置 900，包括数据获取模块 910、数据处理模块 920、数据筛选模块 930 和模型更新模块 940。其中：

数据获取模块 910 可以用于获取初始训练数据，初始训练数据包括图文输入数据以及图文输入数据对应的真值答案，其中，图文输入数据包括图像特征数据和文本特征数据；数据处理模块 920 可以用于将图文输入数据输入至第一初始模型得到第一参考答案；数据筛选模块 930 可以用于基于第一参考答案和真值答案对初始训练数据进行筛选得到第一目标训练数据；模型更新模块 940 可以用于第一目标训练数据的第一参考答案以及对应的真值答案更新第二初始模型得到视觉问答模型。其中，第一初始模型和第二初始模型的模型架构一致。

在一种示例实施方式中，数据获取模块 910 可以被配置为获取初始图像、初始问题文本以及初始问题文本对应的真值答案；多初始图像和初始问题文本进行特征提取得到图像特征数据和文本特征数据。

在一种示例实施方式中，数据筛选模块 930 可以被配置为计算第一参考答案与真值答案的损失值；将损失值小于预设阈值的第一参考答案对应的初始训练数据作为第一目标训练数据。

在另一种示例实施方式中，数据筛选模块 930 可以被配置为将初始训练数据划分为多组初始子数据；基于第一参考答案和真值答案对各初始子数据进行筛选得到目标子数据；将多组目标子数据作为第一目标训练数据。

在一种示例实施方式中，模型更新模块 940 可以被配置为分别利用各目标子数据中的第二参考答案以及对应的真值答案更新第二初始模型得到视觉问答模型。

进一步的，参考图 10 所示，本示例的实施方式中还提供一种视觉问答任务处理装置 1000，包括特征获取模块 1010 和任务处理模块 1020。其中：

特征获取模块 1010 可以用于获取待提问图像对应的参考图像特征以及对应的参考文本特征；任务处理模块 1020 可以用于将参考图像特征和参考文本特征输入至视觉问答模型得到目标答案；其中，视觉问答模型是根据权利要求 1-6 中任一项的视觉问答模型的训练方法得到。

上述装置中各模块的具体细节在方法部分实施方式中已经详细说明，未披露的细节内容可以参见方法部分的实施方式内容，因而不赘述。

本公开的示例性实施方式还提供一种用于执行上述视觉问答模型的训练方法的电子设备，该电子设备可以是上述终端 110 或服务器 120。一般的，该电子设备可以包括处理器与存储器，存储器用于存储处理器的可执行指令，处理器配置为经由执行可执行指令来执行上述视觉问答模型的训练方法。

下面以图 11 中的移动终端 1100 为例，对该电子设备的构造进行示例性说明。本领域技术人员应当理解，除了特别用于移动目的的部件之外，图 11 中的构造也能够应用于固定类型的设备。

如图 11 所示，移动终端 1100 具体可以包括：处理器 1101、存储器 1102、总线 1103、移动通信模块 1104、天线 1、无线通信模块 1105、天线 2、显示屏 1106、摄像模块 1107、音频模块 1108、电源模块 1109 与传感器模块 1110。

处理器 1101 可以包括一个或多个处理单元，例如：处理器 1101 可以包括 AP（Application Processor，应用处理器）、调制解调处理器、GPU（Graphics Processing Unit，图形处理器）、ISP（Image Signal Processor，图像信号处理器）、控制器、编码器、解码器、DSP（Digital Signal Processor，数字信号处理器）、基带处理器和/或 NPU（Neural-Network Processing Unit，神经网络处理器）等。本示例性实施方式中的视觉问答模型的训练方法可以由 AP、GPU 或 DSP 来执行，当方法涉及到神经网络相关的处理时，可以由 NPU 来执行。

处理器 1101 可以通过总线 1103 与存储器 1102 或其他部件形成连接。

存储器 1102 可以用于存储计算机可执行程序代码，可执行程序代码包括指令。处理器 1101 通过运行存储在存储器 1102 的指令，执行移动终端 1100 的各种功能应用以及数据处理。存储器 1102 还可以存储应用数据，例如存储图像，视频等文件。

移动终端 1100 的通信功能可以通过移动通信模块 1104、天线 1、无线通信模块 1105、天线 2、调制解调处理器以及基带处理器等实现。天线 1 和天线 2 用于发射和接收电磁波信号。移动通信模块 1104 可以提供应用在移动终端 1100 上 2G、3G、4G、5G 等移动通信解决方案。无线通信模块 1105 可以提供应用在移动终端 1100 上的无线局域网、蓝牙、近场通信等无线通信解决方案。

显示屏 1106 用于实现显示功能，如显示用户界面、图像、视频等。摄像模块 1107 用于实现拍摄功能，如拍摄图像、视频等。音频模块 208 用于实现音频功能，如播放音频，采集语音等。电源模块 209 用于实现电源管理功能，如为电池充电、为设备供电、监测电池状态等。传感器模块 1110 可以包括深度传感器 11101、压力传感器 11102、陀螺仪传感器 11103、气压传感器 11104 等，以实现相应的感应检测功能。

所属技术领域的技术人员能够理解，本公开的各个方面可以实现为系统、方法或程序产品。因此，本公开的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式（包括固件、微代码等），或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。

本公开的示例性实施方式还提供了一种计算机可读存储介质，其上存储有能够实现本说明书上述方法的程序产品。在一些可能的实施方式中，本公开的各个方面还可以实现为一种程序产品的形式，其包括程序代码，当程序产品在终端设备上运行时，程序代码用于使终端设备执行本说明书上述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步骤。

需要说明的是，本公开所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。

在本公开中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：无线、电线、光缆、RF 等等，或者上述的任意合适的组合。

此外，可以以一种或多种程序设计语言的任意组合来编写用于执行本公开操作的程序代码，程序设计语言包括面向对象的程序设计语言—诸如 Java、C++ 等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网（LAN）或广域网（WAN），连接到用户计算设备，或者，可以连接到外部计算设备（例如利用因特网服务提供商来通过因特网连接）。

本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本公开的其他实施例。本申请旨在涵盖本公开的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本公开的真正范围和精神由权利要求指出。

应当理解的是，本公开并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限定。

权利要求

1. 一种视觉问答模型的训练方法，其特征在于，包括：

获取初始训练数据，所述初始训练数据包括图文输入数据以及所述图文输入数据对应的真值答案，其中，所述图文输入数据包括图像特征数据和文本特征数据；

5 将所述图文输入数据输入至第一初始模型得到第一参考答案；

基于第一参考答案和所述真值答案对所述初始训练数据进行筛选得到第一目标训练数据；

利用所述第一目标训练数据的第一参考答案以及对应的所述真值答案更新第二初始模型得到所述视觉问答模型；

10 其中，所述第一初始模型和所述第二初始模型的模型架构一致。

2. 根据权利要求1所述的方法，其特征在于，所述方法还包括：

将所述图文输入数据输入至所述第二初始模型得到第二参考答案；

基于第二参考答案和所述真值答案对所述初始训练数据进行筛选得到第二目标训练数据；

15 利用所述第二目标训练数据的第二参考答案以及对应的所述真值答案更新所述第一初始模型得到所述视觉问答模型。

3. 根据权利要求1所述的方法，其特征在于，所述基于第一参考答案和所述真值答案对所述初始训练数据进行筛选得到第一目标训练数据包括：

计算所述第一参考答案与所述真值答案的损失值；

20 将损失值小于预设阈值的第一参考答案对应的初始训练数据作为所述第一目标训练数据。

4. 根据权利要求1所述的方法，其特征在于，所述基于第一参考答案和所述真值答案对所述初始训练数据进行筛选得到第一目标训练数据包括：

将所述初始训练数据划分为多组初始子数据；

25 基于所述第一参考答案和所述真值答案对各所述初始子数据进行筛选得到目标子数据；

将多组所述目标子数据作为所述第一目标训练数据。

5. 根据权利要求4所述的方法，其特征在于，所述利用所述第一目标训练数据的第一参考答案以及对应的所述真值答案更新第二初始模型得到所述视觉问答模型包括：

30 分别利用各所述目标子数据中的第二参考答案以及对应的所述真值答案更新所述第二初始模型得到所述视觉问答模型。

6. 根据权利要求1所述的方法，其特征在于，所述获取初始训练数据包括：

获取初始图像、初始问题文本以及所述初始问题文本对应的真值答案；

35 对所述初始图像和所述初始问题文本进行特征提取得到所述图像特征数据和文本特征数据。

7. 一种视觉问答任务处理方法，其特征在于，包括：

获取待提问图像对应的参考图像特征以及对应的参考文本特征；

将所述参考图像特征和所述参考文本特征输入至视觉问答模型得到目标答案；

40 其中，所述视觉问答模型可以根据权利要求1至6中任一项所述的视觉问答模型的训练方法得到。

8. 一种视觉问答模型的训练装置，其特征在于，包括：

数据获取模块，用于获取初始训练数据，所述初始训练数据包括图文输入数据以及所述图文输入数据对应的真值答案，其中，所述图文输入数据包括图像特征数据和文本特征数据；

45 数据处理模块，用于将所述图文输入数据输入至第一初始模型得到第一参考答案；

数据筛选模块，用于基于第一参考答案和所述真值答案对所述初始训练数据进行筛选得到第一目标训练数据；

模型更新模块，利用所述第一目标训练数据的第一参考答案以及对应的所述真值答案更新第二初始模型得到所述视觉问答模型；

5 其中，所述第一初始模型和第二初始模型的模型架构一致。

9. 一种视觉问答任务处理装置，其特征在于，包括：

特征获取模块，用于获取待提问图像对应的参考图像特征以及对应的参考文本特征；

任务处理模块，用于将所述参考图像特征和所述参考文本特征输入至视觉问答模型得到目标答案；

10 其中，所述视觉问答模型是根据权利要求 1 至 6 中任一项所述的视觉问答模型的训练方法得到。

10. 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，所述程序被处理器执行时实现如权利要求 1 至 7 中任一项所述的方法。

11. 一种电子设备，其特征在于，包括：

15 一个或多个处理器；以及

存储器，用于存储一个或多个程序，当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或多个处理器实现如权利要求 1 至 7 中任一项所述的方法。

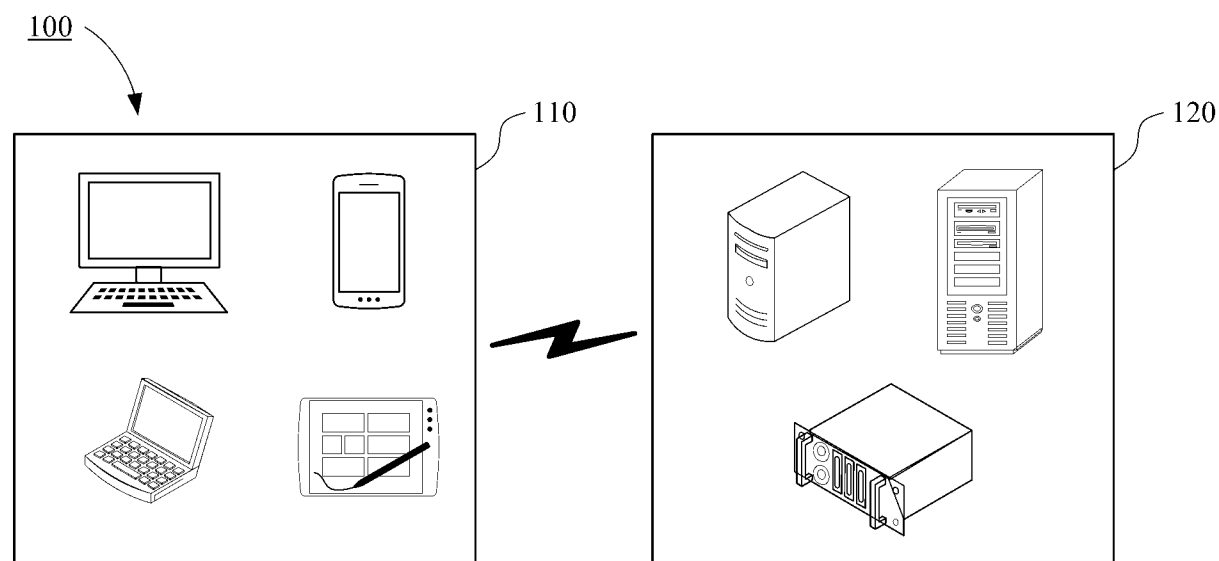


图 1

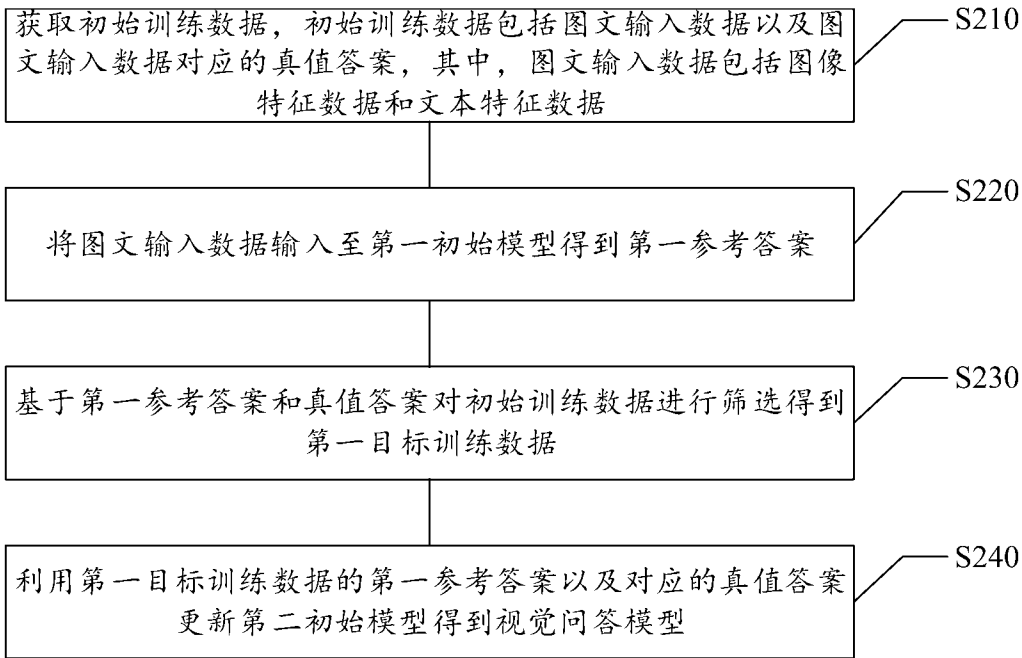


图 2

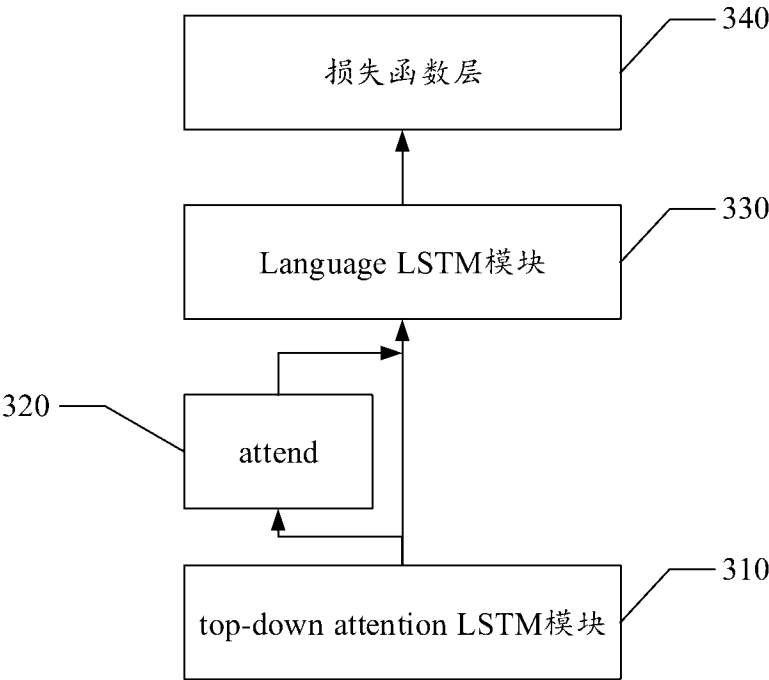


图 3

初始训练数据

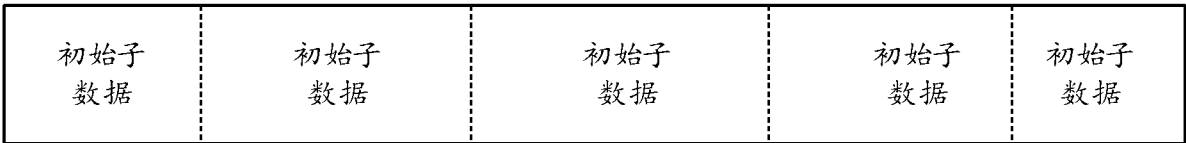


图 4

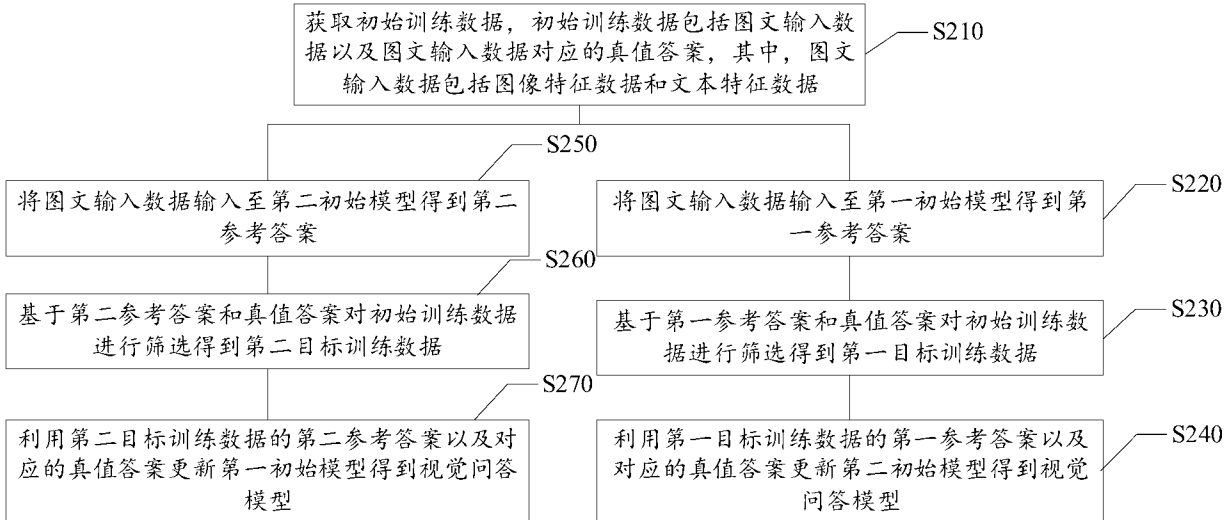


图 5

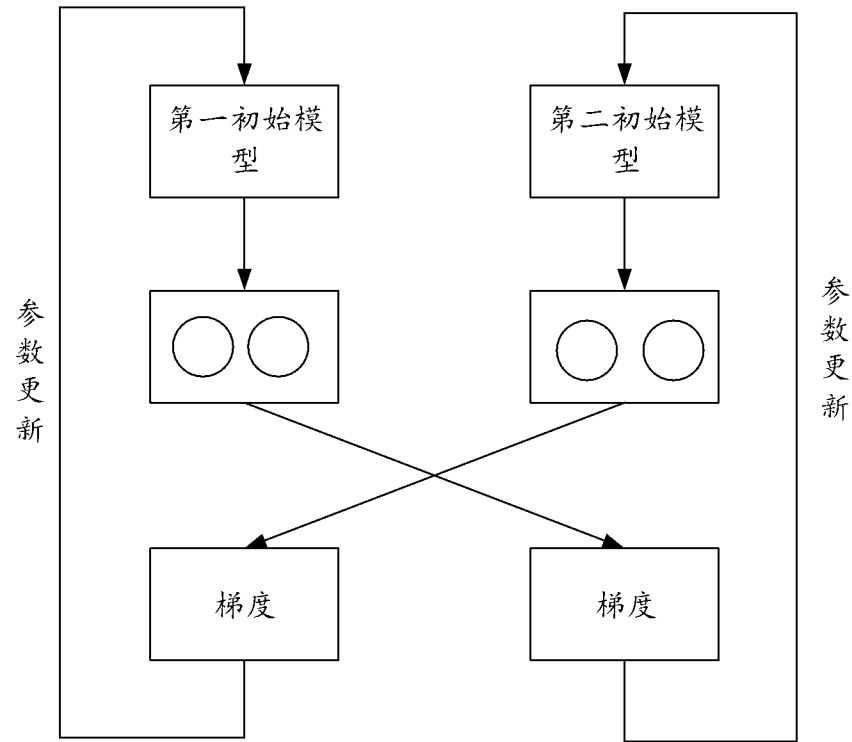


图 6

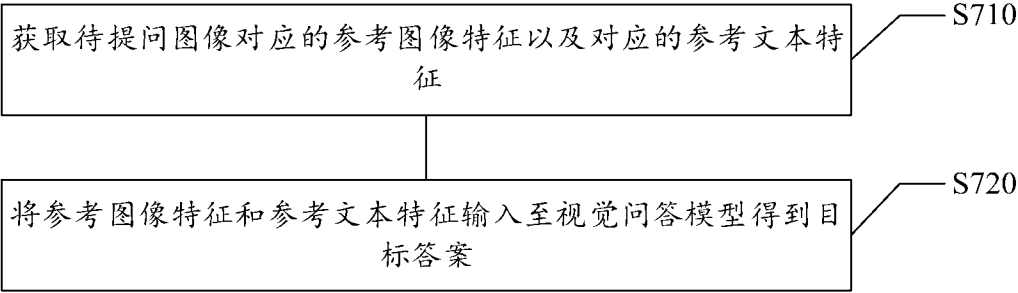


图 7

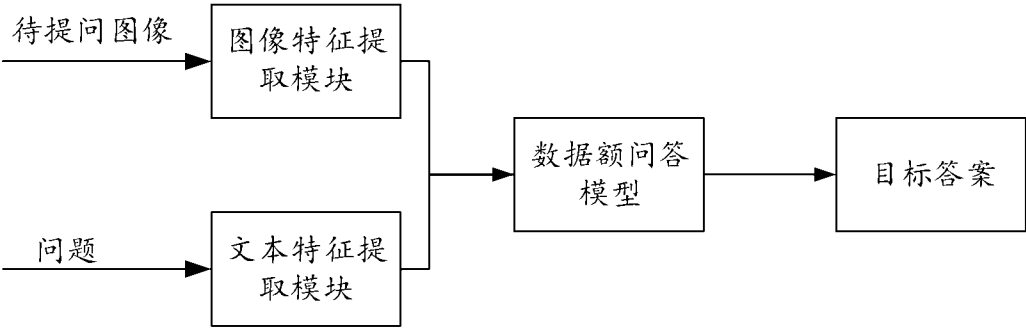


图 8

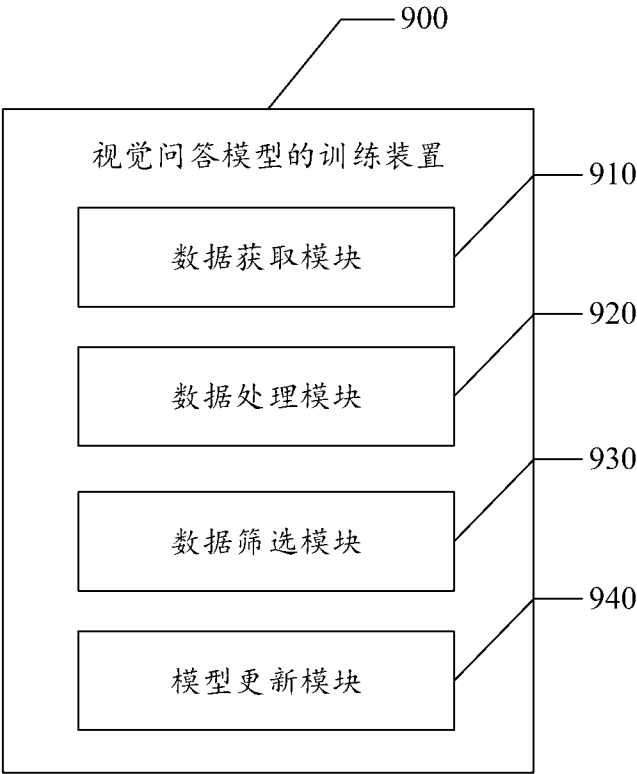


图 9

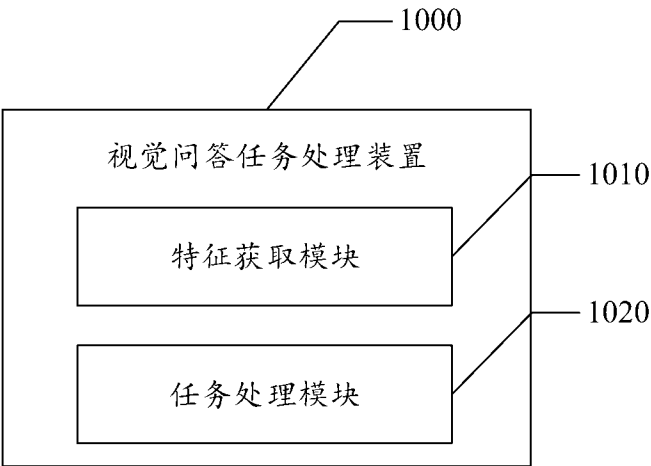


图 10

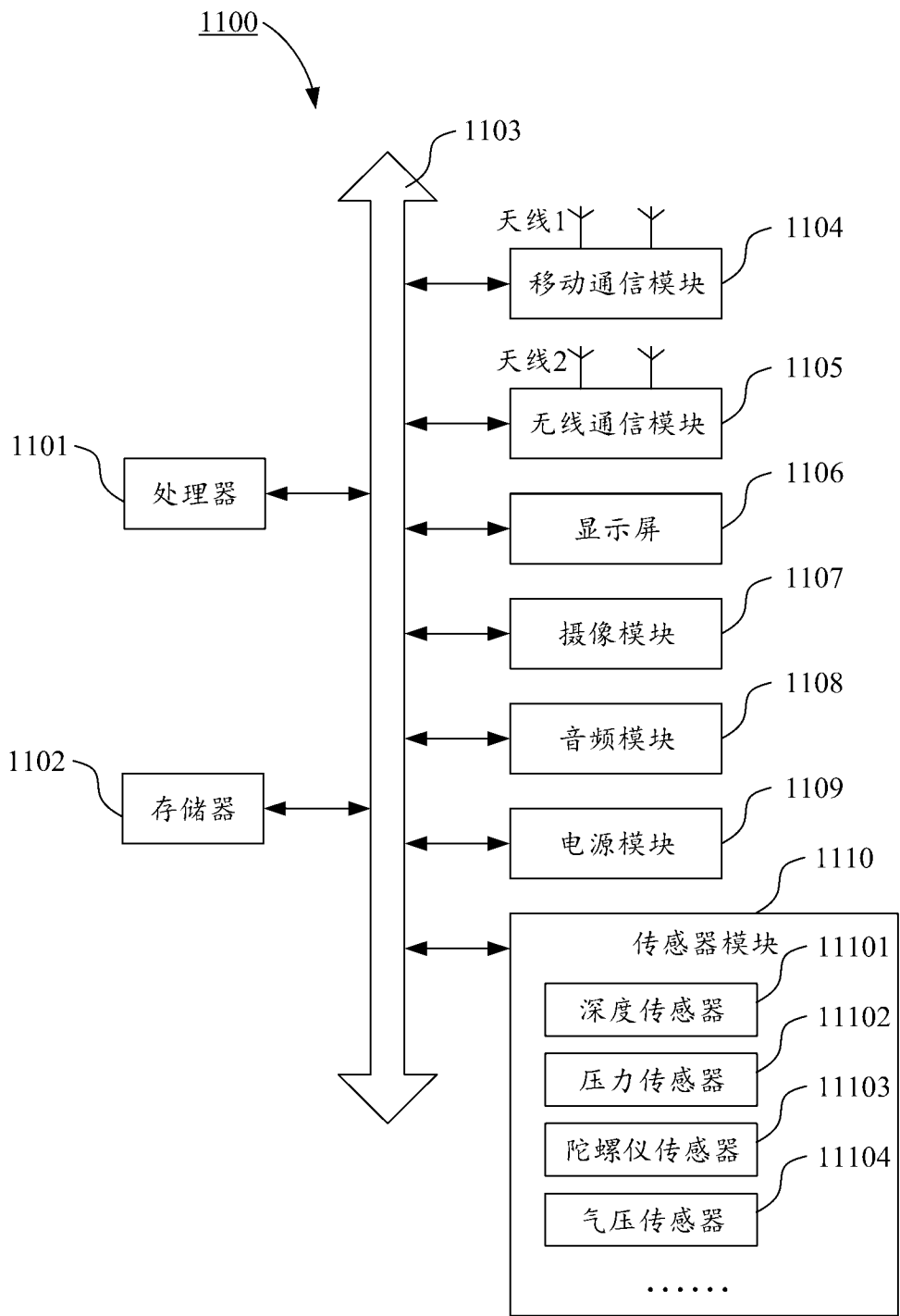


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/138801

A. CLASSIFICATION OF SUBJECT MATTER

G06F16/31(2019.01)i; G06V30/413(2022.01)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)

IPC: G06F, G06V

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)

CNTXT, ENTXTC, ENTXT, DWPI, IEEE, CNKI: 视觉问答, 模型, 训练, 图像, 文本, 特征, 答案, 收敛, 约束, 损失, 分组, 反馈, 更新, Visual Question Answering, Model, Training, Image, Text, Features, Answers, Convergence, Constraints, Loss, Grouping, Feedback, Update

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116894996 A (TECHNOLOGY INNOVATION CENTER OF CHINA TELECOM CORPORATION LIMITED et al.) 17 October 2023 (2023-10-17) claims 1-11, and description, paragraphs 2-136	1-11
Y	CN 110837548 A (TAIKANG INSURANCE GROUP CO., LTD. et al.) 25 February 2020 (2020-02-25) description, paragraphs 2-129, and figures 1-7	1, 3-11
Y	WO 2023024412 A1 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 02 March 2023 (2023-03-02) description, paragraphs 39-111, and figures 1-5	1, 3-11
Y	WO 2021217935 A1 (SHENZHEN ONECONNECT SMART TECHNOLOGY CO., LTD.) 04 November 2021 (2021-11-04) description, paragraphs 120-144, and figures 1-9	4-5



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

“D” document cited by the applicant in the international application

“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

01 March 2024

Date of mailing of the international search report

06 March 2024

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	王峰等 (WANG, Feng et al.). "融合答案掩码的视觉问答模型 (Answer Mask-fused Visual Question Answering Model)" <i>中国图象图形学报 (Journal of Image and Graphics)</i> , 16 November 2023 (2023-11-16), entire document	1-11
A	Ma, Jie et al. "Multitask Learning for Visual Question Answering" <i>IEEE Transactions on Neural Networks and Learning Systems</i> , Vol. 34, No. 3, 31 March 2023 (2023-03-31), entire document	1-11

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/138801

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116894996	A	17 October 2023	None			
CN	110837548	A	25 February 2020	None			
WO	2023024412	A1	02 March 2023	CN	113656570	A	16 November 2021
WO	2021217935	A1	04 November 2021	CN	111639163	A	08 September 2020

A. 主题的分类

G06F16/31(2019.01)i; G06V30/413(2022.01)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)

IPC: G06F, G06V

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))

CNTEXT,ENTXTC,ENTXT,DWPI,IEEE,CNKI: 视觉问答, 模型, 训练, 图像, 文本, 特征, 答案, 收敛, 约束, 损失, 分组, 反馈, 更新, Visual Question Answering, Model, Training, Image, Text, Features, Answers, Convergence, Constraints, Loss, Grouping, Feedback, Update

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
PX	CN 116894996 A (中国电信股份有限公司技术创新中心等) 2023年10月17日 (2023 - 10 - 17) 权利要求1-11, 说明书第2-136段	1-11
Y	CN 110837548 A (泰康保险集团股份有限公司等) 2020年2月25日 (2020 - 02 - 25) 说明书第2-129段, 附图1-7	1,3-11
Y	WO 2023024412 A1 (平安科技(深圳)有限公司) 2023年3月2日 (2023 - 03 - 02) 说明书第39-111段, 附图1-5	1,3-11
Y	WO 2021217935 A1 (深圳壹账通智能科技有限公司) 2021年11月4日 (2021 - 11 - 04) 说明书第120-144段, 附图1-9	4-5
A	王峰等. "融合答案掩码的视觉问答模型" 《中国图象图形学报》, 2023年11月16日 (2023 - 11 - 16), 全文	1-11

☒ 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

2024年3月1日

国际检索报告邮寄日期

2024年3月6日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

于雷

电话号码 (+86) 010-53961719

PCT/ISA/210 表(第2页) (2022年7月)

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
A	Ma, Jie 等. "Multitask Learning for Visual Question Answering" IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, 第34卷，第3期，2023年3月31日 (2023 - 03 - 31)， 全文	1-11

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/138801

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	116894996	A	2023年10月17日	无	
CN	110837548	A	2020年2月25日	无	
WO	2023024412	A1	2023年3月2日	CN	113656570 A 2021年11月16日
WO	2021217935	A1	2021年11月4日	CN	111639163 A 2020年9月8日