

(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
2023 年 2 月 2 日 (02.02.2023)



(10) 国际公布号
WO 2023/005287 A1

- (51) 国际专利分类号:
G06K 9/62 (2022.01)
- (21) 国际申请号: PCT/CN2022/088612
- (22) 国际申请日: 2022 年 4 月 22 日 (22.04.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110866809.2 2021 年 7 月 29 日 (29.07.2021) CN
- (71) 申请人: 北京百度网讯科技有限公司 (BEIJING BAIDU NETCOM SCIENCE TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦二层, Beijing 100085 (CN)。
- (72) 发明人: 希滕(XI, Teng); 中国北京市海淀区上地十街 10 号百度大厦二层, Beijing 100085 (CN)。张刚(ZHANG, Gang); 中国北京市海淀区上地十街 10 号百度大厦二层, Beijing 100085 (CN)。丁二锐(DING, Errui); 中国北京市海淀区上地十

街 10 号百度大厦二层, Beijing 100085 (CN)。吴甜(WU, Tian); 中国北京市海淀区上地十街 10 号百度大厦二层, Beijing 100085 (CN)。王海峰(WANG, Haifeng); 中国北京市海淀区上地十街 10 号百度大厦二层, Beijing 100085 (CN)。

(74) 代理人: 北京清亦华知识产权代理事务所(普通合伙)(TSINGYIHUA INTELLECTUAL PROPERTY LLC); 中国北京市海淀区北洼路 45 号 1 号楼 2 层 201, Beijing 100142 (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, IT, JM, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(54) Title: MODEL PRE-TRAINING METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 模型预训练方法、装置、电子设备和存储介质

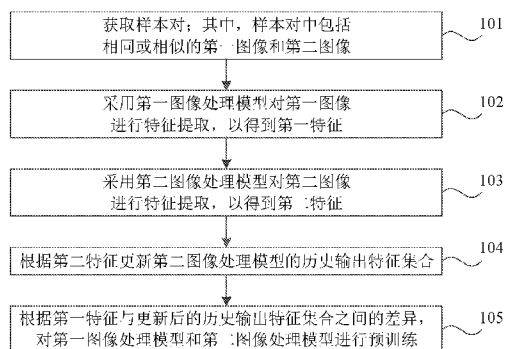


图 1

- 101 Obtain a sample pair, the sample pair comprising the same or similar first and second images
- 102 Perform feature extraction on the first image by using a first image processing model to obtain a first feature
- 103 Perform feature extraction on the second image in the sample pair by using a second image processing model to obtain a second feature
- 104 Update a historical output feature set of the second image processing model according to the second feature
- 105 Pre-train the first image processing model and the second image processing model according to a difference between the first feature and the updated historical output feature set

(57) Abstract: Provided are a model pre-training method and apparatus, an electronic device, a non-transitory computer-readable storage medium, a computer program product, and a computer program. The model pre-training method comprises: obtaining a sample pair; performing feature extraction on a first image in the sample pair by using a first image processing model to obtain a first feature; performing feature extraction on a second image in the sample pair by using a second image processing model to obtain a second feature; updating a historical output feature set of the second image processing model according to the second feature; and pre-training the first image processing model and the second image processing model according to a difference between the first feature and the updated historical output feature set.

(57) 摘要: 提供了模型预训练方法和装置、电子设备、非瞬时计算机可读存储介质、计算机程序产品和计算机程序。该模型预训练方法包括: 获取样本对; 采用第一图像处理模型对样本对中的第一图像进行特征提取, 以得到第一特征; 采用第二图像处理模型对样本对中的第二图像进行特征提取, 以得到第二特征; 根据第二特征更新第二图像处理模型的历史输出特征集合; 和根据第一特征与更新后的历史输出特征集合之间的差异, 对第一图像处理模型和第二图像处理模型进行预训练。

(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))。

模型预训练方法、装置、电子设备和存储介质

相关申请的交叉引用

本申请要求在 2021 年 07 月 29 日在中国提交的中国专利申请号 202110866809.2 的优先权，其全部内容通过引用并入本文。

技术领域

本公开涉及人工智能领域，具体涉及深度学习和计算机视觉等领域，可应用于图像处理、图像识别等场景，更具体地涉及模型预训练方法和装置、电子设备、非瞬时计算机可读存储介质、计算机程序产品和计算机程序。

背景技术

预训练模型被广泛应用于改进上层人工智能任务的效果，在上游任务中，通过大量的训练数据对预训练模型进行预训练，即可实现在下游任务中，利用少量训练数据对模型进行训练的情况下，就可以取得较好的预测结果。因此，如何对预训练模型进行预训练是非常重要的。

发明内容

本公开提供了一种用于模型预训练方法、装置、电子设备、非瞬时计算机可读存储介质、计算机程序产品和计算机程序。

根据本公开一方面的实施例，提供了一种模型预训练方法，包括：

获取样本对；其中，所述样本对中包括相同或相似的第一图像和第二图像；

采用第一图像处理模型对所述第一图像进行特征提取，以得到第一特征；

采用第二图像处理模型对所述第二图像进行特征提取，以得到第二特征；

根据所述第二特征更新所述第二图像处理模型的历史输出特征集合；

根据所述第一特征与更新后的所述历史输出特征集合之间的差异，对所述第一图像处理模型和所述第二图像处理模型进行预训练。

根据本公开另一方面的实施例，提供了一种模型预训练装置，包括：

获取模块，用于获取样本对；其中，所述样本对中包括相同或相似的第一图像和第二图像；

第一提取模块，用于采用第一图像处理模型对所述第一图像进行特征提取，以得到第一特征；

第二提取模块，用于采用第二图像处理模型对所述第二图像进行特征提取，以得到第二特征；

更新模块，用于根据所述第二特征更新所述第二图像处理模型的历史输出特征集合；

5 预训练模块，用于根据所述第一特征与更新后的所述历史输出特征集合之间的差异，对所述第一图像处理模型和所述第二图像处理模型进行预训练。

根据本公开又一方面的实施例，提供了一种电子设备，包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

10 所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行本公开上述一方面提出的模型预训练方法。

根据本公开再一方面的实施例，提供了一种计算机指令的非瞬时计算机可读存储介质，所述计算机指令用于使所述计算机执行本公开上述一方面提出的模型预训练方法。

根据本公开还一方面的实施例，提供了一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现本公开上述一方面提出的模型预训练方法。

根据本公开还一方面的实施例，提供了一种计算机程序，包括计算机程序代码，当所述计算机程序代码在计算机上运行时，使得计算机执行本公开上述一方面提出的模型预训练方法。

应当理解，本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征，也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

25 附图用于更好地理解本方案，不构成对本公开的限定。其中：

图 1 是本公开实施例一所提供的模型预训练方法的流程示意图；

图 2 为本公开实施例二所提供的模型预训练方法的流程示意图；

图 3 为本公开实施例中基于多队列的大规模预训练架构示意图；

图 4 为本公开实施例三所提供的模型预训练方法的流程示意图；

30 图 5 为本公开实施例四所提供的模型预训练方法的流程示意图；

图 6 为本公开实施例五所提供的模型预训练装置的结构示意图；

图 7 示出了可以用来实施本公开的实施例的示例电子设备的示意性框图。

具体实施方式

以下结合附图对本公开的示范性实施例做出说明，其中包括本公开实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本公开的范围和精神。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

现有的预训练模型主要在单任务上进行预训练，或者在类似的任务上进行迁移，比如 cifar10、cifar100、imagenet（上述三个是为了解决分类识别问题而设计的数据集）等分类任务。然而很少有工作研究对预训练模型进行跨任务的预训练，比如跨粗粒度的分类任务和细粒度的人脸识别任务的预训练，因此上述预训练方式在应用场景中具有很大的局限性。

针对上述存在的问题，本公开提出一种模型预训练方法、装置、电子设备、非瞬时计算机可读存储介质、计算机程序产品和计算机程序。

下面参考附图描述本公开实施例的模型预训练方法、装置、电子设备、非瞬时计算机可读存储介质、计算机程序产品和计算机程序。

图 1 为本公开实施例一所提供的模型预训练方法的流程示意图。

本公开实施例以该模型预训练方法被配置于模型预训练装置中来举例说明，该模型预训练装置可以应用于任一电子设备中，以使该电子设备可以执行模型预训练功能。

其中，电子设备可以为任一具有计算能力的设备，例如可以为个人电脑（Personal Computer，简称 PC）、移动终端、服务器等，移动终端例如可以为手机、平板电脑、个人数字助理、穿戴式设备等具有各种操作系统、触摸屏和/或显示屏的硬件设备。

如图 1 所示，该模型预训练方法可以包括以下步骤：

步骤 101，获取样本对；其中，样本对中包括相同或相似的第一图像和第二图像。

在本公开实施例中，样本对，可以从现有的数据集、测试集或训练集中获取，或者，也可以在线采集，或者，也可以线下采集，比如可以通过用户手持图像采集设备拍摄得到，或者，也可以获取样本对中的其中一个图像，通过图像处理技术，对上述其中一个图像进行图像处理，以得到样本对中的另一个图像，等等，本公开对此并不做限制。

步骤 102，采用第一图像处理模型对第一图像进行特征提取，以得到第一特征。

在本公开实施例中，对第一图像处理模型的结构不作限制，比如，该第一图像处理模型可以为以 Transformer 为基本结构的模型，或者，也可以为其他结构的模型，比如为 Transformer 的变种结构的模型。

在本公开实施例中，可以通过第一图像处理模型对第一图像进行特征提取，将提取到的第一图像的特征，作为第一特征。

步骤 103，采用第二图像处理模型对第二图像进行特征提取，以得到第二特征。

在本公开实施例中，对第二图像处理模型的结构不作限制，比如，该第二图像处理模型可以为以 Transformer 为基本结构的模型，或者，也可以为其他结构的模型，比如为 Transformer 的变种结构的模型。

在本公开实施例中，可以通过第二图像处理模型对第二图像进行特征提取，将提取到的第二图像的特征，作为第二特征。

步骤 104，根据第二特征更新第二图像处理模型的历史输出特征集合。

在本公开实施例中，可以根据第二特征更新第二图像处理模型的历史输出特征集合。

作为一种示例，可以预先设置历史输出特征集合中包括的历史输出特征数量的上限值（比如 10^6 个），在历史输出特征集合中的历史输出特征的个数低于上限值的情况下，可以将第二特征直接添加至历史输出特征集合中，而在历史输出特征集合中的历史输出特征的个数等于上限值的情况下，可以将第二特征添加至历史输出特征集合中，并删除历史输出特征集合中的历史输出特征，比如可以删除历史输出特征集合中最早添加的历史输出特征。

举例而言，以历史输出特征集合包含的历史输出特征数量的上限值为 4 进行示例性说明，假设历史输出特征集合中包括按照时间先后顺序，依次添加的历史输出特征 1、历史输出特征 2、历史输出特征 3 以及历史输出特征 4，则可以将第二特征添加至历史输出特征集合，并删除历史输出特征 1。

步骤 105，根据第一特征与更新后的历史输出特征集合之间的差异，对第一图像处理模型和第二图像处理模型进行预训练。

在本公开实施例中，第一特征与更新后的历史输出特征集合之间的差异，可以根据第一特征与更新后的历史输出特征集合中的每个历史输出特征之间的差异确定。其中，第一特征与更新后的历史输出特征集合之间的差异，与第一特征与各历史输出特征之间的差异成正向关系。

应当理解的是，更新后的历史输出特征集合中各个特征中只有第二特征是与第一特征相似的，而其他特征可能与第一特征并不相似，为了使得模型能够正确对各图像进行识别，在对模型进行预训练时，可以根据第一特征与更新后的历史输出特征集合之间的差异，对第一图像处理模型和第二图像处理模型进行预训练，以使上述差异最大化。

本公开实施例的模型预训练方法，通过获取样本对，并采用第一图像处理模型对样本对中的第一图像进行特征提取，以得到第一特征，以及采用第二图像处理模型对样本对中的第二图像进行特征提取，以得到第二特征，根据第二特征更新第二图像处理模型的历史输出特征集合，以根据第一特征与更新后的历史输出特征集合之间的差异，对第一图像处理模型和第二图像处理模型进行预训练。由此，可以实现对不同的模型进行联合预训练，以提升模型预测效果。

为了清楚说明上述实施例中是如何提取第一特征和第二特征的，本公开提供了另一种模型预训练方法。

图 2 为本公开实施例二所提供的模型预训练方法的流程示意图。

如图 2 所示，该模型预训练方法可以包括以下步骤：

步骤 201，获取样本对；其中，样本对中包括相同或相似的第一图像和第二图像。

步骤 201 的执行过程可以参见上述实施例的执行过程，在此不做赘述。

步骤 202，采用待执行的任务对应的第一转换层将第一图像转化为第一输入序列。

在本公开实施例中，待执行的任务可以包括粗粒度的分类任务，和/或，包括细粒度的人脸识别等任务。

在本公开实施例中，转换层用于对图像进行序列化表示。

在本公开实施例中，不同任务对应的序列长度可以不同，可以通过待执行的任务对应的第一转换层将第一图像进行转化，得到该第一图像对应的第一输入序列。

作为一种示例，如图 3 所示，可以将第一图像进行分块，得到 9 个图块，通过平坦图块的线性投影（Linear Projection of Flattened Patches）将 9 个图块进行转化，得到第一输入序列。

步骤 203，将第一输入序列添加到第一图像处理模型的第一目标队列中。

其中，第一目标队列是第一图像处理模型的多个队列中上述待执行的任务所对应的队列。

需要说明的是，不同任务对应的序列长度可以不同，为了便于对同一任务对应的输入序列进行处理，可以通过不同队列，来暂存不同任务对应的输入序列。

即，本公开中，可以从第一图像处理模型的多个队列中，确定待执行的任务对应的第一目标队列，将第一输入序列添加到第一目标队列中。

步骤 204，在第一目标队列轮询到第一输入序列的情况下，将第一输入序列输入到第一图像处理模型进行特征提取，得到第一特征。

在本公开实施例中，可以利用第一图像处理模型依次对各第一目标队列中的各输

入序列进行特征提取，在第一目标队列轮询到第一输入序列的情况下，可以将第一输入序列输入到第一图像处理模型中进行特征提取，得到该第一图像对应的第一特征。

步骤 205，采用任务对应的第二转化层将第二图像转化为第二输入序列。

同理，由于不同任务对应的序列长度可以不同，可以通过待执行的任务对应的第二转换层将第二图像进行转化，得到该第二图像对应的第二输入序列。

作为一种示例，如图 3 所示，可以将第二图像进行分块，得到 9 个图块，通过平坦图块的线性投影（Linear Projection of Flattened Patches）将 9 个图块进行转化，得到第二输入序列。

步骤 206，将第二输入序列添加到第二图像处理模块对应的第二目标队列中。

同理，由于不同任务对应的序列长度可以不同，为了便于对同一任务对应的输入序列进行处理，可以通过不同队列，来暂存不同任务对应的输入序列。

即，本公开中，可以从第二图像处理模型的多个队列中，确定待执行的任务对应的第二目标队列，将第二输入序列添加到第二目标队列中。

步骤 207，在第一目标队列将第一输入序列输入第一图像处理模型的情况下，同步将属于相同样本对的第二输入序列输入到第二图像处理模型，得到第二特征。

在本公开实施例中，第二输入序列可以与第一输入序列同步进行特征提取，即在第一目标队列将第一输入序列输入到第一图像处理模型进行特征提取的情况下，可以同步将属于相同样本对（比如同一 ID 的样本对）的第二输入序列输入到第二图像处理模型进行特征提取，得到第二图像的第二特征。

步骤 208，根据第二特征更新第二图像处理模型的历史输出特征集合。

步骤 209，根据第一特征与更新后的历史输出特征集合之间的差异，对第一图像处理模型和第二图像处理模型进行预训练。

步骤 208 至 209 的执行过程可以参见上述实施例的执行过程，在此不做赘述。

本公开实施例的模型预训练方法，通过采用待执行的任务对应的第一转换层将第一图像转化为第一输入序列，将第一输入序列添加到第一图像处理模型的第一目标队列中，在第一目标队列轮询到第一输入序列的情况下，将第一输入序列输入到第一图像处理模型进行特征提取，得到第一特征。由此，通过将不同任务对应的输入序列，添加至不同的队列中，每个任务维护单独的队列，从而第一图像处理模型可以对各任务进行区分或针对性处理，提升特征提取的准确性。

同理，通过采用任务对应的第二转化层将第二图像转化为第二输入序列，将第二输入序列添加到第二图像处理模块对应的第二目标队列中，在第一目标队列将第一输入序列输入第一图像处理模型的情况下，同步将属于相同样本对的第二输入序列输入

到第二图像处理模型，得到第二特征。由此，通过将不同任务对应的输入序列，添加至不同的队列中，每个任务维护单独的队列，从而第二图像处理模型可以对各任务进行区分或针对性处理，提升特征提取的准确性。

5 为了清楚说明上述实施例中是如何对第一图像处理模型和第二图像处理模型进行预训练的，本公开还提供一种模型预训练方法。

图 4 为本公开实施例三所提供的模型预训练方法的流程示意图。

如图 4 所示，该模型预训练方法可以包括以下步骤：

步骤 401，获取样本对；其中，样本对中包括相同或相似的第一图像和第二图像。

10 步骤 401 的执行过程可以参见上述实施例的执行过程，在此不做赘述。

步骤 402，针对每个待执行的任务，采用该任务对应的第一转换层将第一图像转化为第一输入序列，其中，任务为多个。

在本公开实施例中，不同任务对应的序列长度可以不同，针对每个任务，可以采用该任务对应的第一转换层将第一图像转化为第一输入序列。

15 步骤 403，将第一输入序列添加到第一图像处理模型的第一目标队列中。

其中，第一目标队列是第一图像处理模型的多个队列中该任务所对应的队列。

需要说明的是，不同任务对应的序列长度可以不同，为了便于对同一任务对应的输入序列进行处理，可以通过不同队列，来暂存不同任务对应的输入序列。

20 即，本公开中，针对每个任务，可以从第一图像处理模型的多个队列中，确定该任务对应的第一目标队列，将第一输入序列添加到该任务对应的第一目标队列中。

步骤 404，在第一目标队列轮询到第一输入序列的情况下，将第一输入序列输入到第一图像处理模型进行特征提取，得到第一特征。

在本公开实施例中，针对每个任务，可以采用第一图像处理模型对该任务对应的第一输入序列进行特征提取，得到该任务对应的第一特征。

25 步骤 405，针对每个待执行的任务，采用该任务对应的第二转化层将第二图像转化为第二输入序列。

在本公开实施例中，不同任务对应的序列长度可以不同，针对每个任务，可以采用该任务对应的第二转化层将第二图像转化为第二输入序列。

步骤 406，将第二输入序列添加到第二图像处理模型对应的第二目标队列中。

30 同理，由于不同任务对应的序列长度可以不同，为了便于对同一任务对应的输入序列进行处理，可以通过不同队列，来暂存不同任务对应的输入序列。

即，本公开中，针对每个任务，可以从第二图像处理模型的多个队列中，确定该

任务对应的第二目标队列，将第二输入序列添加到该任务对应的第二目标队列中。

步骤 407，在第一目标队列将第一输入序列输入第一图像处理模型的情况下，同步将属于相同样本对的第二输入序列输入到第二图像处理模型，以得到第二特征。

在本公开实施例中，每个任务均具有对应的第一输入序列和第二输入序列，第一
5 图像处理模型可以对各任务对应的第一输入序列进行特征提取，得到各任务对应的第一特征，相应的，第二图像处理模型可以同步对各任务对应的第二输入序列进行特征提取，得到各任务对应的第二特征。

步骤 408，根据第二特征更新第二图像处理模型的历史输出特征集合。

在公开实施例中，每个任务均可以具有对应的历史输出特征集合，针对每个任务，
10 可以将该任务对应的第二特征添加到该任务对应的历史输出特征集合中。

作为一种示例，针对每个任务，可以预先设置该任务对应的历史输出特征集合中包括的历史输出特征数量的上限值（比如 10^6 个），在历史输出特征集合中的历史输出特征的个数低于上限值的情况下，可以将第二特征直接添加至历史输出特征集合中，而在历史输出特征集合中的历史输出特征的个数等于上限值的情况下，可以将第二特征添加至历史输出特征集合中，并删除历史输出特征集合中的历史输出特征，比如可以删除历史输出特征集合中最早添加的历史输出特征。由此，可以实现对各历史输出特征集合进行动态维护和更新。
15

步骤 409，针对每个任务，根据第一图像处理模型对各任务对应的第一输入序列进行特征提取所得到的第一特征，与对应任务更新后的历史输出特征集合之间的差异，
20 得到各任务的对比损失函数值。

在本公开实施例中，针对每个任务，可以计算该任务对应的第一特征与该任务对应的更新后的历史输出特征集合之间的差异（比如该差异标记为第一差异），例如，可以计算该任务对应的第一特征与该任务对应的更新后的历史输出特征集合中的每个特征之间的差异（比如该差异标记为第二差异），根据第一特征与每个特征之间的第二差异，确定第一差异。其中，第一差异与第二差异成正向关系，即第二差异越大，第一差异越大，反之，第二差异越小，第一差异越小。
25

从而本公开中，可以根据上述第一差异，确定该任务对应的对比损失函数值，其中，对比损失函数值与第一差异成反向关系，即第一差异越大，对比损失函数值越小，反之，第一差异越小，损失函数值越大。

30 步骤 410，对各任务的对比损失函数值加权求和得到目标对比损失函数值。

在本公开实施例中，可以对各任务对应的对比损失函数值进行加权求和，得到目标对比损失函数值 Loss。例如，可以根据各任务对应的权重，对各任务对应的对比损

失函数值进行加权求和，得到目标对比损失函数值。

其中，不同任务对应的权重可以固定为特定的比例，或者，也可以为端到端学习比例，本公开对此并不做限制。

5 步骤 411，根据目标对比损失函数值，对第一图像处理模型和第二图像处理模型进行模型参数调整。

在本公开实施例中，可以根据目标对比损失函数值，对第一图像处理模型和第二图像处理模型进行模型参数调整，以使目标对比损失函数值的取值最小化。

在本公开的任意一个实施例之中，不同任务可以具有不同的序列长度，即编码方式不同，各个任务对应的注意力权重也可以不同，为了使得模型获知输入序列是用于
10 执行哪种任务的，以提升模型的预测效果，可以在输入序列中添加任务对应的位置编码。

即，本公开中，针对每个任务，还可以将该任务对应的第一位置编码添加到该任务对应的第一输入序列中，其中，第一位置编码用于指示第一图像处理模型采用对应的注意力权重。

15 作为一种示例，以第一图像处理模型为基于 Transformer 结构的模型进行示例，在不同任务对应的序列长度不同的情况下，不同任务具有对应的位置编码（position embedding），该位置编码用于指示 Transformer 模型的自注意力机制（self attention）部分可以加入对应 position embedding 的注意力权重。

同理，针对每个任务，还可以将该任务对应的第二位置编码添加到该任务对应的
20 第二输入序列中，其中，第二位置编码用于指示第二图像处理模型采用对应的注意力权重。

作为一种示例，以第二图像处理模型与第一图像处理模型的结构相同进行示例，在不同任务对应的序列长度不同的情况下，不同任务具有对应的位置编码（position embedding），该位置编码用于指示 Transformer 模型的 self attention 部分可以加入对应
25 position embedding 的注意力权重。

由此，在多任务上对模型进行预训练，可以解决相关技术中在单任务上对模型进行预训练的方式，存在应用场景的局限性的技术问题。并且，对模型进行大规模预训练，可以提升小模型的预测精度和速度，使得模型在同样的速度情况下精度更高，或者，在同样精度的情况下，速度更快，可以提升模型在特定硬件或芯片上处理图像或者
30 音视频的速度。

本公开实施例的模型预训练方法，通过针对每个任务，根据第一图像处理模型对各任务对应的第一输入序列进行特征提取所得到的第一特征，与对应任务更新后的历

史输出特征集合之间的差异，得到各任务的对比损失函数值，对各任务的对比损失函数值加权求和得到目标对比损失函数值，可以提升目标对比损失函数值确定结果的可靠性和合理性，从而根据该目标对比损失函数值，对第一图像处理模型和第二图像处理模型进行模型参数调整，可以提升模型的预训练效果。

5

为了清楚说明上述实施例中是如何对第一图像处理模型和第二图像处理模型进行预训练的，本公开还提供一种模型预训练方法。

图 5 为本公开实施例四所提供的模型预训练方法的流程示意图。

如图 5 所示，该模型预训练方法可以包括以下步骤：

- 10 步骤 501，获取样本对；其中，样本对中包括相同或相似的第一图像和第二图像。
- 步骤 502，采用第一图像处理模型对第一图像进行特征提取，以得到第一特征。
- 步骤 503，采用第二图像处理模型对第二图像进行特征提取，以得到第二特征。
- 步骤 504，根据第二特征更新第二图像处理模型的历史输出特征集合。
- 步骤 501 至 504 的执行过程可以参见上述任一实施例的执行过程，在此不做赘述。
- 15 步骤 505，根据第一特征与更新后的历史输出特征集合之间的差异，确定目标对比损失函数值。

在本公开实施例中，可以计算第一特征与更新后的历史输出特征集合之间的差异（比如该差异标记为第一差异），例如，可以计算第一特征与更新后的历史输出特征集合中的每个特征之间的差异（比如该差异标记为第二差异），根据第一特征与每个

20 特征之间的第二差异，确定第一差异，其中，第一差异可以与第二差异成正向关系。从而本公开中，可以根据上述第一差异，确定目标对比损失函数值，其中，目标对比损失函数值与第一差异成反向关系，即第一差异越大，目标对比损失函数值越小，反之，第一差异越小，目标对比损失函数值越大。

步骤 506，根据目标对比损失函数值，对第二图像处理模型进行模型参数调整。

- 25 在本公开实施例中，可以根据目标对比损失函数值，对第二图像处理模型进行模型参数调整，以使目标对比损失函数值的取值最小化。

步骤 507，根据第二图像处理模型参数调整前和调整后的模型参数变化量，对第一图像处理模型进行模型参数调整。

在本公开实施例中，可以采用动量更新方式，对两个模型中的模型参数进行更新，

30 即可以在线对第二图像处理模型中的模型参数进行更新，在第二图像处理模型更新后，利用第二图像处理模型参数调整前和调整后的模型参数变化量，离线对第一图像处理模型进行模型参数调整。也就是说，可以根据第二图像处理模型中模型参数的变化趋

势，对第一图像处理模型中的模型参数进行调整。

需要说明的是，第一特征是当前样本对中第一图像对应的特征，而更新后的历史输出特征集合中的各特征不仅包含当前样本对中第二图像对应的特征，还包含历史输入的各样本对中图像对应的特征，将第一图像处理模型当前提取到的第一特征与第二图像处理模型的历史输出特征集合中各特征进行比对，将导致两个模型在时间维度不对等。

因此，在本公开实施例中，为了使得两个模型在特征提取时，在时间维度上尽量对等，从而提升模型预训练效果，可以根据第二图像处理模型中模型参数的变化趋势，对第一图像处理模型中的模型参数进行调整，以使第一图像处理模型中的模型参数在时间维度上滞后些。

作为一种示例，可以将第二图像处理模型参数调整前和调整后的模型参数变化量，与第一图像处理模型的模型参数进行加权融合，以得到第一图像处理模型调整后的模型参数。比如，可以将上述模型参数变化量与第一图像处理模型的模型参数进行叠加，得到第一图像处理模型调整后的模型参数。

举例而言，第一图像处理模型的模型权重为 1，第二图像处理模型对应的模型参数变化量为 0.1，则调整后的模型权重可以为 0.9。

作为一种示例，如图 3 所示，以任务 1 为物体识别任务，任务 2 为人脸识别任务进行示例性说明，其中，第一图像处理模型和第二图像处理模型的结构可以相同。

图 3 中，任务 1 和任务 2 输入的都是正样本对，即该正样本对中可以包括采用同一 ID 的不同图像，或者包括原始图像以及对原始图像进行处理后得到的新的图像。

图 3 中，预训练架构图中可以同时支持多任务，每个任务维护单独的队列，不同任务具有对应的转换层，用于将图像转换成序列。其中，不同任务对应的序列长度可以不同，在不同任务对应的序列长度不同的情况下，不同任务对应的序列中，需要加入对应任务所对应的位置编码，模型的自注意力机制部分可以加入对应位置编码的注意力权重。

图 3 中，第一图像处理模型和第二图像处理模型可以通过动量的方式更新模型参数，即第二图像处理模型可以在线更新，第一图像处理模型可以离线更新。

本公开实施例的模型预训练方法，通过根据第一特征与更新后的历史输出特征集合之间的差异，确定目标对比损失函数值；根据目标对比损失函数值，对第二图像处理模型进行模型参数调整；根据第二图像处理模型参数调整前和调整后的模型参数变化量，对第一图像处理模型进行模型参数调整。由此，可以实现对第一图像处理模型和第二图像处理模型进行联合预训练，可以提升模型的预测效果。

与上述图 1 至图 5 实施例提供的模型预训练方法相对应，本公开还提供一种模型预训练装置，由于本公开实施例提供的模型预训练装置与上述图 1 至图 5 实施例提供的模型预训练方法相对应，因此在模型预训练方法的实施方式也适用于本公开实施例提供的模型预训练装置，在本公开实施例中不再详细描述。

图 6 为本公开实施例五所提供的模型预训练装置的结构示意图。

如图 6 所示，该模型预训练装置 600 可以包括：获取模块 610、第一提取模块 620、第二提取模块 630、更新模块 640 以及预训练模块 650。

其中，获取模块 610，用于获取样本对；其中，样本对中包括相同或相似的第一图像和第二图像。

第一提取模块 620，用于采用第一图像处理模型对第一图像进行特征提取，以得到第一特征。

第二提取模块 630，用于采用第二图像处理模型对第二图像进行特征提取，以得到第二特征。

更新模块 640，用于根据第二特征更新第二图像处理模型的历史输出特征集合。

预训练模块 650，用于根据第一特征与更新后的历史输出特征集合之间的差异，对第一图像处理模型和第二图像处理模型进行预训练。

在本公开实施例的一种可能的实现方式中，第一提取模块 620，具体用于：采用待执行的任务对应的第一转换层将第一图像转化为第一输入序列；将第一输入序列添加到第一图像处理模型的第一目标队列中；在第一目标队列轮询到第一输入序列的情况下，将第一输入序列输入到第一图像处理模型进行特征提取；其中，第一目标队列是第一图像处理模型的多个队列中任务所对应的队列。

在本公开实施例的一种可能的实现方式中，第二提取模块 630，具体用于：采用任务对应的第二转化层将第二图像转化为第二输入序列；将第二输入序列添加到第二图像处理模型对应的第二目标队列中；在第一目标队列将第一输入序列输入第一图像处理模型的情况下，同步将属于相同样本对的第二输入序列输入到第二图像处理模型。

在本公开实施例的一种可能的实现方式中，任务为多个，对应地，各任务具有对应的第二输入序列，且各任务具有对应的历史输出特征集合；更新模块 640，具体用于：将第二图像处理模型对各任务对应的第二输入序列所提取的第二特征，添加到对应任务的历史输出特征集合，并删除对应任务的历史输出特征集合中的历史输出特征。

在本公开实施例的一种可能的实现方式中，预训练模块 650，具体用于：针对每个任务，根据第一图像处理模型对各任务对应的第一输入序列进行特征提取所得到的第

一特征，与对应任务更新后的历史输出特征集合之间的差异，得到各任务的对比损失函数值；对各任务的对比损失函数值加权求和得到目标对比损失函数值；根据目标对比损失函数值，对第一图像处理模型和第二图像处理模型进行模型参数调整。

在本公开实施例的一种可能的实现方式中，该模型预训练装置 600 还可以包括：

5 添加模块，用于将任务对应的第一位置编码添加到第一输入序列中，其中，第一位置编码用于指示第一图像处理模型采用对应的注意力权重；将任务对应的第二位置编码添加到第二输入序列中，其中，第二位置编码用于指示第二图像处理模型采用对应的注意力权重。

在本公开实施例的一种可能的实现方式中，预训练模块 650，具体用于：根据第一
10 特征与更新后的历史输出特征集合之间的差异，确定目标对比损失函数值；根据目标对比损失函数值，对第二图像处理模型进行模型参数调整；根据第二图像处理模型参数调整前和调整后的模型参数变化量，对第一图像处理模型进行模型参数调整。

在本公开实施例的一种可能的实现方式中，预训练模块 650，具体用于：将第二图
15 像处理模型参数调整前和调整后的模型参数变化量，与第一图像处理模型的模型参数进行加权融合，以得到第一图像处理模型调整后的模型参数。

本公开实施例的模型预训练装置，通过获取样本对，并采用第一图像处理模型对
20 样本对中的第一图像进行特征提取，以得到第一特征，以及采用第二图像处理模型对样本对中的第二图像进行特征提取，以得到第二特征，根据第二特征更新第二图像处理模型的历史输出特征集合，以根据第一特征与更新后的历史输出特征集合之间的差异，对第一图像处理模型和第二图像处理模型进行预训练。由此，可以实现对不同的模型进行联合预训练，以提升模型预测效果。

为了实现上述实施例，本公开还提供一种电子设备，电子设备可以包括上述实施
25 例中的主播客户端或服务端，该电子设备可以包括至少一个处理器；以及与至少一个处理器通信连接的存储器；其中，存储器存储有可被至少一个处理器执行的指令，指令被至少一个处理器执行，以使至少一个处理器能够执行本公开上述任一实施例提出的模型预训练方法。

为了实现上述实施例，本公开还提供一种存储有计算机指令的非瞬时计算机可读
30 存储介质，其中，计算机指令用于使计算机执行本公开上述任一实施例提出的模型预训练方法。

为了实现上述实施例，本公开还提供一种计算机程序产品，该计算机程序产品包
括计算机程序，计算机程序在被处理器执行时实现本公开上述任一实施例提出的模型

预训练方法。

为了实现上述实施例，本公开还提供一种计算机程序，该计算机程序包括计算机程序代码，当所述计算机程序代码在计算机上运行时，使得计算机执行本公开上述任一实施例提出的模型预训练方法。

5 根据本公开的实施例，本公开还提供了一种电子设备、一种非瞬时计算机可读存储介质、一种计算机程序产品和一种计算机程序。

图 7 示出了可以用来实施本公开的实施例的示例电子设备的示意性框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本公开的实现。

如图 7 所示，设备 700 包括计算单元 701，其可以根据存储在 ROM（Read-Only Memory，只读存储器）702 中的计算机程序或者从存储单元 707 加载到 RAM（Random Access Memory，随机访问/存取存储器）703 中的计算机程序，来执行各种适当的动作和处理。在 RAM 703 中，还可存储设备 700 操作所需的各种程序和数据。计算单元 701、ROM 702 以及 RAM 703 通过总线 704 彼此相连。I/O（Input/Output，输入/输出）接口 705 也连接至总线 704。

设备 700 中的多个部件连接至 I/O 接口 705，包括：输入单元 706，例如键盘、鼠标等；输出单元 707，例如各种类型的显示器、扬声器等；存储单元 708，例如磁盘、光盘等；以及通信单元 709，例如网卡、调制解调器、无线通信收发机等。通信单元 709 允许设备 700 通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

计算单元 701 可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元 701 的一些示例包括但不限于 CPU（Central Processing Unit，中央处理单元）、GPU（Graphic Processing Units，图形处理单元）、各种专用的 AI（Artificial Intelligence，人工智能）计算芯片、各种运行机器学习模型算法的计算单元、DSP（Digital Signal Processor，数字信号处理器）、以及任何适当的处理器、控制器、微控制器等。计算单元 701 执行上文所描述的各个方法和处理，例如上述模型预训练方法。例如，在一些实施例中，上述模型预训练方法可被实现为计算机软件程序，其被有形地包含于机器可读介质，例如存储单元 708。在一些实施例中，计算机程序的部分或者全部可以经由 ROM 702 和/或通信单元 709 而被载入和/或安装到设备 700 上。当计算机程序加载

到 RAM 703 并由计算单元 701 执行时，可以执行上文描述的模型预训练方法的一个或多个步骤。备选地，在其他实施例中，计算单元 701 可以通过其他任何适当的方式（例如，借助于固件）而被配置为执行上述模型预训练方法。

本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、FPGA（Field Programmable Gate Array，现场可编程门阵列）、ASIC（Application-Specific Integrated Circuit，专用集成电路）、ASSP（Application Specific Standard Product，专用标准产品）、SOC（System On Chip，芯片上系统的系统）、CPLD（Complex Programmable Logic Device，复杂可编程逻辑设备）、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器，使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行，作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、RAM、ROM、EPROM（Electrically Programmable Read-Only-Memory，可擦除可编程只读存储器）或快闪存储器、光纤、CD-ROM（Compact Disc Read-Only Memory，便捷式紧凑盘只读存储器）、光学储存设备、磁储存设备、或上述内容的任何合适组合。

为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置（例如，CRT（Cathode-Ray Tube，阴极射线管）或者 LCD（Liquid Crystal Display，液晶显示器）监视器）；以及键盘和指向装置（例如，鼠标或者轨迹球），用户可以通过该键盘和该指向装置来将输入提供给计算机。

其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈（例如，视觉反馈、听觉反馈、或者触觉反馈）；并且可以用任何形式（包括声输入、语音输入或者、触觉输入）来接收来自用户的输入。

可以将此处描述的系统和技术实施在包括后台部件的计算系统（例如，作为数据服务器）、或者包括中间件部件的计算系统（例如，应用服务器）、或者包括前端部件的计算系统（例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互）、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信（例如，通信网络）来将系统的部件相互连接。通信网络的示例包括：LAN（Local Area Network，局域网）、WAN（Wide Area Network，广域网）、互联网和区块链网络。

计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器，又称为云服务器或云主机，是云计算服务体系中的一项主机产品，以解决了传统物理主机与 VPS 服务（Virtual Private Server，虚拟专用服务器）中，存在的管理难度大，业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器，或者是结合了区块链的服务器。

其中，需要说明的是，人工智能是研究使计算机来模拟人的某些思维过程和智能行为（如学习、推理、思考、规划等）的学科，既有硬件层面的技术也有软件层面的技术。人工智能硬件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、大数据处理等技术；人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术以及机器学习/深度学习、大数据处理技术、知识图谱技术等几大方向。

根据本公开实施例的技术方案，通过获取样本对，并采用第一图像处理模型对样本对中的第一图像进行特征提取，以得到第一特征，以及采用第二图像处理模型对样本对中的第二图像进行特征提取，以得到第二特征，根据第二特征更新第二图像处理模型的历史输出特征集合，以根据第一特征与更新后的历史输出特征集合之间的差异，对第一图像处理模型和第二图像处理模型进行预训练。由此，可以实现对不同的模型进行联合预训练，以提升模型预测效果。

应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本公开公开的技术方案所期望的结果，本文在此不进行限制。

上述具体实施方式，并不构成对本公开保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等，均应包含在本公开保护范围之内。

权利要求书

1. 一种模型预训练方法，包括：

获取样本对；其中，所述样本对中包括相同或相似的第一图像和第二图像；

5 采用第一图像处理模型对所述第一图像进行特征提取，以得到第一特征；

采用第二图像处理模型对所述第二图像进行特征提取，以得到第二特征；

根据所述第二特征更新所述第二图像处理模型的历史输出特征集合；

根据所述第一特征与更新后的所述历史输出特征集合之间的差异，对所述第一图像处理模型和所述第二图像处理模型进行预训练。

10

2. 根据权利要求1所述的方法，其中，所述采用第一图像处理模型对所述第一图像进行特征提取，以得到第一特征，包括：

采用待执行的任务对应的第一转换层将所述第一图像转化为第一输入序列；

将所述第一输入序列添加到所述第一图像处理模型的第一目标队列中；

15 在所述第一目标队列轮询到所述第一输入序列的情况下，将所述第一输入序列输入到所述第一图像处理模型进行特征提取；其中，所述第一目标队列是所述第一图像处理模型的多个队列中所述任务所对应的队列。

3. 根据权利要求1或2所述的方法，其中，所述采用第二图像处理模型对所述第二图像进行特征提取，以得到第二特征，包括：

20

采用所述任务对应的第二转化层将所述第二图像转化为第二输入序列；

将所述第二输入序列添加到所述第二图像处理模型对应的第二目标队列中；

在所述第一目标队列将所述第一输入序列输入所述第一图像处理模型的情况下，同步将属于相同样本对的所述第二输入序列输入到所述第二图像处理模型。

25

4. 根据权利要求3所述的方法，其中，所述任务为多个，对应地，各所述任务具有对应的所述第二输入序列，且各所述任务具有对应的所述历史输出特征集合；

所述根据所述第二特征更新所述第二图像处理模型的历史输出特征集合，包括：

30 将所述第二图像处理模型对各任务对应的第二输入序列所提取的第二特征，添加到对应任务的所述历史输出特征集合，并删除对应任务的所述历史输出特征集合中的历史输出特征。

5. 根据权利要求 1-4 中任一项所述的方法，其中，所述根据所述第一特征与更新后的所述历史输出特征集合之间的差异，对所述第一图像处理模型和所述第二图像处理模型进行预训练，包括：

针对每个所述任务，根据所述第一图像处理模型对各所述任务对应的第一输入序列进行特征提取所得到的第一特征，与对应任务更新后的所述历史输出特征集合之间的差异，得到各任务的对比损失函数值；

对各所述任务的对比损失函数值加权求和得到目标对比损失函数值；

根据所述目标对比损失函数值，对所述第一图像处理模型和所述第二图像处理模型进行模型参数调整。

6. 根据权利要求 3 所述的方法，还包括：

将所述任务对应的第一位置编码添加到所述第一输入序列中，其中，所述第一位置编码用于指示所述第一图像处理模型采用对应的注意力权重；

将所述任务对应的第二位置编码添加到所述第二输入序列中，其中，所述第二位置编码用于指示所述第二图像处理模型采用对应的注意力权重。

7. 根据权利要求 1-6 中任一项所述的方法，其中，所述根据所述第一特征与更新后的所述历史输出特征集合之间的差异，对所述第一图像处理模型和所述第二图像处理模型进行预训练，包括：

根据所述第一特征与更新后的所述历史输出特征集合之间的差异，确定目标对比损失函数值；

根据所述目标对比损失函数值，对所述第二图像处理模型进行模型参数调整；

根据所述第二图像处理模型参数调整前和调整后的模型参数变化量，对所述第一图像处理模型进行模型参数调整。

8. 根据权利要求 7 所述的方法，其中，所述根据所述第二图像处理模型参数调整前和调整后的模型参数变化量，对所述第一图像处理模型进行模型参数调整，包括：

将所述第二图像处理模型参数调整前和调整后的模型参数变化量，与所述第一图像处理模型的模型参数进行加权融合，以得到所述第一图像处理模型调整后的模型参数。

9. 一种模型预训练装置，包括：

获取模块，用于获取样本对；其中，所述样本对中包括相同或相似的第一图像和第二

图像；

第一提取模块，用于采用第一图像处理模型对所述第一图像进行特征提取，以得到第一特征；

5 第二提取模块，用于采用第二图像处理模型对所述第二图像进行特征提取，以得到第二特征；

更新模块，用于根据所述第二特征更新所述第二图像处理模型的历史输出特征集合；

预训练模块，用于根据所述第一特征与更新后的所述历史输出特征集合之间的差异，对所述第一图像处理模型和所述第二图像处理模型进行预训练。

10 10. 根据权利要求 9 所述的装置，其中，所述第一提取模块，具体用于：
采用待执行的任务对应的第一转换层将所述第一图像转化为第一输入序列；
将所述第一输入序列添加到所述第一图像处理模型的第一目标队列中；

15 在所述第一目标队列轮询到所述第一输入序列的情况下，将所述第一输入序列输入到所述第一图像处理模型进行特征提取；其中，所述第一目标队列是所述第一图像处理模型的多个队列中所述任务所对应的队列。

11. 根据权利要求 9 或 10 所述的装置，其中，所述第二提取模块，具体用于：
采用所述任务对应的第二转化层将所述第二图像转化为第二输入序列；
将所述第二输入序列添加到所述第二图像处理模型对应的第二目标队列中；

20 在所述第一目标队列将所述第一输入序列输入所述第一图像处理模型的情况下，同步将属于相同样本对的所述第二输入序列输入到所述第二图像处理模型。

12. 根据权利要求 11 所述的装置，其中，所述任务为多个，对应地，各所述任务具有对应的所述第二输入序列，且各所述任务具有对应的所述历史输出特征集合；

25 所述更新模块，具体用于：

将所述第二图像处理模型对各任务对应的第二输入序列所提取的第二特征，添加到对应任务的所述历史输出特征集合，并删除对应任务的所述历史输出特征集合中的历史输出特征。

30 13. 根据权利要求 9-12 中任一项所述的装置，其中，所述预训练模块，具体用于：

针对每个所述任务，根据所述第一图像处理模型对各所述任务对应的第一输入序列进行特征提取所得到的第一特征，与对应任务更新后的所述历史输出特征集合之间的差异，

得到各任务的对比损失函数值；

对各所述任务的对比损失函数值加权求和得到目标对比损失函数值；

根据所述目标对比损失函数值，对所述第一图像处理模型和所述第二图像处理模型进行模型参数调整。

5

14. 根据权利要求 11 所述的装置，还包括：

添加模块，用于将所述任务对应的第一位置编码添加到所述第一输入序列中，其中，所述第一位置编码用于指示所述第一图像处理模型采用对应的注意力权重；将所述任务对应的第二位置编码添加到所述第二输入序列中，其中，所述第二位置编码用于指示所述第二图像处理模型采用对应的注意力权重。

10

15. 根据权利要求 9-14 中任一项所述的装置，其中，所述预训练模块，具体用于：

根据所述第一特征与更新后的所述历史输出特征集合之间的差异，确定目标对比损失函数值；

15

根据所述目标对比损失函数值，对所述第二图像处理模型进行模型参数调整；

根据所述第二图像处理模型参数调整前和调整后的模型参数变化量，对所述第一图像处理模型进行模型参数调整。

16. 根据权利要求 15 所述的装置，其中，所述预训练模块，具体用于：

20

将所述第二图像处理模型参数调整前和调整后的模型参数变化量，与所述第一图像处理模型的模型参数进行加权融合，以得到所述第一图像处理模型调整后的模型参数。

17. 一种电子设备，包括：

至少一个处理器；以及

25

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求 1-8 中任一项所述的模型预训练方法。

30

18. 一种存储有计算机指令的非瞬时计算机可读存储介质，其中，所述计算机指令用于使所述计算机执行根据权利要求 1-8 中任一项所述的模型预训练方法。

19. 一种计算机程序产品, 包括计算机程序, 所述计算机程序在被处理器执行时实现根据权利要求 1-8 中任一项所述的模型预训练方法。

20. 一种计算机程序, 包括计算机程序代码, 当所述计算机程序代码在计算机上运行时,
5 使得计算机执行根据权利要求 1-8 中任一项所述的模型预训练方法。

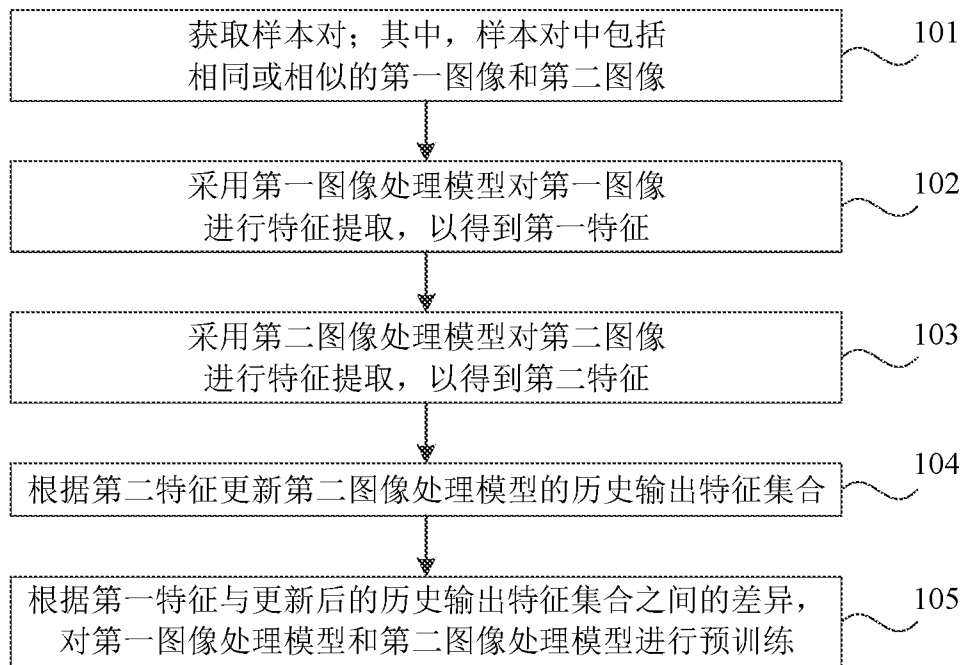


图 1

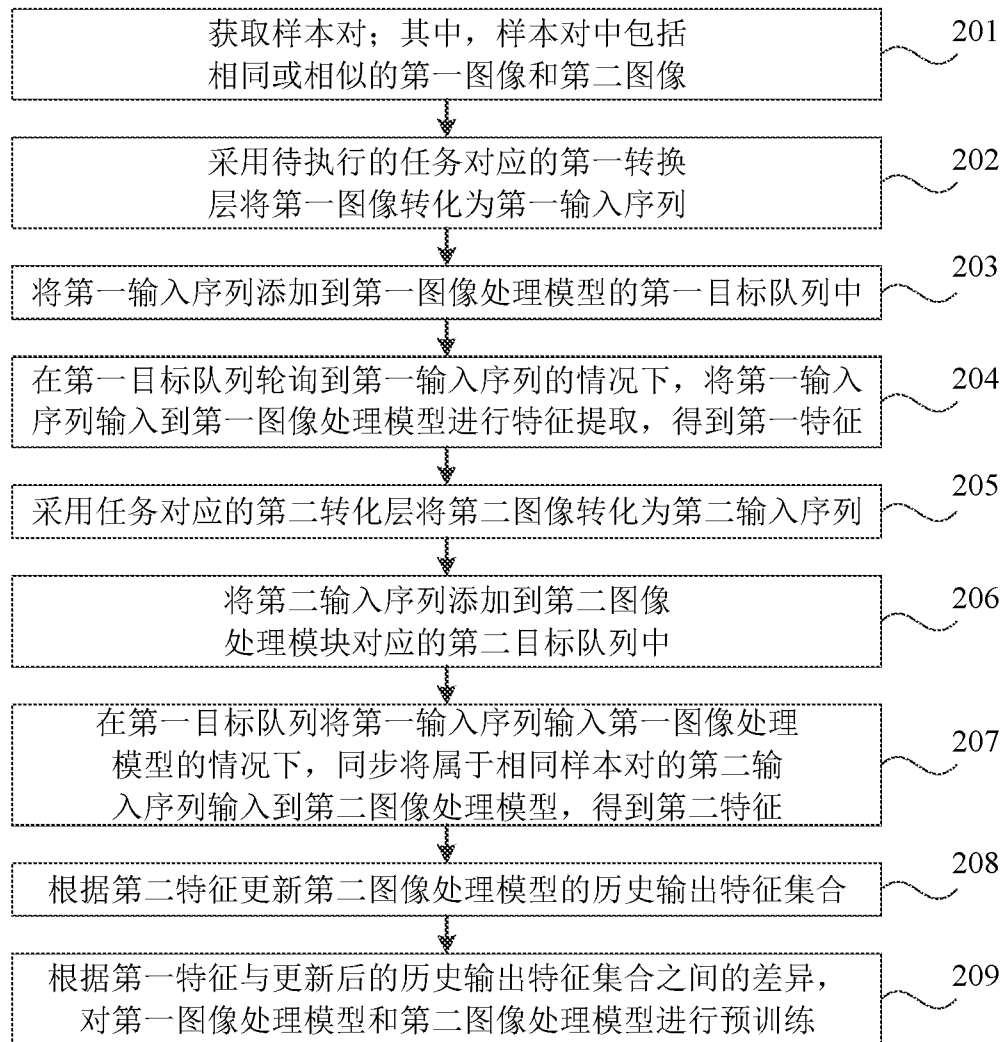


图 2

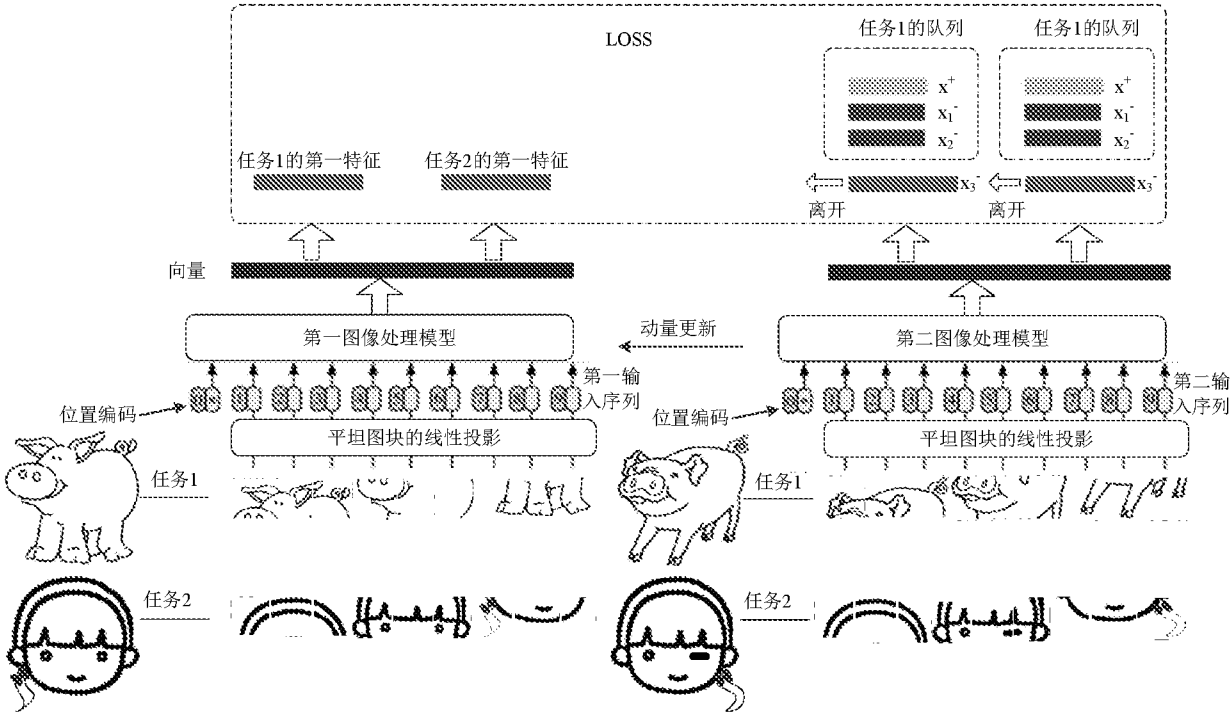


图 3

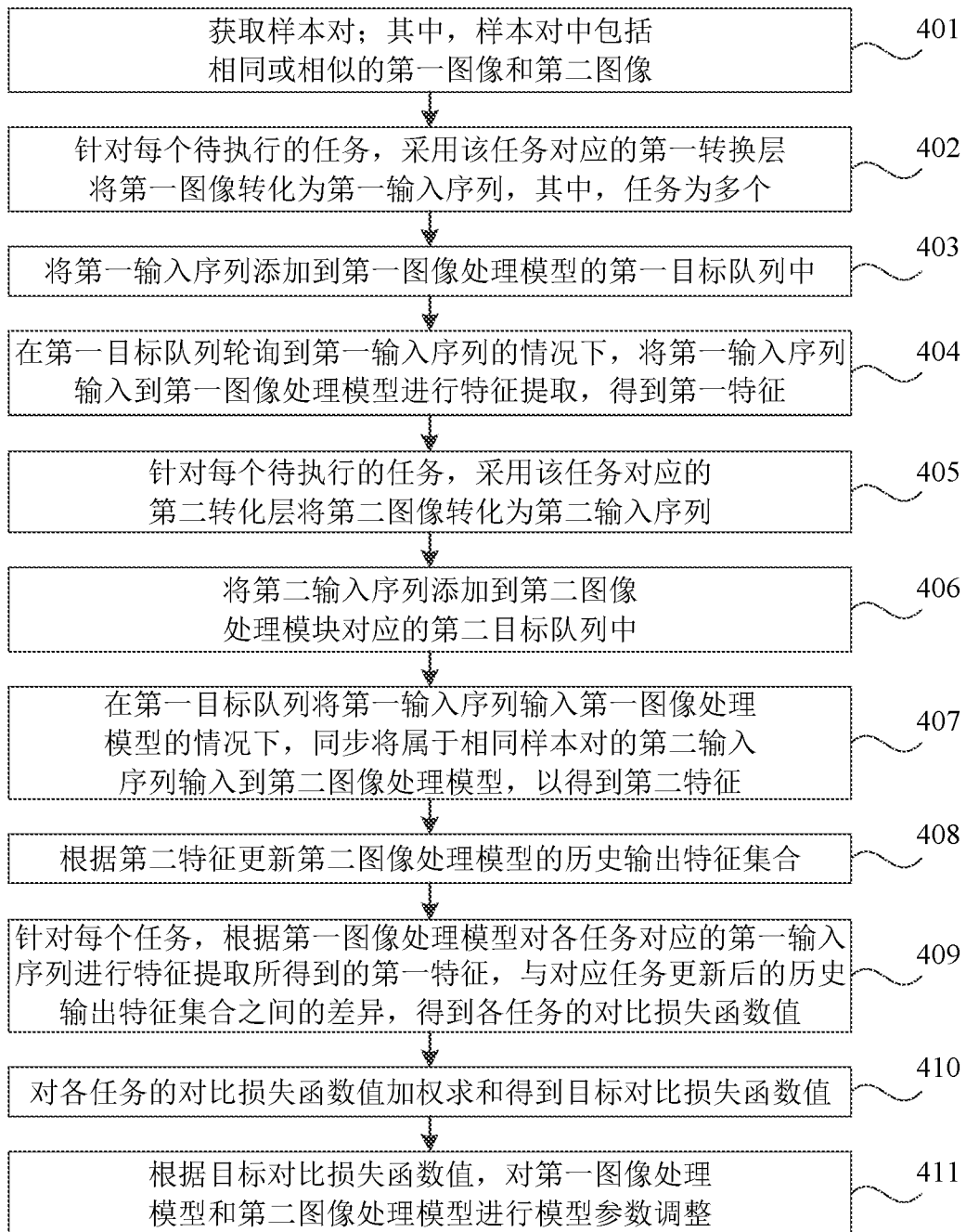


图 4

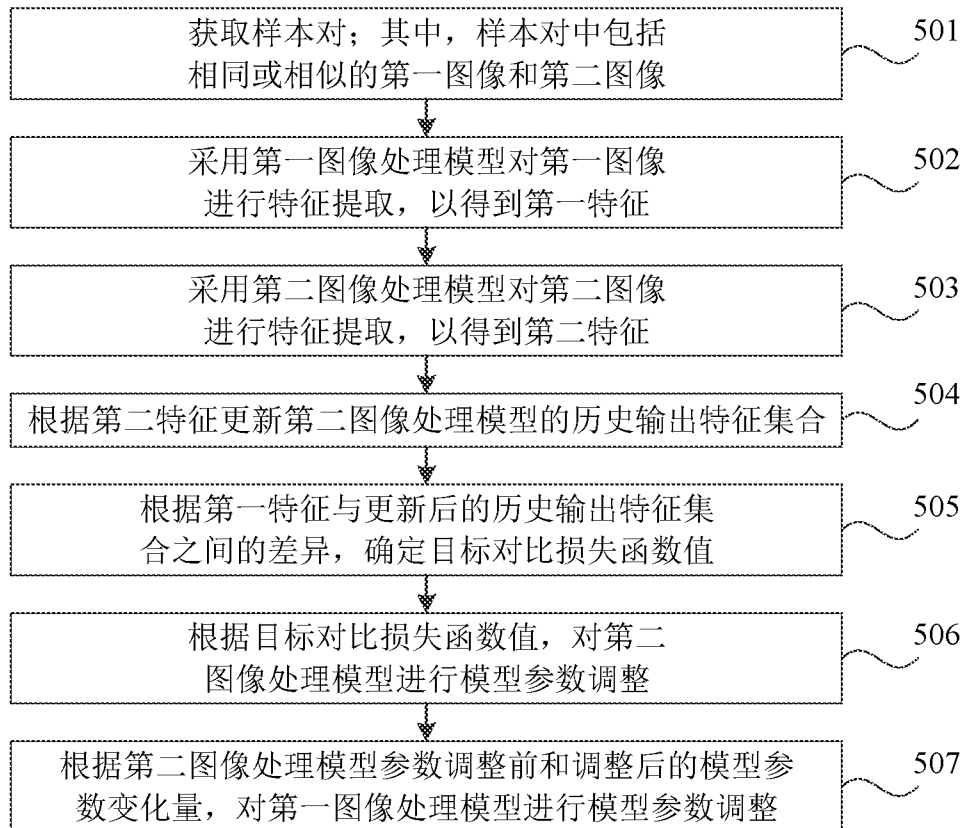


图 5

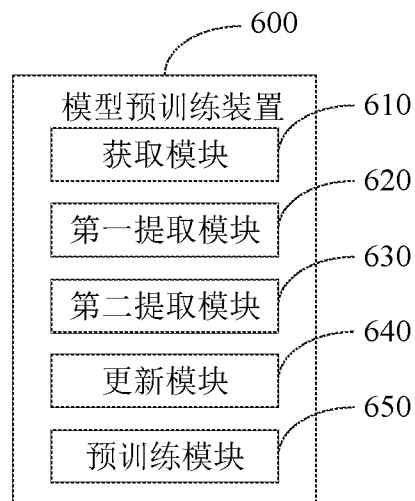


图 6

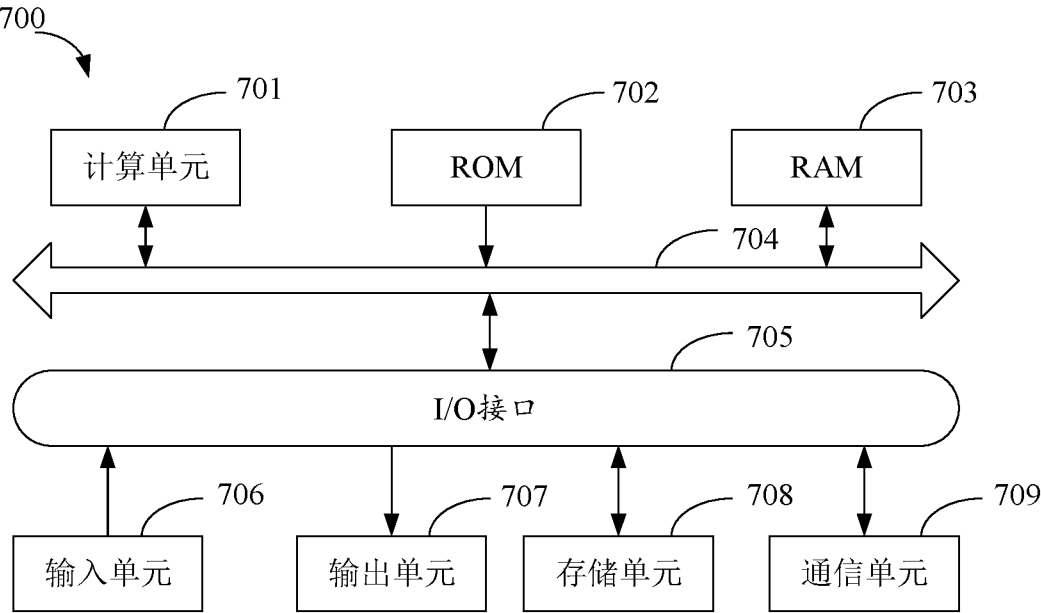


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/088612

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(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)

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)

CNTXT, ENTXT, ENTXTC, DWPI, CNKI: 图像, 样本, 训练, 模型, image, sample, train+, model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113657467 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 16 November 2021 (2021-11-16) claims 1-19, and description, paragraphs 0032-0177	1-20
X	LI, Bi et al. "Dynamic Class Queue for Large Scale Face Recognition In the Wild" <i>arxiv.org</i> , 24 May 2021 (2021-05-24), text, section 3, and figure 2	1, 5, 7, 8, 9, 13, 15, 16, 17-20
Y	LI, Bi et al. "Dynamic Class Queue for Large Scale Face Recognition In the Wild" <i>arxiv.org</i> , 24 May 2021 (2021-05-24), text, section 3, and figure 2	2-8,10-16,17-20
Y	DOSOVITSKIY, A. et al. "An Image is Worth 16*16 Words: Transformers for Image Recognition at Scale" <i>arXiv.org</i> , 03 June 2021 (2021-06-03), text, section 3, and figure 1	2-8, 10-16, 17-20



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

12 June 2022

Date of mailing of the international search report

16 June 2022

Name and mailing address of the ISA/CN

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/088612

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	113657467	A	16 November 2021	None	

国际检索报告

国际申请号

PCT/CN2022/088612

A. 主题的分类

G06K 9/62 (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)

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))

CNTXT, ENTXT, ENTXTC, DWPI, CNKI:图像, 样本, 训练, 模型, image, sample, train+, model

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 113657467 A (北京百度网讯科技有限公司) 2021年11月16日 (2021 - 11 - 16) 权利要求1-19, 说明书第0032-0177段	1-20
X	Li B等. "Dynamic Class Queue for Large Scale Face Recognition In the Wild" arxiv.org, 2021年5月24日 (2021 - 05 - 24), 正文第3部分, 附图2	1, 5, 7, 8, 9, 13, 15, 16, 17-20
Y	Li B等. "Dynamic Class Queue for Large Scale Face Recognition In the Wild" arxiv.org, 2021年5月24日 (2021 - 05 - 24), 正文第3部分, 附图2	2-8, 10-16, 17-20
Y	Dosovitskiy A等. "An Image is Worth 16*16 Words: Transformers for Image Re- cognition at Scale" arXiv.org, 2021年6月3日 (2021 - 06 - 03), 正文第3部分, 附图1	2-8, 10-16, 17-20

☐ 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

“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

2022年6月12日

国际检索报告邮寄日期

2022年6月16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

传真号 (86-10)62019451

受权官员

韩鲜萍

电话号码 62411841

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/088612

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 113657467 A	2021年11月16日	无	