

(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
2022年8月4日 (04.08.2022)



(10) 国际公布号
WO 2022/161380 A1

- (51) 国际专利分类号:
G06K 9/62 (2022.01)
- (21) 国际申请号: PCT/CN2022/073923
- (22) 国际申请日: 2022年1月26日 (26.01.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110132038.4 2021年1月30日 (30.01.2021) CN
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 李炜棉 (LI, Weimian); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 朱凯强 (ZHU, Kaiqiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 黄飞 (HUANG, Fei); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 许松岑 (XU, Songcen); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (74) 代理人: 深圳市深佳知识产权代理事务所 (普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦B座18C2、18D、18E、18E2, Guangdong 518001 (CN)。

(54) Title: MODEL TRAINING METHOD AND APPARATUS, AND IMAGE RETRIEVAL METHOD AND APPARATUS

(54) 发明名称: 一种训练模型的方法、图像检索的方法以及装置

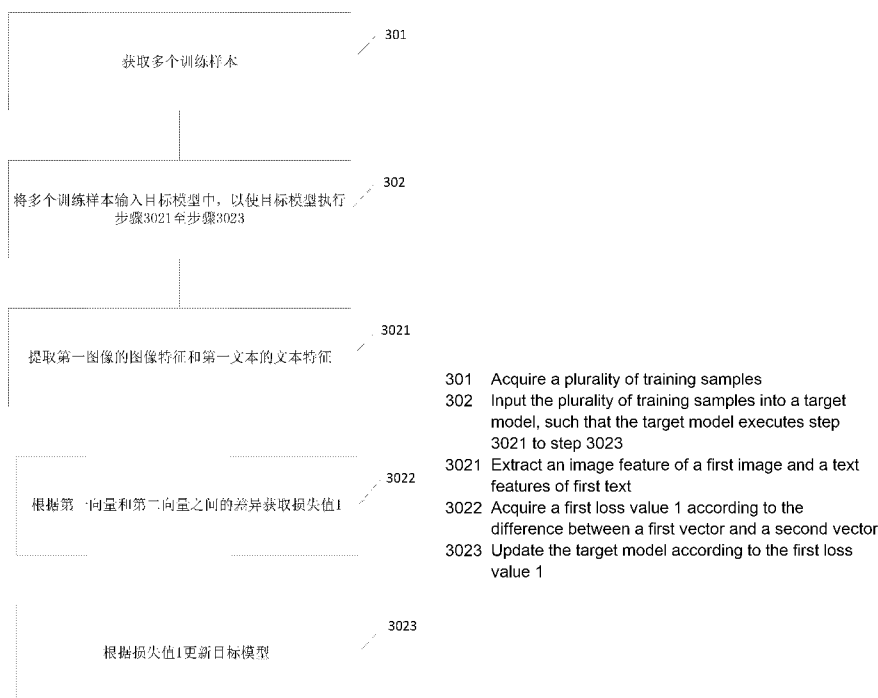


图3

(57) Abstract: The present application relates to the field of artificial intelligence. Disclosed is a model training method. The method comprises: acquiring a plurality of training samples, wherein each training sample comprises an image and text, and the text is used for describing a target object in the image; and inputting the plurality of training samples into a target model, such that the target model executes the following process until a preset stop condition is satisfied: extracting an image feature of a first image and a text feature of first text; acquiring a first loss value according to the difference between a first vector and a second vector, wherein the dimension



WO 2022/161380 A1

(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, 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。

(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))。

of the first vector is the same as that of the second vector, the first vector is used for indicating the image feature of the first image, and the second vector is used for indicating the text feature of the first text; and updating the target model according to the first loss value. By means of the solution provided in the present application, the learning of an image feature is guided by using a text feature, such that an image feature extracted by means of a target model has a better representation capability for an image.

(57) 摘要: 本申请公开了一种训练模型的方法, 涉及人工智能领域, 包括: 获取多个训练样本, 每个训练样本包括图像和文本, 文本用于描述图像中的目标对象。将多个训练样本输入目标模型中, 以使目标模型执行下述流程, 直至满足预设的停止条件: 提取第一图像的图像特征和第一文本的文本特征。根据第一向量和第二向量之间的差异获取第一损失值。第一向量的维度和第二向量的维度相同, 第一向量用于指示第一图像的图像特征, 第二向量用于指示第一文本的文本特征。根据第一损失值更新目标模型。本申请提供的方案利用文本特征指导图像特征的学习, 使通过目标模型提取的图像特征对图像的表征能力更好。

一种训练模型的方法、图像检索的方法以及装置

本申请要求于 2021 年 1 月 30 日提交中国专利局、申请号为 202110132038.4、申请名称为“一种训练模型的方法、图像检索的方法以及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及图像处理领域，尤其涉及一种训练模型的方法、图像检索的方法以及装置。

背景技术

10 人工智能 (artificial intelligence, AI) 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分支，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作出反应的智能机器。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。人工
15 智能领域的研究包括机器人，自然语言处理，计算机视觉，决策与推理，人机交互，推荐与搜索，AI 基础理论等。

图像检索是计算机视觉中备受关注的问题之一，而产品的图像检索是图像检索中非常重要的部分。产品的图像检索是指根据用户输入的产品图像，获取该产品图像的图像特征，根据该图像特征从图像数据库中检索出对应的产品，并向用户展示。产品的图像检索技术
20 在电商行业中有广泛的应用，商业价值巨大。

产品图像检索要求返回的结果与查询图像对应的产品完全一致。然而，图像数据库中可能包括大量外观相似的产品，这给产品图像检索带来了很大的挑战。此外，由于不同用户拍摄产品时拍照角度、拍照环境存在差异，导致输入的产品图像和图像数据库中的产品图像可能有很大差异。因此，如何提升产品图像检索的精度亟待解决。

25 发明内容

本申请实施例提供一种训练模型的方法，图像检索的方法以及装置。通过本申请实施例提供的一种训练模型的方法获取的图像特征执行图像检索任务，可以有效提升图像检索的精度、图像检索的速度、并且降低对图像数据库的存储性能的要求。

为达到上述目的，本申请实施例提供如下技术方案：

30 本申请第一方面提供一种训练模型的方法，可以包括：获取多个训练样本，每个训练样本可以包括图像和文本，文本用于描述图像中的目标对象。其中目标对象可以是一个也可以是多个。将多个训练样本输入目标模型中，以使目标模型执行下述流程，直至满足预设的停止条件：提取第一图像的图像特征和第一文本的文本特征。选择不同的图像特征提取模型提取各个训练样本中的图像的图像特征。比如可以采用 Resnet50 或者 SEResnet50
35 提取图像特征，再比如，可以采用 Efficient-NetB3 模型提取图像特征。选择不同的文本特征提取模型提取各个训练样本中的文本的文本特征比如，可以采用 GPT 模型提取文本特征，再比如可以采用 Bert 模型提取文本特征。第一图像是多个训练样本中任意一个训练样本中可以包括的图像，第一文本用于描述第一图像中的目标对象。根据第一向量和第二向

量之间的差异获取第一损失值，第一向量的维度和第二向量的维度相同，第一向量用于指示第一图像的图像特征，第二向量用于指示第一文本的文本特征。根据第一损失值更新目标模型。本申请提供的方案通过在同构空间内减少图像特征与文本特征差异性。其中，预设的停止条件可以理解为目标模型收敛了，或者达到预设数目的迭代轮次。由于模型的训练过程是不断的降低损失值的过程，当通过损失值 1 更新目标模型时，损失值 1 不断减小，则同一个产品的图像特征对应的向量和文本特征对应的向量之间的距离也会更靠近。相比于只通过图像特征对目标模型进行训练，本申请提供的方案在目标模型的训练过程中，融合了文本特征。通过本申请提供的方案训练后的目标模型提取的图像特征对图像的表征能力更好。

在一种可能的实施方式中，该方法还可以包括：根据第一图像的图像特征，预测第一图像中的目标对象的至少一个属性。根据至少一个属性和预设标签之间的差异获取第二损失值，预设标签是对第一文本进行分词预处理后获取的。根据第一损失值更新目标模型，可以包括：根据第一损失值和第二损失值更新目标模型。为了加深利用文本特征对目标模型提取图像特征的指导，在预测图像中包括的目标对象的多个属性时，利用文本中的关键词作为预设标签进行多分类监督。使目标模型根据图像特征对文本描述的目标对象进行属性预测时，预测结果（至少一个属性）更接近预设标签。

在一种可能的实施方式中，该方法还可以包括：根据第一图像的图像特征，预测第一图像中的目标对象属于对象集合中各个目标对象的第一概率，对象集合可以包括多个训练样本中全部图像中可以包括的目标对象。根据第一文本的文本特征预测第一文本用于描述对象集合中各个目标对象的第二概率。根据第一概率和第二概率之间的差异获取第三损失值。根据第一损失值更新目标模型，可以包括：根据第一损失值、第二损失值以及第三损失值更新目标模型。为了更进一步的提升目标模型的性能，进一步提升目标模型提取的图像特征对图像的表征能力，还可以使根据图像特征预测的分类结果和根据文本特征预测的分类结果更接近。其中，分类结果是指预测的产品属于对象集合中各个目标对象的概率。其中，对象集合是指全部训练样本中全部图像中包括的目标对象。

在一种可能的实施方式中，该方法还可以包括：根据第一概率和第一预设值之间的差异，获取第四损失值，第一预设值指示预设的第一图像中的目标对象属于对象集合中各个目标对象的概率。根据第二概率和第二预设值之间的差异，获取第五损失值，第二预设值指示预设的第一文本用于描述对象集合中各个目标对象的概率。第一损失值更新目标模型，可以包括：根据第一损失值、第二损失值、第三损失值、第四损失值以及第五损失值更新目标模型。为了更进一步的提升目标模型的性能，进一步提升目标模型提取的图像特征对图像的表征能力，还可以使根据图像特征预测的分类结果与图像中真实包括的目标对象的分类更接近，还可以使根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类更接近。

在一种可能的实施方式中，该方法还可以包括：根据第一图像的图像特征和第二图像的图像特征之间的差异获取第六损失值，第二图像和第一图像从不同角度展示相同的目标对象。根据第一损失值更新目标模型，可以包括：根据第一损失值、第二损失值、第三损

失值、第四损失值、第五损失值以及第六损失值更新目标模型。为了更进一步的提升目标模型的性能，进一步提升目标模型提取的图像特征对图像的表征能力，还可以使同款产品的图像特征更加靠近，不同款产品的图像特征更远离。

5 在一种可能的实施方式中，该方法还可以包括：获取第一向量和第三向量之间的第一距离，获取第一向量和第四向量之间的第二距离，第三向量用于指示第二图像的图像特征，第四向量用于指示第三图像的图像特征，第三图像和第一图像中可以包括的目标对象不同。根据第一距离和第二距离之间的差值与预设阈值的差异获取第七损失值。根据第一损失值更新目标模型，可以包括：根据第一损失值、第二损失值、第三损失值、第四损失值、第五损失值、第六损失值以及第七损失值更新目标模型。为了更进一步的提升目标模型的性能，进一步提升目标模型提取的图像特征对图像的表征能力，使不同产品的图像特征具有足够的区分性。

10 在一种可能的实施方式中，获取多个训练样本，可以包括：获取产品的第一目标图像、产品的文本标题、产品的第二目标图像，第一目标图像来自于产品的提供者，第二目标图像来自产品的购买者。将第一目标图像作为分割模型的输入，以获取分割结果，分割结果指示产品在第一目标图像中的轮廓。根据分割结果提取第三目标图像的目标区域，第三目标图像是通过将第一目标图像输入生成器中获取的，其中生成器在训练过程中利用了来自产品的提供者提供的图像，以及来自产品的购买者提供的图像。对目标区域进行随机仿射变换，以获取变换后的目标区域。将变换后的目标区域与预设背景进行组合，以获取第四目标图像。获取第一类型的训练样本、第二类型的训练样本以及第三类型的训练样本，每个第一类型的训练样本可以包括第一目标图像和文本标题，每个第二类型的训练样本可以包括第二目标图像和文本标题，每个第三类型的训练样本可以包括第四目标图像和文本标题。在这种实施方式中，在买家秀缺失的情况下，通过卖家秀生成大量的买家秀数据，使得训练后的目标模型对于复杂的买家秀图像风格具有一定的鲁棒性，并且在一定程度上能够缓解用户拍摄视角大角度变化造成的检索难度。

25 在一种可能的实施方式中，该方法还可以包括：根据目标模型提取的多个图像特征，对哈希模型进行多次训练，以获取训练后的哈希模型，其中，多次训练中的任意一次训练，可以包括：获取每个图像特征对应的哈希特征。对多个哈希特征进行聚类处理，以获取多个聚类数据集。从多个哈希特征中获取第一哈希特征，第一哈希特征是多个哈希特征中的任意一个。根据各个聚类数据集的聚类中心和第一哈希特征的相似性，从多个聚类数据集中获取第一聚类数据集，多个聚类数据中第一聚类数据集的聚类中心和第一哈希特征的相似性最高。根据第一聚类数据集的聚类中心和第一哈希特征的差异获取第八损失值。根据第八损失值更新哈希模型。为了在提升图像检索精度的同时，还可以提升图像检索的速度，降低对图像数据库的存储能力的要求，本申请提供的方案提供了一种端到端的设计，对哈希特征和倒排码本进行联合的优化。

35 在一种可能的实施方式中，根据第一损失值更新目标模型，可以包括：根据第一损失值、第二损失值、第三损失值、第四损失值、第五损失值、第六损失值、第七损失值以及第八损失值更新目标模型。

在一种可能的实施方式中，多次训练中的任意一次训练，还可以包括：对目标哈希特征施加第一约束和第二约束，以获取更新后的目标哈希特征，目标哈希特征是根据上一次更新后的哈希模型获取的哈希特征，第一约束为哈希特征中的第三预设值和第四预设值的数目是平均的，第二约束为目标矩阵的第一对角线上的元素为第四预设值，目标矩阵中除
5 第一对角线之外的其他元素为第三预设值，目标矩阵中第 i 行第 j 列的元素表示多个目标元素的平均值，多个目标元素表示各个哈希特征的第 i 个维度的取值和各个哈希特征各自的第 j 个维度的取值的乘积， i 和 j 为正整数。对目标哈希特征中每一维度的取值和第二均值进行比较，第二均值是第三预设值和第四预设值的平均值。每一维度的取值小于第二均值时，根据哈希特征的每一维度的取值和第三预设值之间的差异获取第九损失值。每一
10 维度的取值不小于第二均值时，根据哈希特征的每一维度的取值和第四预设值之间的差异获取第九损失值。根据第八损失值更新哈希模型，可以包括：根据第八损失值和第九损失值更新哈希模型。

本申请第二方面提供一种图像检索的方法，可以包括：获取待检索图像。通过目标模型提取待检索图像的图像特征。根据待检索图像的图像特征从图像数据库中获取预测图像，
15 预测图像的图像特征和待检索图像的图像特征的相似度达到阈值。其中，目标模型是通过第一损失值更新初始目标模型后获取的模型。初始目标模型可以是预先获取的模型。第一损失值是通过第一向量和第二向量之间的差异获取的，第一向量的维度和第二向量的维度相同，第一向量用于指示第一图像的图像特征，第二向量用于指示第一文本的文本特征，第一图像是多个训练样本中任意一个训练样本中可以包括的图像，第一文本用于描述第一
20 图像中的目标对象，每个训练样本可以包括图像和文本，文本用于描述图像中的目标对象。

在一种可能的实施方式中，目标模型具体是通过第一损失值和第二损失值更新初始目标模型后获取的模型，第二损失值是根据至少一个属性和预设标签之间的差异获取的，至少一个属性是根据第一图像的图像特征，预测的第一图像中的目标对象的至少一个属性。

在一种可能的实施方式中，目标模型具体是通过第一损失值、第二损失值以及第三损失值更新初始目标模型后获取的，第三损失值是根据第一概率和第二概率之间的差异获取的，第一概率是根据第一图像的图像特征，预测第一图像中的目标对象属于对象集合中各个目标对象的概率，第二概率是根据第一文本的文本特征预测第一文本用于描述对象集合
25 中各个目标对象的概率。

在一种可能的实施方式中，目标模型具体是通过第一损失值、第二损失值、第三损失值、第四损失值以及第五损失值更新初始目标模型后模型的，第四损失值是根据第一概率和第一预设值之间的差异获取的，第一预设值指示预设的第一图像中的目标对象属于对象集合中各个目标对象的概率，第五损失值是根据第二概率和第二预设值之间的差异获取的，第二预设值指示预设的第一文本用于描述对象集合中各个目标对象的概率。
30

在一种可能的实施方式中，目标模型具体是根据第一损失值、第二损失值、第三损失值、第四损失值、第五损失值以及第六损失值更新初始目标模型后获取的，第六损失值是根据第一图像的图像特征和第二图像的图像特征之间的差异获取的，第二图像和第一图像从不同角度展示相同的目标对象。
35

在一种可能的实施方式中，目标模型具体是根据第一损失值、第二损失值、第三损失值、第四损失值、第五损失值、第六损失值以及第七损失值更新初始目标模型后获取的，第七损失值是通过第一距离和第二距离之间的差值与预设阈值的差异获取的，第一距离是根据第一向量和第三向量之间的距离，第二距离是第一向量和第四向量之间的距离，第三向量用于指示第二图像的图像特征，第四向量用于指示第三图像的图像特征，第三图像和第一图像中可以包括的目标对象不同。

在一种可能的实施方式中，多个训练样本可以包括第一类型的训练样本、第二类型的训练样本以及第三类型的训练样本，每个第一类型的训练样本可以包括第一目标图像和文本标题，每个第二类型的训练样本可以包括第二目标图像和文本标题，每个第三类型的训练样本可以包括第四目标图像和文本标题，第一目标图像来自于产品的提供者，第二目标图像来自产品的购买者，文本标题是产品的文本标题，第四目标图像是将变换后的目标区域与预设背景进行组合后获取的，变换后的目标区域是对目标区域进行随机仿射变换后获取的，目标区域是根据分割结果提取第三目标图像后获取的，第三目标图像是通过将第一目标图像输入生成器中获取的，其中生成器在训练过程中利用了来自产品的提供者提供的图像，以及来自产品的购买者提供的图像，分割结果是将第一目标图像作为分割模型的输入后获取的结果，分割结果指示产品在第一目标图像中的轮廓。

在一种可能的实施方式中，该方法还可以包括：将待检索图像的图像特征输入至哈希模型中，以获取所待检索图像的哈希特征。根据待检索图像的图像特征从图像数据库中获取预测图像，可以包括：根据所待检索图像的哈希特征从图像数据库中获取预测图像，其中哈希模型是通过第八损失值更新初始哈希模型后获取的，初始哈希模型是预先获取的。第八损失值是根据第一聚类数据集的聚类中心和第一哈希特征的差异获取的，第一聚类数据集的聚类中心是多个聚类数据中各个聚类数据集的聚类中心中和第一哈希特征的相似性最高的聚类中心，多个聚类数据集是对多个哈希特征进行聚类处理后获取的，第一哈希特征是多个哈希特征中的任意一个。

在一种可能的实施方式中，目标模型具体是根据第一损失值、第二损失值、第三损失值、第四损失值、第五损失值、第六损失值、第七损失值以及第八损失值更新初始目标模型后获取的。

本申请第三方面提供一种训练模型的装置，可以包括：存储器，用于存储计算机可读指令。与存储器耦合的处理器，用于执行存储器中的计算机可读指令从而执行如第一方面所描述的方法。

本申请第四方面提供一种图像检索的设备，可以包括：存储器，用于存储计算机可读指令。与存储器耦合的处理器，用于执行存储器中的计算机可读指令从而执行如第二方面所描述方法。

本申请第五方面提供一种芯片系统，芯片系统可以包括处理器和通信接口，处理器通过通信接口获取程序指令，当程序指令被处理器执行时实现第一方面所描述方法。

本申请第六方面提供一种芯片系统，芯片系统可以包括处理器和通信接口，处理器通过通信接口获取程序指令，当程序指令被处理器执行时实现第二方面所描述方法。

本申请第七方面提供一种计算机可读存储介质，可以包括程序，当其被处理单元所执行时，执行如第一方面所描述方法。

本申请第八方面提供一种计算机可读存储介质，可以包括程序，当其被处理单元所执行时，执行如第二方面所描述方法。

5 本申请第九方面一种计算机程序产品，当计算机程序产品在计算机上运行时，使得计算机执行如第一方面所描述方法。

本申请第十方面提供一种计算机程序产品，当计算机程序产品在计算机上运行时，使得计算机执行如第二方面所描述方法。

10 本申请提供的方案利用文本特征指导图像特征的学习，提升通过目标模型提取的图像特征对图像的表征能力。此外，还可以通过对文本进行预处理获取关键词，在预测图像中包括的目标对象的多个属性时，利用文本中的关键词作为预设标签进行多分类监督，进一步提升图像特征的表征能力。通过提升图像特征的表征能力，使图像特征能够更准确的表示图像的特征，有利于提升根据图像特征进行图像检索的精度。

附图说明

- 15 图 1 为本申请实施例提供的方案的一种应用场景示意图；
图 2 为本申请实施例提供的方案的一种应用场景示意图；
图 3 为本申请实施例提供的一种训练模型的方法的流程示意图；
图 4 为申请实施例提供的一种目标模型的执行流程示意图；
图 5 为申请实施例提供的获取均衡的训练样本集合的流程示意图；
20 图 6 为循环生成式对抗网络的结构示意图；
图 7 为申请实施例提供的另一种目标模型的执行流程示意图；
图 8 为申请实施例提供的另一种目标模型的执行流程示意图；
图 9 为申请实施例提供的另一种目标模型的执行流程示意图；
图 10 为本申请实施例提供的一种图像检索方法的流程示意图；
25 图 11 为本申请实施例提供的方案的一种应用场景示意图；
图 12 为本申请实施例提供的方案的一种应用场景示意图；
图 13 为本申请实施例提供的一种训练装置的框图；
图 14 为本申请实施例提供的一种执行设备的框图。

具体实施方式

30 本申请实施例提供了一种训练模型的方法、图像检索的方法以及装置。通过本申请实施例提供的方案，可以提升图像检索的精度、图像检索的速度，降低对图像数据库存储空间的要求。本申请提供的方案，尤其可以提升商品图像检索的精度、商品图像检索的速度，以及降低对商品图像数据库存储空间的要求。

为了更好的理解本申请提供的方案，下面首先对本申请提供的方案可能适用的典型场景进行介绍。

35 如图 1 所示，为本申请实施例提供的方案的一种应用场景示意图。本申请提供的方案的一种典型的应用场景为通过拍照进行产品检索。当用户遇到心仪的产品时，可以通过手

机等终端设备对该产品进行拍照，终端设备可以根据获取到的图像执行图像检索任务，或者终端设备和其他设备（比如云端服务器）共同执行图像检索任务后，可以向用户展示该产品的同款产品的链接。用户可以对展示的至少一款同款产品进行挑选和购买。

以终端设备和云端服务器共同执行图像检索任务为例，对图像检索任务进行解释。终端获取到产品的图像后，可以将该图像上传至云端服务器。云端服务器提取该图像的图像特征后，根据提取的图像特征对图像数据库中存储的数据进行检索，以获取该产品的同款产品，并将同款产品的链接发送至终端上。

目前，产品图像检索至少存在以下困境：

（一）检索精度难以提升

由于图像数据库中往往包括大量的相似产品，各个相似产品在外观上的差异很小，给产品检索带来很大的挑战。比如，参照图 2 进行理解，假设图 2 中的 a 是用户输入的产品图像。由于图 2 中的 b 展示的首饰和图 2 中的 a 展示的首饰在外观上十分相似，当执行了图像检索任务后，可能优先向用户展示图 2 中的 b 对应的产品。此外，用户拍摄的产品图像与图像数据库中存储的产品的图像可能有很大的差异。比如针对于同一款产品，用户拍摄的产品图像是图 2 中的 a，图像数据库中存储的产品图像是图 2 中的 c，由于二者之间的差异，导致在执行图像检索任务时，检索精度低，难以根据图 2 中的 a 从图像数据库中检索到图 2 中的 c。其中差异可以理解为拍摄角度不同，拍摄环境不同等等。

（二）检索速度难以提升

对于一些大规模的图像检索系统，其中包括了大量的产品类目和产品图像。当需要从这些大规模的图像检索系统中检索需要的图像中，往往需要耗费大量的时间。检索时间过长，影响用户的体验，因此如何提升图像检索的速度，特别是提升大规模的图像检索系统的检索速度亟待解决。

（三）对图像数据库的存储能力要求高

提取的图像的图像特征一般是浮点型数据，为了降低对图像数据库的存储能力的要求，可以对图像特征进行转换，比如将浮点型数据转换为二值哈希值。在一种可能的实现方式中，可以通过对图像特征进行哈希处理，以获取哈希特征。图像数据库中存储的是各个图像的哈希特征。通过这样的方式，可以在一定程度上降低对图像数据库的存储能力的要求。然而，对图像特征进行哈希处理，导致图像的表达精度下降，这会带来误差。此外，为了加快检索速度、执行图像检索任务时需要利用倒排码本，而利用倒排码本会再次引入误差。具体的，对多个哈希特征进行聚类处理，以获取多个聚类数据集，多个聚类数据集中的每一个聚类数据集都对应有一个聚类中心，可以将聚类中心理解为一个聚类数据集中包括的全部哈希特征的平均值。全部聚类数据集的全部聚类中心的集合称为倒排码本。在执行图像检索的任务时，获取图像的哈希特征，根据该图像的哈希特征从倒排码本中查找和该哈希特征相似度最高的一个或者多个聚类中心，获取该相似度最高的一个或者多个聚类中心对应的聚类数据集，从该聚类数据集中包括的哈希特征中查找输入图像中包括的同款产品。这种方式，在执行图像检索的任务时，只会对某一个或某几个聚类数据集中包括的哈希特征进行检索，对于其他聚类数据集中包括的哈希特征不再进行检索。如果输入图像中包括

的同款产品在其他聚类数据中，通过这种方式将无法检索到同款产品。因此这种方式，虽然能够降低对图像数据库的存储能力的要求，但是也带来了很大的误差，导致检索精度的下降或者检索效率低。为了解决上述问题，本申请实施例提供了一种模型训练的方法，以及图像检索的方法。通过本申请提供的方案可以在提升检索的精度、检索的效率的同时，降低对图像数据库的存储能力的要求。

为了便于更好的理解本申请，下面具体阐述本申请所描述的技术方案的研究思路：

针对于上述提到的（一）中的检索精度难以提升的问题，一般是因为训练样本集合存在数据不均衡的问题。具体的，来自用户提供的图像和来自产品提供者提供的图像是不均衡的。举例说明，通常可以采用爬虫技术爬取购物类的网页，获取用户提供的图像和来自产品提供者提供的图像。其中，本申请有时也将用户提供的图像称为来自产品购买者提供的图像，或者称为买家秀，他们表示相同的意思。本申请有时也将产品提供者提供的图像称为卖家秀，二者表示相同的意思。其中，一般可以通过爬取产品的评论区获取买家秀，在一些场景中，可能某些产品的评论区没有买家秀，或者买家秀与产品无关，导致针对于这些产品存在买家秀缺失的问题。基于这些问题，导致获取到的买家秀和卖家秀是不均衡的，由获取到的买家秀和卖家秀组成的训练样本集合则是不均衡的训练样本集合。通过不均衡的训练样本集合去训练模型，会使训练后的模型无法很好的提取图像特征。比如，训练数据集中没有足够数量的买家秀，导致模型在训练过程中，无法很好的学习买家秀图像特征的提取。当应用训练后的模型执行图像检索任务时，如果提取的买家秀的图像特征不理想，会影响检索的精度。所以为了能够提升图像检索的精度，应当尽量使训练样本集合是均衡的，换句话说，应当使训练样本集合中卖家秀和买家秀的数据是均衡的。此外，针对上述提到的（一）中的检索精度难以提升的问题，还可能是因为相似产品过多。针对这一问题，其中一种方式为对相似产品进行人工标注，使每个产品有更精准的标签或者属性。然而这种方式，需要投入大量的人力和物力，获取训练样本的周期长，不利于图像检索系统的快速上线。基于这些考虑，本申请提供的方案通过卖家秀生成买家秀，对生成的买家秀根据本申请提供的方案进行进一步的处理，获取大量的接近真实买家秀的图像。通过这种方式，有效的解决了训练样本集合中买家秀数据缺失或者买家秀和产品不匹配的问题，使训练样本集合是均衡的。此外，本申请提供的方案，获取产品的文本描述，比如获取卖家设置的产品的标题。通过在同构空间内减少同款产品的图像特征和文本特征差异，达到通过文本特征指导模型训练的目的，使模型提取的图像特征更理想，可以更好的表达产品的特征。此外，本申请提供的方案不需要人工标注，本申请提供的方案可以利用产品的文本描述获取关键词，通过关键词作为产品的标注。通过这些设计可以进一步提升图像检索的精度。

针对于上述提到的（二）中的检索速度难以提升的问题以及（三）中的对图像数据库的存储能力要求高的问题，可以通过（三）中提到的方式来解决。当通过哈希特征表示图像特征时，表示图像的单位所需要占据的存储空间更小了，进而可以提升检索速度，也会降低对图像数据库的存储能力的要求。其中表示图像的单位可以理解为包括图像特征和哈希特征。然而上文也提到了（三）中的方式会导致多次引入误差的问题，导致检索精度下

降的问题，所以本申请提供的方案是一种端到端的设计，对哈希特征和倒排码本进行联合的优化。

基于上面的研究思路，下面对本申请提供的技术方案进行具体的介绍。

5 以下将从训练侧和应用侧两个方面对本申请提供的方案进行介绍，首先从训练侧对本申请提供的方案进行介绍。

(一) 模型的训练

参阅图3，为本申请实施例提供的一种训练模型的方法的流程示意图。

如图3所示，本申请实施例提供的一种训练模型的方法，可以包括以下步骤：

301、获取多个训练样本。

10 每个训练样本包括图像和文本，文本用于描述图像中的目标对象。换句话说，每个训练样本中包括两种类型的数据，分别是图像数据和文本数据，并且文本数据用于描述图像数据中的目标对象。其中目标对象可以是一个也可以是多个。此外，本申请中的目标对象是产品。比如训练样本 1 中包括图像 1 和文本 1，其中图像 1 中包括人和衣服，其中目标对象是衣服，文本可以是“短款白色羽绒服 2020 年新款亮面洋气轻薄高端爆款”。再比如，
15 训练样本 2 中包括图像 2 和文本 2，其中图像 2 中包括筷子、叉子、碗和盘子，其中目标对象包括筷子、叉子、碗和盘子，文本可以是“轻奢金边骨瓷碗碟套装家用景德镇碗筷组合餐具乔迁之礼”。

本申请提供的方案可以通过多种途径获取多个训练样本。比如，可以通过爬虫工具爬取包括购物信息的互联网网页，以多个训练样本。再比如，可以通过已有的公开的数据集
20 获取多个训练样本。

302、将多个训练样本输入目标模型中，以使目标模型执行步骤 3021 至步骤 3023。

3021、提取第一图像的图像特征和第一文本的文本特征。

第一图像是多个训练样本中任意一个训练样本中包括的图像，第一文本用于描述第一图像中的目标对象。本领域的技术人员可以根据需求，选择不同的图像特征提取模型提取
25 各个训练样本中的图像的图像特征，比如可以采用 Resnet50 或者 SEResnet50 提取图像特征。本领域的技术人员可以根据需求，选择不同的文本特征提取模型提取各个训练样本中的文本的文本特征，比如可以采用 GPT 模型提取文本特征。在一个优选的实施方式中，可以采用 Efficient-NetB3 模型提取各个训练样本中的图像的图像特征。在一个优选的实施方式中，可以采用 Bert 模型提取各个训练样本中的文本的文本特征。

30 3022、根据第一向量和第二向量之间的差异获取损失值 1。

将图像特征和文本特征转换为相同维度的向量。比如将第一图像的图像特征通过第一向量进行表示，将第一文本的文本特征通过第二向量进行表示。即第一向量用于指示第一图像的图像特征，第二向量用于指示第一文本的文本特征，第一向量的维度和第二向量的
35 维度相同。根据第一向量和第二向量之间的差异获取损失值 1。示例性的，下面公式 1-1 给出一种损失值 1 的计算公式：

-10-

$$J_1 = \frac{1}{n^2} \sum_{i,j=1}^n (\log(1 + e^{\alpha_{i,j}}) - l\{u_i, v_j\} * \alpha_{i,j}) \quad (1-1)$$

其中, n 表示训练样本的总数目, u_i 表示第 i 个样本中的图像特征对应的向量, v_j 表示第 j 个样本中的文本特征对应的向量, $\alpha_{i,j}$ 表示第 i 个样本中的图像特征对应的向量和第 j 个样本中的文本特征对应的向量之间的距离。

5 3023、根据损失值 1 更新目标模型。

本申请提供的方案通过在同构空间内减少图像特征与文本特征的差异性。由于模型的训练过程是不断的降低损失值的过程, 当通过损失值 1 更新目标模型时, 损失值 1 不断减小, 则同一个产品的图像特征对应的向量和文本特征对应的向量之间的距离也会更靠近。相比于只通过图像特征对目标模型进行训练, 本申请提供的方案在目标模型的训练过程中, 10 融合了文本特征。通过本申请提供的方案训练后的目标模型提取的图像特征对图像的表征能力更好。

为了更进一步的提升目标模型的性能, 进一步提升目标模型提取的图像特征对图像的表征能力, 还可以使同款产品的图像特征更加靠近, 不同款产品的图像特征更远离, 同款产品的文本特征更加靠近, 不同款产品的文本特征更加远离。因此, 在一个可能的实施方式中, 本申请提供的方案还可以引入损失值 2 和损失值 3。其中, 损失值 2 用于表示同款产品的图像特征之间的差异, 损失值 3 用于表示同款产品的文本特征之间的差异。比如, 15 根据第一图像的图像特征和第二图像的图像特征之间的差异获取损失值 2, 第二图像和第一图像从不同角度展示相同的目标对象, 即第二图像和第一图像中包括的目标对象是同款。其中不同角度可以理解为不同视角、不同属性等等。比如第一图像和第二图像中的目标对象是同款体重秤, 第一图像从俯视角度展示了该款体重秤, 第二图像从侧视角度展示了该款体重秤; 假设该款体重秤包括银色和金色, 第一图像展示了银色的该款体重秤, 第二图像展示了金色的该款体重秤。示例性的, 损失 2 可以参照公式 1-2 进行理解, 损失值 3 可以参照公式 1-3 进行理解。

$$J_2 = \frac{1}{n^2} \sum_{i,j=1}^n (\log(1 + e^{\beta_{i,j}}) - l\{u_i, u_j\} * \beta_{i,j}) \quad (1-2)$$

25 其中, n 表示训练样本的总数目, u_i 表示第 i 个样本中的图像特征对应的向量, u_j 表示第 i 个样本中的图像特征对应的向量, $\beta_{i,j}$ 表示第 i 个样本中的图像特征对应的向量和第 j 个样本中的图像特征对应的向量之间的距离。

$$J_3 = \frac{1}{n^2} \sum_{i,j=1}^n (\log(1 + e^{\gamma_{ij}}) - 1\{v_i, v_j\} * \gamma_{ij}) \quad (1-3)$$

其中，n 表示训练样本的总数目， v_i 表示第 i 个样本中的文本特征对应的向量， v_j 表示第 j 个样本中的文本特征对应的向量， γ_{ij} 表示第 i 个样本中的文本特征对应的向量和第 j 个样本中的文本特征对应的向量之间的距离。

5 因此，在图 3 对应的实施例的基础上，在一个可能的实施方式中，可以根据损失值 1 和损失值 2 更新目标模型，则总的损失是 J1 和 J2 之和；在一个可能的实施方式中，可以根据损失值 1、损失值 2 以及损失值 3 更新目标模型，则总的损失是 J1、J2 以及 J3 之和。

为了更进一步的提升目标模型的性能，进一步提升目标模型提取的图像特征对图像
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000

其中，分类结果是指预测的产品属于对象集合中各个目标对象的概率。其中，对象集合是指全部训练样本中全部图像中包括的目标对象。在一个可能的实施方式中，根据第一图像的图像特征，预测第一图像中的目标对象属于对象集合中各个目标对象的第一概率，对象集合包括多个训练样本中全部图像中包括的目标对象。根据第一文本的文本特征预测第一文本用于描述对象集合中各个目标对象的第二概率。根据第一概率和第二概率之间的差异获取损失值 4。举例说明，假设训练样本包括训练样本 1，训练样本 2 以及训练样本 3，其中，训练样本 1 包括图像 1 和文本 1，训练样本 2 包括图像 2 和文本 2，训练样本 3 包括图像 3 和文本 3。其中，图像 1 包括目标对象电饭煲，文本 1 为“B 品牌电饭煲家用 4L 智能大容量多功能饭锅 3-4 人”；图像 2 包括目标对象高压锅，文本 2 为“A 品牌电压力锅家用智能 5L 电高压锅多功能全自动”；图像 3 包括目标对象加热饭盒，文本 3 为“C 品牌加热电饭盒便当盒”。在这个例子中，对象集合包括全部图像中包括的目标对象，则对象集合包括电饭煲、电压力锅以及加热饭盒。假设当前目标模型提取训练样本 1 中包括的图像的图像特征和文本的文本特征，即提取图像 1 的图像特征和文本 1 的文本特征。假设根据图像 1 的图像特征预测图像 1 中的目标对象属于电饭煲的概率是 0.6，属于电压力锅的概率是 0.3，属于加热饭盒的概率是 0.1；假设根据文本 1 的文本特征预测文本 1 描述的目标对象属于电饭煲的概率是 0.9，描述的目标对象属于电压力锅的概率是 0.1，描述的目标对象属于加热饭盒的概率是 0。则使根据图像特征预测的分类结果和根据文本特征预测的分类结果更接近，同时也是根据文本特征预测的分类结果和根据图像特征预测的分类结果更接近。以使根据图像特征预测的分类结果和根据文本特征预测的分类结果更接近为例进行说明，是使根据图像 1 的图像特征预测图像 1 中的目标对象属于电饭煲的概率，接近根据文本 1 的文本特征预测文本 1 描述的目标对象属于电饭煲的概率，根据图像 1 的图像特征预测图像 1 中的目标对象属于电压力锅的概率，接近根据文本 1 的文本特征预测文本 1 描述的目标对象属于电压力锅的概率，根据图像 1 的图像特征预测图像 1 中的目标对象属于加热饭盒的概率，接近根据文本 1 的文本特征预测文本 1 描述的目标对象属于加热饭盒的

概率。通过损失值 4 对目标模型进行更新后，下一次根据图像特征预测的分类结果和根据文本特征预测的分类结果可能更接近。比如，通过多轮迭代训练之后，当前目标模型提取第 100 个样本中包括的图像（为了区分，以下称为图像 100）的图像特征（为了区分，以下称为图像特征 100）和文本的文本特征（为了区分，以下称为文本特征 100），由于已经经过了多轮迭代训练，根据图像特征预测的分类结果和根据文本特征预测的分类结果可能更接近，则这一次的预测结果可能是根据图像特征 100 预测图像 100 中的目标对象属于电饭煲的概率是 0.9，属于电压力锅的概率是 0，属于加热饭盒的概率是 0；根据文本特征 100 预测图像 100 中描述的目标对象属于电饭煲的概率是 0.9，属于电压力锅的概率是 0，属于加热饭盒的概率是 0，在这个例子中，通过损失值 4 更新目标模型，进行多次迭代训练，根据图像特征预测的分类结果和根据文本特征预测的分类结果已经相同。

在一个优选的实施方式中，本申请提供的方案可以使用双向的 KL 散度使根据图像特征预测的分类结果和根据文本特征预测的分类结果更接近，示例性的，可以参照公式 1-4 进行理损失值 4。

$$J_4 = \frac{1}{2} [KL(pred_{img}, pred_{text}) + KL(pred_{text}, pred_{img})] \quad (1-4)$$

其中， $pred_{img}$ 表示根据图像特征预测的产品属于对象集合中一个目标对象的概率， p_{text} 表示根据文本特征预测的的产品属于对象集合中该一个目标对象的概率。

因此，在图 3 对应的实施例的基础上，在一个可能的实施方式中，可以根据损失值 1、损失值 4 更新目标模型，则总的损失是 J_1 和 J_4 之和。需要说明的是，本申请提供的方案中任意两个可能的实施方式都可以进行结合，以下对此不再重复赘述。比如，在一个可能的实施方式中，可以根据损失值 1、损失值 2 以及损失 4 更新目标模型，则总的损失是 J_1 、 J_2 以及 J_4 之和，在一个可能的实施方式中，可以根据损失值 1、损失值 2、损失值 3 以及损失 4 更新目标模型，则总的损失是 J_1 、 J_2 、损失值 3 以及 J_4 之和。

为了更进一步的提升目标模型的性能，进一步提升目标模型提取的图像特征对图像的表征能力，还可以使根据图像特征预测的分类结果与图像中真实包括的目标对象的分类更接近，还可以使根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类更接近。在一个可能的实施方式中，根据第一图像的图像特征，预测第一图像中的目标对象属于对象集合中各个目标对象的第一概率，对象集合包括多个训练样本中全部图像中包括的目标对象。根据第一文本的文本特征预测第一文本用于描述对象集合中各个目标对象的第二概率。根据第一概率和第一预设值之间的差异，获取损失值 5，第一预设值指示预设的第一图像中的目标对象属于对象集合中各个目标对象的概率。根据第二概率和第二预设值之间的差异，获取损失值 6，第二预设值指示预设的第一文本用于描述对象集合中各个目标对象的概率。举例说明，假设训练样本包括训练样本 1，训练样本 2 以及训练样本 3，其中，训练样本 1 包括图像 1 和文本 1，训练样本 2 包括图像 2 和文本 2，训练样本 3 包括图像 3 和文本 3。其中，图像 1 包括目标对象 A 品牌第一款电饭煲，文本 1 为“A 品牌电饭煲家用智能迷你 1.6L 压力电饭锅单人小型宿舍”；图像 2 包括目标对象 A 品牌的

第二款电饭煲，文本 2 为“A 品牌电饭煲 4L 家用智能多功能饭煲 5 人”；图像 3 包括目标对象 C 品牌的一款电饭煲，文本 3 为“B 品牌小饭煲 1.6L 电饭煲智能迷你电饭锅”。则对象集合包括 A 品牌第一款电饭煲、A 品牌的第二款电饭煲以及 C 品牌的一款电饭煲。假设当前目标模型提取训练样本 1 中包括的图像的图像特征和文本的文本特征，即提取图像 1 的图像特征和文本 1 的文本特征。假设根据图像 1 的图像特征预测图像 1 中的目标对象属于 A 品牌第一款电饭煲的概率是 0.6，属于 A 品牌的第二款电饭煲的概率是 0，属于 C 品牌的一款电饭煲的概率是 0.4；假设根据文本 1 的文本特征预测文本 1 中的目标对象属于 A 品牌第一款电饭煲的概率是 0.7，属于 A 品牌的第二款电饭煲的概率是 0，属于 C 品牌的一款电饭煲的概率是 0.3；由于图像 1 中真实包括的目标对象是 A 品牌的第一款电饭煲，则对于样本 1，则预设的图像 1 中的目标对象属于 A 品牌的第一款电饭煲的概率为 1，属于 A 品牌的第二款电饭煲的概率为 0，属于 C 品牌的一款电饭煲的概率为 0。在这次训练中根据 0.6 和 1 之间的差异，0 和 0 之间的差异，以及 0.4 和 0 之间的差异获取损失值 5，在这次训练中根据 0.7 和 1 之间的差异，0 和 0 之间的差异，以及 0.3 和 0 之间的差异获取损失值 6。通过损失值 5 和损失值 6 更新目标模型的目的在于，可以使经过多轮迭代训练之后，根据图像特征预测的分类结果与图像中真实包括的目标对象的分类更接近，根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类更接近。比如，通过多轮迭代训练之后，当前目标模型提取第 100 个样本中包括的图像（为了区分，以下称为图像 100）的图像特征（为了区分，以下称为图像特征 100）和文本的文本特征（为了区分，以下称为文本特征 100）。假设图像 100 中真实包括的目标对象是 A 品牌的第一款电饭煲，则对于第 100 个样本，预设的图像 100 中的目标对象属于 A 品牌的第一款电饭煲的概率为 1，属于 A 品牌的第二款电饭煲的概率为 0，属于 C 品牌的一款电饭煲的概率为 0。由于已经经过了多轮迭代训练，根据图像特征预测的分类结果与图像中真实包括的目标对象的分类更接近，根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类更接近。则这一次的预测结果可能是根据图像特征 100 预测图像 100 中的目标对象属于 A 品牌的第一款电饭煲的概率为 0.9，属于 A 品牌的第二款电饭煲的概率为 0，属于 C 品牌的一款电饭煲的概率为 0.1。根据文本特征 100 预测描述的图像 100 中的目标对象属于 A 品牌的第一款电饭煲的概率为 0.9，属于 A 品牌的第二款电饭煲的概率为 0，属于 C 品牌的一款电饭煲的概率为 0.1。在这个例子中，通过损失值 5 和损失值 6 更新目标模型，进行多次迭代训练，根据图像特征预测的分类结果与图像中真实包括的目标对象的分类已经很接近，根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类也已经很接近。其中损失值 5 和损失值 6 可以是交叉熵损失值或者其他类型的损失值。

因此，在一个可能的实施方式中，可以根据损失值 1、损失 5 和损失值 6 更新目标模型，或者根据损失值 1 至损失值 4 中的多个损失值、损失值 5 以及损失值 6 更新目标模型，比如根据损失值 1 只损失值 6 更新目标模型。

为了更进一步的提升目标模型的性能，进一步提升目标模型提取的图像特征对图像的代表能力，还可以引入损失值 7，使不同产品的图像特征具有足够的区分性。在一个可能的实施方式中，获取第一向量和第三向量之间的第一距离，获取第一向量和第四向量之间

的第二距离，第三向量用于指示第二图像的图像特征，第四向量用于指示第三图像的图像特征，第三图像和第一图像中包括的目标对象不同。本申请有时也将获取两个向量之间的距离称为获取两个图像特征之间的距离，二者表示相同的意思，此外可以将图像特征看做一个向量。根据第一距离和第二距离之间的差值与预设阈值的差异获取损失值 7。举例说明，假设同款商品两个不同图像 a 与 p 的特征距离（也是向量的距离）为 d1，不同款商品的图像 a 与 n 的特征距离（也是向量的距离）为 d2，根据损失值 7 更新模型的目的是使得 $d2-d1 > m$ ，m 是预设阈值，即同款商品的图像特征距离 d1 会更小，不同款商品的图像特征距离 d2 会更大。示例性的，可以参照公式 1-5 进行理解损失值 7₅：

$$J_5 = \sum_i^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + m] \quad (1-5)$$

10 其中， $f(x_i^a)$ 是指当前待处理的样本中图像的图像特征， $f(x_i^p)$ 表示与当前待处理样本中图像包括同款产品的图像的图像特征， $f(x_i^n)$ 表示与当前待处理样本中图像包括不同款产品的图像的图像特征。m 表示预设阈值。

15 在一个可能的实施方式中，可以根据损失值 1 和损失值 7 更新目标模型，或者根据损失值 1 至损失 6 中多个损失值、损失值 7 更新目标模型，比根据损失值 1 至损失值 7 更新目标模型。

20 为了加深利用文本特征对目标模型提取图像特征的指导，在预测图像中包括的目标对象的多个属性时，利用文本中的关键词作为预设标签进行多分类监督。在一个可能的实施方式中，根据第一图像的图像特征，预测第一图像中的目标对象的至少一个属性。根据至少一个属性和预设标签之间的差异获取损失值 8，预设标签是对第一文本进行分词预处理后获取的。在一个可能的实施方式中，分词预处理包括分词操作和词频分析操作，具体的，可以预先对各个训练样本中包括的各个文本进行分词处理，以获取多个备选关键词，再对多个备选关键词进行词频分析操作，获取其中最能表示目标对象属性的备选关键词作为最终的关键词，利用最终的关键词作为预设标签。举例说明，假设对文本“A 品牌电饭煲家用智能迷你 1.6L 压力电饭锅单人小型宿舍”进行分词处理后，获取备选关键词“A 品牌”、“电饭煲”、“家用”、“智能”、“迷你”、“1.6L”、“压力”、“电饭锅”、“单人”、“小型”、“宿舍”。通过对这些备选关键词进行词频分析操作，可能从备选关键词中筛选出最终关键词，假设最终关键词包括“A 品牌”、“电饭煲”、“迷你”、“小型”、“1.6L”、“电饭锅”。将这些最终关键词作为预设标签，根据图像特征对该文本描述的目标对象进行属性预测时，预测结果（至少一个属性）应当尽量预测接近预设标签。

30 根据至少一个属性和预设标签之间的差异获取损失值 8，根据损失值 8 对目标模型进行多伦迭代训练后，可以使目标模型根据图像特征对文本描述的目标对象进行属性预测时，预测结果（至少一个属性）更接近预设标签。

35 在一个可能的实施方式中，可以根据损失值 1 和损失值 8 对目标模型进行更新。在一个可能的实施方式中，可以根据损失值 1、损失值 2 和损失值 8 对目标模型进行更新。在一个可能的实施方式中，可以根据损失值 1、损失值 2、损失值 3 和损失值 8 对目标模型进行更新。在一个可能的实施方式中。在一个可能的实施方式中，可以根据损失值 1、损失

值 2、损失值 3、损失值 4、以及损失值 8 对目标模型进行更新。在一个可能的实施方式中，可以根据损失值 1、损失值 2、损失值 3、损失值 4、损失值 5、损失值 6 以及损失值 8 对目标模型进行更新。在一个可能的实施方式中，可以根据损失值 1、损失值 2、损失值 3、损失值 4、损失值 5、损失值 6、损失值 7 以及损失值 8 对目标模型进行更新。

5 为了更好的理解本申请提供的方案，下面结合一个实施例对基于图 3 对应的实施例进行优化的一个实施例进行梳理。如图 4 所示，为申请实施例提供的一种目标模型的执行流程图示意图。该目标模块可以包括两个特征提取模块，分别用于提取训练样本中图像的图像特征以及训练样本中文本的文本特征。在同构空间内减少图像特征与文本特征的差异性，参照损失值 1 进行理解。使同款产品的图像特征更加靠近，同款产品的文本特征更加靠近，
10 参照损失 2 和损失 3 进行理解。使根据图像特征预测的分类结果和根据文本特征预测的分类结果更接近，参照损失值 4 进行理解。使根据图像特征预测的分类结果与图像中真实包括的目标对象的分类更接近，使根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类更接近，参照损失值 5 和损失值 6 进行理解。使不同产品的图像特征具有足够的区分性，参照损失值 7 进行理解。在预测图像中包括的目标对象的多个属性时，
15 利用文本中的关键词作为预设标签进行多分类监督根据至少一个属性和预设标签之间的差异获取损失值 8。根据上述损失值 1 至损失值 8 的总和获取总的损失值，根据总的损失值更新目标模型。

由图 3 对应的实施例，以及基于图 3 对应的实施例进行优化的其他实施例可知，本申请提供的方案利用文本特征指导图像特征的学习，提升通过目标模型提取的图像特征对图像
20 的表征能力。此外，还可以通过对文本进行预处理获取关键词，在预测图像中包括的目标对象的多个属性时，利用文本中的关键词作为预设标签进行多分类监督，进一步提升图像特征的表征能力。通过提升图像特征的表征能力，使图像特征能够更准确的表示图像的特征，有利于提升根据图像特征进行图像检索的精度。

图 3 对应的实施例中介绍到本申请提供的方案可以通过多种途径获取多个训练样本。
25 其中，通过爬虫工具爬取包括购物信息的互联网网页，获取多个训练样本这种方案可能存在噪声大的问题，还可能导致训练样本集合不均衡的问题。这些已经在研究思路中进行了介绍，这里不再重复赘述。下面结合一个具体的实施例，对如何解决这些问题进行说明。

参阅图 5，本申请实施例可以通过如下方法获取均衡的训练样本集合，具体的，如图 5 所示，可以包括以下步骤：

30 501、获取产品的第一目标图像、产品的文本标题、产品的第二目标图像。

第一目标图像来自于产品的提供者，第二目标图像来自产品的购买者。换句话说，可以将第一目标图像理解为产品的买家秀，将第二目标图像理解为卖家秀，产品的文本标题可以理解卖家设置的产品的标题，比如“A 品牌电饭煲家用智能迷你 1.6L 压力电饭锅单人
小型宿舍”。

35 502、将第一目标图像作为分割模型的输入，以获取分割结果。

通过对图像中每一个像素点进行前背景分类，确定每个点的类别，从而进行前景主体区域划分。

本领域的技术人员可以根据需求选择不同的分割模型对第一目标图像进行分割处理，以获取分割结果。在一个优选的实施方式中，该分割模型是 DeepLab V3。分割结果可以指示产品在第一目标图像中的轮廓。需要说明的是，由于卖家秀一般具有简单的背景，所有采用通用的分割模型对卖家秀进行分割处理，就可以获取较好的分割效果。换句话说，本申请中使用的分割模型不需要以大量真实的卖家秀作为训练数据，对分割模型进行训练，本申请中使用的分割模型可以通过其他类型的训练数据进行训练后得到的分割模型。

503、根据分割结果提取第三目标图像的目标区域，第三目标图像是根据第一目标图像生成的。

根据第一目标图像可以生成第三目标图像。可以将第一目标图像看做源域数据中的一个数据，将第三目标图像看做目标域数据中的一个数据，本领域的技术人员可以根据需求采用不同的方式根据源域数据生成目标域数据。比如，在一个可能的实施方式中，参阅图 6，通过来自于产品的提供者提供的图像（源域数据）和来自产品的购买者提供的图像（目标域数据）训练循环生成式对抗网络（cycle-generative adversarial network, cycle-GAN）。cycle-GAN 由 2 个生成器 (G, F) 和 2 个判别器 (DX, DY) 组成。其中，生成器 G 根据源域数据生成目标域数据，判别器 DX 要判别输入的是数据是真实的目标域数据，还是根据生成的目标域数据。生成器 F 根据生成的目标域数据生成源域数据，判别器 DY 要判别输入的数据是真实的源域数据，还是生成的源域数据。通过对抗训练，生成器 G 要尽可能欺骗判别器 DX，而生成器 F 要尽可能欺骗判别器 DY，当判别器 DX、DY 不在有能力分出真假时，可以认为训练完成，此时，将真实的源域数据输入生成器 G，生成器 G 可以根据输入的真实源域数据生成目标域数据。在本申请提供的方案，将第一目标图像输入至生成器 G，生成器 G 可以根据第一目标图像生成第三目标图像。继续参阅图 7，可以根据目标对象的轮廓从第三目标图像中提取第三目标图像中的目标对象所在的区域。

504、对目标区域进行随机仿射变换，以获取变换后的目标区域。

为了增加生成的第三目标图像的多样性，可以对目标区域进行随机仿射变换。通过对目标区域进行随机仿射变换，也可以获取更接近真实情况的买家秀。比如不同用户拍摄同款产品的角度可能不同，拍摄环境可能不同，对目标区域进行随机仿射变换后，可以模拟买家秀拍摄角度的差异。其中，对目标区域进行随机仿射变换可以理解为对目标区域进行旋转、平移等操作。

505、将变换后的目标区域与预设背景进行组合，以获取第四目标图像。

其中，预设背景可以从其他买家秀中提取的背景区域或者是随机收集的背景区域。通过将预设背景区域和变换后的目标区域进行组合，以获取第四目标图像。

在一个可能的实施方式中，获取第一类型的训练样本、第二类型的训练样本以及第三类型的训练样本，每个第一类型的训练样本包括第一目标图像和文本标题，每个第二类型的训练样本第二目标图像和文本标题，每个第三类型的训练样本包括第四目标图像和文本标题。

在一个可能的实施方式中，通过图 5 对应的实施例获取的训练样本也可以只包括图像不包括文本。即每个第一类型的训练样本包括产品的购买者提供的图像，产品的提供者提

供的图像，以及根据图 5 所描述的方案根据产品的提供者提供的图像生成的图像。将这些图像数据作为训练数据，去训练已有的图像特征提取模型，以执行后续的图像检索任务，也可以提升图像的精度，将在下文通过实验数据进行说明。

通过图 5 对应的实施例的方案，在买家秀缺失的情况下，通过卖家秀生成大量的买家秀数据，使得训练后的目标模型对于复杂的买家秀图像风格具有一定的鲁棒性，并且在一定程度上能够缓解用户拍摄视角大角度变化造成的检索难度。

需要说明的是，图 5 对应的实施例可以作为一个单独的实施例，也可以和图 3 对应的实施例进行结合。比如通过图 5 对应的实施例中获取的第一类型的训练样本、第二类型的训练样本以及第三类型的训练样本作为图 3 对应的实施例中的多个训练样本。

此外，为了在提升图像检索精度的同时，还可以提升图像检索的速度，降低对图像数据库的存储能力的要求，本申请提供的方案提供了一种端到端的设计，对哈希特征和倒排码本进行联合的优化，以下结合一个具体的实施例对此进行说明。

参阅图 7，为本申请实施例提供的一种训练模型的方法的流程示意图。

如图 7 所示，本申请实施例提供的一种训练模型的方法，可以包括以下步骤：

701、根据目标模型提取的多个图像特征，对哈希模型进行多次训练，以获取训练后的哈希模型，其中，多次训练中的任意一次训练，包括步骤 7011 至步骤 7015。

其中，目标模型可以参照图 3 对应的实施例中描述的目标模型进行理解，这里不再重复赘述。

在一些可能的实施方式中，目标模型可能是已经训练好的目标模型，比如已经完成预设次数的训练，或者目标模型已经收敛。在一些可能的实施方式中，目标模型也可能是还没有训练好的目标模型。

7011、获取每个图像特征对应的哈希特征。

将目标模型提取的图像特征作为哈希模型的输入，以使哈希模型输出图像特征对应的哈希特征。

7012、对多个哈希特征进行聚类处理，以获取多个聚类数据集。

比如根据多个哈希特征的大小将多个哈希特征划分为不同的组，使每一组包括的哈希特征的大小是相近的。多个聚类数据集中的每一个聚类数据集都对应有一个聚类中心，可以将聚类中心理解为一个聚类数据集中包括的全部哈希特征的平均值。全部聚类数据集的全部聚类中心的集合称为倒排码本。

7013、从多个哈希特征中获取第一哈希特征，根据各个聚类数据集的聚类中心和第一哈希特征的相似性，从多个聚类数据集中获取第一聚类数据集。

第一哈希特征是多个哈希特征中的任意一个。多个聚类数据中第一聚类数据集的聚类中心和第一哈希特征的相似性最高。

7014、根据第一聚类数据集的聚类中心和第一哈希特征的差异获取损失值 9。

7015、根据损失值 9 更新哈希模型。

为了同时优化倒排码本和哈希模型，本申请引入损失值 9，对哈希模型进行迭代训练。假设当前是对哈希模型进行第 r 次迭代训练，基于第 $r-1$ 轮迭代获取的哈希模型对图像特

征进行哈希特征抽取，采用聚类的方式获取多个聚类数据集，根据该多个聚类数据集的聚类中心的集合获得倒排码本 C。计算各个聚类数据集的聚类中心，通过第 i 个样本的哈希特征和第一聚类数据集的聚类中心的差异获取损失值，通过该损失值更新哈希模型，以获取经过第 r 次迭代训练后获取的哈希模型。示例性的，可以参照公式 1-6 理解损失值 $9L_C$ 。

$$5 \quad L_C = \|u_i - c_{i^*}\|_2 \quad (1-6)$$

其中， u_i 表示第 i 个样本的哈希特征， c_{i^*} 表示与 u_i 最接近的聚类中心。

当哈希模型收敛了，或者训练次数到达预设数目时，可获得优化后的哈希特征模型和倒排码本。本申请提供的方案将二者联合优化，端到端训练倒排码本和哈希特征。通过本申请提供的方案，在执行图像检索的任务时，获取图像的哈希特征，根据该图像的哈希特征从倒排码本中查找和该哈希特征相似度最高的聚类中心，获取该相似度最高的一个或者多个聚类中心对应的聚类数据集，提升了从该聚类数据集中包括的哈希特征中查找输入图像中包括的同款产品的概率，进而提升了图像检索的精度。

此外，步骤 7011 中提到将目标模型提取的图像特征作为哈希模型的输入，以使哈希模型输出图像特征对应的哈希特征。在一个可能的实施方式中，通过对哈希模型的迭代训练，以使图像特征通过二值化的向量（假设二值化的向量包括第三预设值和第四预设值）进行表示，为了实现这一目的，本申请引入损失值 10，可以参照公式 1-7 理解损失值 $10L_r$ 。图像数据库中存储的是各个图像的二值化的向量，可以在一定程度上降低对图像数据库的存储能力的要求。此外，在一个可能的实施方式中应当使该二值化的向量中第三预设值和第四预设值的数目是均衡的，以使二值化的向量可以表示更多的信息。这是因为二值化向量全部都是第三预设值或者全部都是第四预设值时，表示的信息量少。为了实现这一目的，本申请引入了第一约束，可以参照公式 1-8 理解第一约束 L_b 。此外，在一个可能的实施方式中，还应当使二值化向量中的每个维度的取值是独立的，以使二值化的向量可以携带更多的信息。为了实现这一目的，本申请引入了第二约束，可以参照公式 1-9 理解第二约束 L_i 。

结合上述分析，在一个可能的实施实施方式中，对哈希模型的任意一次训练，包括：

对目标哈希特征施加第一约束和第二约束，以获取更新后的目标哈希特征，目标哈希特征是根据上一次更新后的哈希模型获取的哈希特征，第一约束为哈希特征中的第三预设值和第四预设值的数目是平均的，第二约束为目标矩阵的第一对角线上的元素为第四预设值，目标矩阵中除第一对角线之外的其他元素为第三预设值，目标矩阵中第 i 行第 j 列的元素表示多个目标元素的平均值，多个目标元素表示各个哈希特征的第 i 个维度的取值和各个哈希特征各自的第 j 个维度的取值的乘积。i 和 j 为正整数。对目标哈希特征中每一维度的取值和第二均值进行比较，第二均值是第三预设值和第四预设值的平均值。每一维度的取值小于第二均值时，根据哈希特征的每一维度的取值和第三预设值之间的差异获取损失值 10。每一维度的取值不小于第二均值时，根据哈希特征的每一维度的取值和第四预设值之间的差异获取损失值 10。

$$L_f = - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^L [1(u_{ij} \geq 0.5)(1-u_{ij})^a \cdot \log(u_{ij}) + 1(u_{ij} < 0.5)(u_{ij})^a \cdot \log(1-u_{ij})] \quad (1-7)$$

其中, n 表示训练样本的总数目, L 表示哈希特征的长度 (比如 1024 比特), 也可以称为二值化向量的长度。 a 表示预设的参数, u_{ij} 表示第 i 样本的哈希特征的第 j 个维度。

$$L_b = \|U^T \mathbf{1}\|_1 \quad (1-8)$$

5 其中, U 表示哈希模型输出的哈希特征。

$$L_i = \|U^T U - I\|_F \quad (1-9)$$

其中, U 表示哈希模型输出的哈希特征, I 表示对角线上的元素为第四预设值, 其余元素为第三预设值的矩阵。其中, 该对角线的一端是第一行第一列的元素, 另一端是最后一行最后一列的元素。

10 上文提到目标模型可能是已经训练好的目标模型, 比如已经完成预设次数的训练, 或者目标模型已经收敛。目标模型也可能是还没有训练好的目标模型。在一种可能的实施方式中, 如果目标模型是已经训练好的目标模型, 则根据损失值 9 和损失值 10 更新哈希模型时, 不再更新目标模型。在一种可能的实施方式中, 如果目标模型还不是已经训练好的目标模型, 则根据损失值 9 和损失值 10 更新哈希模型时, 还需要根据损失值 9 和损失值 10
15 更新目标模型。比如根据损失值 1 至损失值 10 更新目标模型。此外, 通过损失值 9 和损失值 10 更新哈希模型时, 还可以通过损失值 1 至损失值 8 中的一个或者多个损失值一起更新哈希模型。

需要说明的是, 图 7 对应的实施例可以和其他实施例结合, 也可以不和其他实施例进行结合, 作为一个单独的实施例。下面结合图 8 进行说明。参阅图 8, 为本申请实施例提供的一种训练模型的方法的流程示意图。在这种实施方式中, 图 7 对应的实施例可以不和
20 图 3 对应的实施例相结合, 不利用文本特征指导图像特征的学习。在这种实施方式中, 目标模型的输入只包括图像, 目标模型仅提取图像的图像特征。目标模型的输出的图像特征作为哈希模型的输入, 以使哈希模型将图像特征转换为预设维度的二值化向量。其中, 根据图像特征对哈希模型进行训练的过程中, 还对倒排码本进行了联合的优化。关于如何同时优化倒排码本和哈希模型, 已经在上文图 7 对应的实施例进行了介绍, 这里不再重复说明, 也可以结合图 8 进行理解。此外, 在对哈希模型进行训练的过程中, 还考虑了损失值
25 10、第一约束、第二约束, 这里不再重复说明。在这种实施方式中, 为了提升哈希模型提取的图像的哈希特征对图像的代表能力, 还可以使不同产品的哈希特征具有足够的区分性, 这里可以参照损失值 7 进行理解, 即参照使不同的图像特征具有足够的区分性进行理解,
30 将损失值 7 的相关描述中的图像特征替换成哈希特征即可。此外, 为了进一步提升目标模型提取的哈希特征对图像的代表能力, 还可以使根据哈希特征预测的分类结果与图像中真实包括的目标对象的分类更接近, 还可以使根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类更接近。这里可以参照损失值 5 和损失值 6 进行理解, 此处不再重复赘述, 即将损失值 5 和损失值 6 中的相关描述中的图像特征替换成哈希特征即

可。

此外，各个实施例之间的组合方式可能有多种，下面结合图 9 进行说明。参阅图 9，为本申请实施例提供的一种训练模型的方法的流程示意图。在这种实施方式中，图 6 对应的实施方式可以和图 3 对应的的实施例进行结合，以及和图 5 对应的实施例进行结合。如图 9 所示，可以将根据图 5 所描述的方法获取的第一类型的训练样本、第二类型的训练样本以及第三类型的训练样本作为目标模型的输入。在这种实施方式中，目标模型和哈希模型可以同步训练，在一次迭代训练的过程中，目标模型可以提取训练样本的文本特征和图像特征，并获取损失值 1、损失值 2、损失值 3 以及损失值 4；目标模型提取的图像特征作为哈希模型的输入，以使哈希模型将图像特征转换为预设维度的二值化向量（或者说哈希特征）。此外，在对哈希模型进行训练的过程中，还考虑了损失值 10、第一约束、第二约束。在这种实施方式中，通过哈希特征获取分类损失和三元组损失，其中分类损失为了使不同产品的哈希特征具有足够的区分性，这里可以参照损失值 7 进行理解，即参照使不同的图像特征具有足够的区分性进行理解，将损失值 7 的相关描述中的图像特征替换成哈希特征即可。三元组损失为了使根据哈希特征预测的分类结果与图像中真实包括的目标对象的分类更接近，还可以使根据文本特征预测的分类结果描述的对象与图像中真实包括的目标对象的分类更接近。这里可以参照损失值 5 和损失值 6 进行理解，此处不再重复赘述，即将损失值 5 和损失值 6 中的相关描述中的图像特征替换成哈希特征即可。还可以利用文本中的关键词作为预设标签进行多分类监督。在一个可能的实施方式中，根据第一图像的哈希特征，预测第一图像中的目标对象的至少一个属性。根据至少一个属性和预设标签之间的差异获取损失值，可以参照损失值 8 进行理解，将图像特征替换为哈希特征即可。

以上对如何训练模型进行了说明，下面对如何应用训练后的模型执行图像检索任务进行说明。

（二）通过训练好的模型执行图像检索任务

参阅图 10，为本申请实施例提供的一种图像检索方法的流程示意图。

如图 10 所示，本申请实施例提供的一种图像检索方法，可以包括以下步骤：

1001、获取待检索图像。

本申请提供的方案可以适用于通过拍照进行产品检索的场景。具体的，用户通过手机等终端设备对该产品进行拍照，进行图像检索。本申请提供的方案可以通过终端设备获取待检索的图像。

1002、通过目标模型提取待检索图像的图像特征。

通过图 3 所描述的模型训练方法训练得到的目标模型提取待检索图像的图像特征。其中，目标模型的训练方法可以参照图 3 对应的实施例进行理解，这里不再重复赘述。

1003、根据待检索图像的图像特征从图像数据库中获取预测图像。

预测图像的图像特征和待检索图像的图像特征的相似度达到阈值。即检索的目的要获取与待检索图像中包括的产品的同款产品或者相似产品，通过本申请实施例提供的方案，可以使获取到的产品更多的是同款产品。图像数据库中存储了大量的图像特征或者哈希特征。图像数据库中存储的图像特征或者哈希特征来自产品的提供者提供的图像。具体的，

将产品的提供者提供的图像输入至目标模型中，以获取各个产品的提供者提供的图像的图像特征，图像数据库中保存各个产品的提供者提供的图像的图像特征。目标模型输出的图像特征可以作为哈希模型的输入，以输出各个图像特征对应的哈希特征，则可以获取各个产品的提供者提供的图像的哈希特征，图像数据库中包括各个产品的提供者提供的图像的哈希特征。其中，哈希模型的训练过程已经上文进行了介绍，这里不再重复说明。当图像数据库中保存的是图像的图像特征时，根据目标模型提取的待检索图像的图像特征从图像数据库中获取预测图像。当图像数据库中保存的是图像的哈希特征时，还需要将目标模型输出的图像特征输入至哈希模型中，以获取待检索图像的哈希特征，根据待检索图像的哈希特征从图像数据库中获取预测图像。

通过本申请实施例提供的图像检索方法，可以有效提升图像检索的精度、图像检索的速度。

通过本申请提供的方案执行图像检索任务，可以显著提升图像检索的精度、图像检索的速度，在降低对图像数据库的存储能力的要求的基础上，不降低图像检索的精度。下面结合几组实验数据对本申请提供的方案的性能进行展示。参照表 1，实验过程，针对饰品和玩具两个类型的产品进行测试。其中，产品数目代表测试数据集中包括的产品的总数目。查询数量表示测试数据集中来自产品的购买者提供的图像的总数目，底裤数量表示测试数据集中来自产品的提供者提供的图像的总数目。

表 1:

测试集	产品数量	查询数量	底库数量
饰品-测试数据集 1	203	691	754
饰品-测试数据集 2	800	4179	5634
玩具-测试数据集 3	206	1110	1462
玩具-测试数据集 4	541	2784	3813

表 2 为采用了本申请图 5 对应的方案获取的训练样本集合作为训练样本（只包括图像，不包括文本的情况），对已有的图像提取模型进行训练后，执行图像检索任务后的 Top1 精度测试指标（Top1 精度测试指标是指检索出的置信度最高的产品的精度）。由于通过卖家秀生成大量的买家秀数据，使得训练后的目标模型对于复杂的买家秀图像风格具有一定的鲁棒性，并且在一定程度上能够缓解用户拍摄视角大角度变化造成的检索难度。因此相比

于传统方案，本申请提供的方案在饰品和玩具品类上的图像检索均有明显的精度提升，其中在饰品上提升显著，在饰品较大规模的测试集上有 4.58%的精度提升。Top1 指标越高，表示检索的精度越高，Top1 精度指检索结果中置信度最高的结果和查询图像是否是同一款商品。如果一共 10 张查询图像，8 张查询图像检索出的置信度最高的结果和查询图像是同款，那么 Top1 精度为 8/10=80%。

表 2

方法	饰品-测试数据集 1	饰品-测试数据集 2	玩具-测试数据集 3	玩具-测试数据集 4
传统方案	87.93	64.34	88.10	74.72
本方案	90.72(+2.79)	68.92(+4.5)	89.53 (+ 1.43)	75.21 (+ 0.49)

如图 11 所示，图 11 展示了通过传统方案进行图像检索任务后返回的同款产品的列表，以及通过本方案进行图像检索任务后返回的同款产品的示意图。本申请提供的方案列表的第一个产品（检索出的置信度最高的产品）即为输入产品的同款产品，而传统方案检索出的置信度最高的产品并不是输入产品的同款产品。

表 3 展示了通过本申请提供的方案利用文本特征指导图像特征的学习后，在饰品和玩具两个商品品类上的 Top1 精度测试指标增益；其中传统方法仅使用图像数据训练模型，而本发明方法在模型训练过程中利用文本特征指导图像特征的学习。相比传统方法，利用文本特征指导图像特征的学习后，在饰品和玩具品类上有明显的精度提升，其中在两个较大规模测试集上分别提升了 1.57%和 0.53%。在这一基础上，本申请提供的方案还在饰品上进一步测试了利用文本中的关键词作为预设标签进行多分类监督的方案，模型在饰品上的精度有了更大的提升，相比传统方法最终提升了 2.77%。

表 3

方法	饰品-测试数据集 1	饰品-测试数据集 2	玩具-测试数据集 3	玩具-测试数据集 4
传统方案	93.1	75.63	93.16	81.19
本方案 (利用文本特征指导图像特征的学习)	93.5 (+ 0.4)	77.72(+1.5)	94.12 (+ 0.96)	81.72 (+ 0.53)
本方案 (利用文本特征指导图像特征的学习，并	93.1	78.40(+2.7)	-	-

且利用文本中的关键词作为预设标签进行多分类监督)

如图 12 所示，在利用文本特征指导图像特征的学习之后，提升了模型在应对图像视觉特征造成检索干扰时的鲁棒性，对于用户多变的拍摄条件有更好的适应性。例如，用户在所要检索的玩具车上摆放了公仔，这对检索造成了非常大的干扰，如果只通过图像特征进行检索，很容易检索出同样带有公仔的玩具车；而利用文本特征指导图像特征的学习之后，

5

表 4 展示了通过本申请提供的哈希模型获取的哈希特征去执行图像检索任务，在 9 个产品类目上的测试结果；本申请提供的方法将图像特征压缩了 32 倍（将 1024 维度 32 位浮点数的图像特征压缩至 1024 维度的二值哈希特征），而检索精度下降保持在 2%以内，在保证检索精度的同时大幅度减少了图像数据库的储存空间需求。同时在亿级产品底库检索

10

表 4

产品类目	数据类型	Top1 精度测试指标	产品类目	Model	Top1 精度测试指标
玩具类	浮点型	91.36	食品类	浮点型	98.71
	深度哈希	91.88		深度哈希	98.71
饰品类	浮点型	77.32	美妆类	浮点型	93.28
	深度哈希	80.24		深度哈希	91.91
箱包类	浮点型	96.36	数码类	浮点型	69.38
	深度哈希	96.35		深度哈希	68.17
鞋子类	浮点型	76.58	家居类	浮点型	76.97
	深度哈希	78.22		深度哈希	75.86

服装类	浮点型	82.24	/	/	/
	深度哈希	81.42			

表 5 展示了本申请提供的方案在鞋子类产品的大规模测试，实现结果表明，本申请提供的方案在检索精度、检索时间、对图像数据库存储性能的限制都比传统方案具有优势。

表 5

方法	top1 精度 测试指标	(Top1+mAP @10)/2)	图像数据库 的大小	总时长/检索 次数
本方案（采用图像特征进行图像检索，未采用倒排索引的方式）	63.7	40.11	4.9G	18min/427
	64.87	40.49	157M	297s/427
本方案（采用哈希特征进行图像检索，未采用倒排索引的方式）				
传统方案	51.29	31.87	363M	143s/427
本方案（采用哈希特征进行图像检索，并采用倒排索引的方式）	63.93	39.79	167M	100s/427

mAP@10 表示通过检索结果中前十个置信度最高的结果和查询图像是否是同一款商品获取的精度。其中，倒排索引是一种利用了根据本申请提供的方案获取的倒排码本的一索引方式。

以上对本申请实施例提供的一种训练模型的方法以及图像检索的方法进行了介绍，通过本申请实施例提供的一种训练模型的方法获取的图像特征执行图像检索任务，可以有效提升图像检索的精度、图像检索的速度、并且降低对图像数据库的存储性能的要求。

可以理解的是，可以通过训练装置执行上述训练模型的方法，通过执行设备执行上述图像检索的方法。本申请有时也训练装置称为训练模型的装置，本申请有时也将执行设备称为终端设备或者移动终端或者图像检索的设备。为了实现上述功能，训练装置和执行设备包含了执行各个功能相应的硬件结构和/或软件模块。本领域技术人员应该很容易意识到，结合本文中所公开的实施例描述的各示例的模块及算法步骤，本申请能够以硬件或硬件和计算机软件的结合形式来实现。某个功能究竟以硬件还是计算机软件驱动硬件的方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

例如，可以通过图13中的训练装置来实现。图13所示为本申请实施例提供的训练装置的硬件结构示意图。包括：通信接口1301和处理器1302，还可以包括存储器1303。

通信接口1301可以使用任何收发器一类的装置，用于与其他设备或通信网络通信，如以太网，无线接入网（radio access network, RAN），无线局域网(wireless local area networks, WLAN)等。

处理器1302包括但不限于中央处理器(central processing unit, CPU)，网络处理器(network processor, NP)，专用集成电路(application-specific integrated circuit, ASIC)或者可编程逻辑器件(programmable logic device, PLD)中的一个或多个。上述PLD可以是复杂可编程逻辑器件(complex programmable logic device, CPLD)，现场可编程逻辑门阵列(field-programmable gate array, FPGA)，通用阵列逻辑(generic array logic, GAL)或其任意组合。处理器1302负责通信线路1304和通常的处理，还可以提供各种功能，包括定时，外围接口，电压调节，电源管理以及其他控制功能。存储器1303可以用于存储处理器1302在执行操作时所使用的数据。

存储器1303可以是只读存储器(read-only memory, ROM)或可存储静态信息和指令的其他类型的静态存储设备，随机存取存储器(random access memory, RAM)或者可存储信息和指令的其他类型的动态存储设备，也可以是电可擦可编程只读存储器(electrically erasable programmable read-only memory, EEPROM)、只读光盘(compact disc read-only memory, CD-ROM)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质，但不限于此。存储器可以是独立存在，通过通信线路1304与处理器1302相连接。存储器1303也可以和处理器1302集成在一起。如果存储器1303和处理器1302是相互独立的器件，存储器1303和处理器1302相连，例如存储器1303和处理器1302可以通过通信线路通信。通信接

口1301和处理器1302可以通过通信线路通信，通信接口1301也可以与处理器1302直连。

通信线路1304可以包括任意数量的互联的总线和桥，通信线路1304将包括由处理器1302代表的一个或多个处理器1302和存储器1303代表的存储器的各种电路链接在一起。通信线路1304还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起，这些都是本领域所公知的，因此，本申请不再对其进行进一步描述。

在一个具体的实施方式中，该训练装置包括存储器和处理器，处理器获取存储器中存储的指令，执行上述图3至图9中所描述的模型训练的方法。该训练装置还可以包括通信接口，以通过通信接口和其他设备进行交互，比如通过通信接口获取训练样本。

在本申请实施例中，可以将通信接口视为接收模块或者发送模块或者收发模块，将处理器视为处理模块，将存储器视为存储模块。

参见图14，是本申请实施例提供的一种执行设备的结构示意图。如图14所示，执行设备可以包括处理器1401、存储器1402、通信线路1403、通信接口1404，人工智能处理器1405通过所述通信总线连接所述存储器1402和所述通信接口1404。

处理器1401可以是中央处理单元(central processing unit, CPU)，该处理器1401还可以是其他通用处理器、数字信号处理器(digital signal processor, DSP)、专用集成电路(application specific integrated circuit, ASIC)、现成可编程门阵列(field-programmable gate array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。处理器1401可以是微处理器或者该处理器1401也可以是任何常规的处理器等。

处理器1401还可以是一种集成电路芯片，具有信号的处理能力。在实现过程中，本申请的神经网络处理方法的各个步骤可以通过处理器1401中的硬件的集成逻辑电路或者软件形式的指令完成。

存储器1402可以是只读存储器(read-only memory, ROM)、随机存取存储器(random access memory, RAM)或其他存储器。本申请实施例中，存储器1402用于存储数据以及各种软件程序，例如本申请实施例中根据确定好的目标拆分路径对神经网络模型进行拆分的程序等。

可选的，在本申请实施例中，所述存储器可以包括用于存储信息的物理装置，通常是将信息数字化后再以利用电、磁或者光学等方法的媒体加以存储。本实施方式所述的存储器又可以包括：利用电能方式存储信息的装置，如RAM、ROM等；利用磁能方式存储信息的装置，如硬盘、软盘、磁带、磁芯存储器、磁泡存储器、U盘；利用光学方式存储信息的装置，如CD或DVD。当然，还有其他方式的存储器，例如量子存储器、石墨烯存储器等等。

通信接口1404使用例如但不限于收发器一类的收发装置，来实现执行设备与其他设备或通信网络之间的通信。例如，可以通过通信接口1404接收其他设备发送的模型文件。

在一个具体的实施方式中，该执行设备包括存储器和处理器，处理器获取存储器中存储的指令，执行上述图10中所描述的图像检索的方法。该训练装置还可以包括通信接口，以通过通信接口和其他设备进行交互。

在本申请实施例中，可以将通信接口视为接收模块或者发送模块或者收发模块，将处

理器视为处理模块，将存储器视为存储模块。

应当理解，执行设备仅为本申请实施例提供的一个例子，并且，执行设备可具有比示出的部件更多或更少的部件，可以组合两个或更多个部件，或者可具有部件的不同配置实现。

5 在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。

应当理解，上述仅为本申请实施例提供的一个例子，并且，端侧设备可具有比示出的部件更多或更少的部件，可以组合两个或更多个部件，或者可具有部件的不同配置实现。

10 在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。

本申请实施例提供还提供一种芯片，该芯片包括：处理单元和通信单元，所述处理单元例如可以是处理器，所述通信单元例如可以是输入/输出接口、管脚或电路等。该处理单元可执行存储单元存储的计算机执行指令，以使芯片执行上述图 3 至图 9 中所描述的方法。或者以使芯片执行上述图 10 所描述的方法。可选地，所述存储单元为所述芯片内的存储单元，如寄存器、缓存等，所述存储单元还可以是所述无线接入设备端内的位于所述芯片外部的存储单元，如只读存储器（read-only memory, ROM）或可存储静态信息和指令的其他类型的静态存储设备，随机存取存储器（random access memory, RAM）等。具体地，前述的处理单元或者处理器可以是中央处理器（central processing unit, CPU）、网络处理器（neural-network processing unit, NPU）、图形处理器（graphics processing unit, GPU）、数字信号处理器（digital signal processor, DSP）、专用集成电路（application specific integrated circuit, ASIC）或现场可编程逻辑门阵列（field programmable gate array, FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者也可以是任何常规的处理器等。

25 另外需说明的是，以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。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。另外，本申请提供的装置实施例附图中，模块之间的连接关系表示它们之间具有通信连接，具体可以实现为一条或多条通信总线或信号线。

30 通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件的方式来实现，当然也可以通过专用硬件包括专用集成电路、专用 CPU、专用存储器、专用元器件等来实现。一般情况下，凡由计算机程序完成的功能都可以很容易地用相应的硬件来实现，而且，用来实现同一功能的具体硬件结构也可以是多种多样的，例如模拟电路、数字电路或专用电路等。但是，对本申请而言更多情况下软件程序实现是
35 更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在可读取的存储介质中，如计算机的软盘、U 盘、移动硬盘、只读存储器（read only memory, ROM）、随机存取存

存储器 (random access memory, RAM)、磁碟或者光盘等, 包括若干指令用以使得一台计算机设备 (可以是个人计算机, 服务器, 或者网络设备等等) 执行本申请各个实施例所述的方法。

在上述实施例中, 可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。

5 当使用软件实现时, 可以全部或部分地以计算机程序产品的形式实现。

本申请实施例中还提供一种计算机可读存储介质, 该计算机可读存储介质中存储有用于训练模型的程序, 当其在计算机上运行时, 使得计算机执行上述图 3 至图 9 中所描述的方法。或者使得计算机执行上述图 10 中所描述的方法。

10 本申请实施例还提供一种数字处理芯片。该数字处理芯片中集成了用于实现上述处理器, 或者处理器的功能的电路和一个或者多个接口。当该数字处理芯片中集成了存储器时, 该数字处理芯片可以完成前述实施例中的任一个或多个实施例的方法步骤。当该数字处理芯片中未集成存储器时, 可以通过通信接口与外置的存储器连接。该数字处理芯片根据外置的存储器中存储的程序代码来实现上述实施例中车辆执行的动作。

15 本申请实施例中还提供一种计算机程序产品, 所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时, 全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中, 或者从一个计算机可读存储介质向另一计算机可读存储介质传输, 例如, 所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线 (例如同轴电缆、光纤、数字用户线 (DSL)) 20 或无线 (例如红外、无线、微波等) 方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存储的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质, (例如, 软盘、硬盘、磁带)、光介质 (例如, DVD)、或者半导体介质 (例如固态硬盘 Solid State Disk(SSD)) 等。

25 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成, 该程序可以存储于一计算机可读存储介质中, 存储介质可以包括: ROM、RAM、磁盘或光盘等。

30 本申请的说明书和权利要求书及上述附图中的术语“第一”, “第二”等是用于区别类似的对象, 而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换, 以便这里描述的实施例能够以除了在这里图示或描述的内容以外的顺序实施。本申请中术语“和/或”, 仅仅是一种描述关联对象的关联关系, 表示可以存在三种关系, 例如, A 和/或 B, 可以表示: 单独存在 A, 同时存在 A 和 B, 单独存在 B 这三种情况, 另外, 本文中字符“/”, 一般表示前后关联对象是一种“或”的关系。此外, 术语“包括” 35 和“具有”以及他们的任何变形, 意图在于覆盖不排他的包含, 例如, 包含了一系列步骤或模块的过程, 方法, 系统, 产品或设备不必限于清楚地列出的那些步骤或模块, 而是可包括没有清楚地列出的或对于这些过程, 方法, 产品或设备固有的其它步骤或模块。在本申请中出现的对步骤进行的命名或者编号, 并不意味着必须按照命名或者编号所指示的时

间/逻辑先后顺序执行方法流程中的步骤，已经命名或者编号的流程步骤可以根据要实现的技术目的变更执行次序，只要能达到相同或者相类似的技术效果即可。本申请中所出现的模块的划分，是一种逻辑上的划分，实际应用中实现时可以有另外的划分方式，例如多个模块可以结合成或集成在另一个系统中，或一些特征可以忽略，或不执行，另外，所显示

5 的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端口，模块之间的间接耦合或通信连接可以是电性或其他类似的形式，本申请中均不作限定。并且，作为分离部件说明的模块或子模块可以是也可以不是物理上的分离，可以是也可以不是物理模块，或者可以分布到多个电路模块中，可以根据实际的需要选择其中的部分或全部模块来实现本申请方案的目的。

权利要求

1. 一种训练模型的方法，其特征在于，包括：

获取多个训练样本，每个所述训练样本包括图像和文本，所述文本用于描述所述图像中的目标对象；

5 将所述多个训练样本输入目标模型中，以使所述目标模型执行下述流程，直至所述满足预设的停止条件：

提取第一图像的图像特征和第一文本的文本特征，所述第一图像是所述多个训练样本中任意一个所述训练样本中包括的图像，所述第一文本用于描述所述第一图像中的目标对象；

10 根据第一向量和第二向量之间的差异获取第一损失值，所述第一向量的维度和所述第二向量的维度相同，所述第一向量用于指示所述第一图像的图像特征，所述第二向量用于指示所述第一文本的文本特征；

所述根据所述第一损失值更新所述目标模型。

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：

15 根据所述第一图像的图像特征，预测所述第一图像中的目标对象的至少一个属性；

根据所述至少一个属性和预设标签之间的差异获取第二损失值，所述预设标签是对所述第一文本进行分词预处理后获取的；

所述根据所述第一损失值更新所述目标模型，包括：

根据所述第一损失值和所述第二损失值更新所述目标模型。

20 3. 根据权利要求2所述的方法，其特征在于，所述方法还包括：

根据所述第一图像的图像特征，预测所述第一图像中的目标对象属于对象集合中各个目标对象的第一概率，所述对象集合包括所述多个训练样本中全部所述图像中包括的目标对象；

25 根据所述第一文本的文本特征预测所述第一文本用于描述对象集合中各个目标对象的第二概率；

根据所述第一概率和所述第二概率之间的差异获取第三损失值；

所述根据所述第一损失值更新所述目标模型，包括：

根据所述第一损失值、所述第二损失值以及所述第三损失值更新所述目标模型。

4. 根据权利要求3所述的方法，其特征在于，所述方法还包括：

30 根据所述第一概率和第一预设值之间的差异，获取第四损失值，所述第一预设值指示预设的所述第一图像中的目标对象属于对象集合中各个目标对象的概率；

根据所述第二概率和第二预设值之间的差异，获取第五损失值，所述第二预设值指示预设的所述第一文本用于描述对象集合中各个目标对象的概率；

所述根据所述第一损失值更新所述目标模型，包括：

35 根据所述第一损失值、所述第二损失值、所述第三损失值、所述第四损失值以及所述第五损失值更新所述目标模型。

5. 根据权利要求4所述的方法，其特征在于，所述方法还包括：

根据所述第一图像的图像特征和第二图像的图像特征之间的差异获取第六损失值，所

述第二图像和所述第一图像从不同角度展示相同的目标对象；

所述根据所述第一损失值更新所述目标模型，包括：

根据所述第一损失值、所述第二损失值、所述第三损失值、所述第四损失值、所述第五损失值以及所述第六损失值更新所述目标模型。

5 6. 根据权利要求 5 所述的方法，其特征在于，所述方法还包括：

获取所述第一向量和第三向量之间的第一距离，获取所述第一向量和第四向量之间的第二距离，所述第三向量用于指示所述第二图像的图像特征，所述第四向量用于指示第三图像的图像特征，所述第三图像和所述第一图像中包括的目标对象不同；

根据所述第一距离和所述第二距离之间的差值与预设阈值的差异获取第七损失值；

10 所述根据所述第一损失值更新所述目标模型，包括：

根据所述第一损失值、所述第二损失值、所述第三损失值、所述第四损失值、所述第五损失值、所述第六损失值以及所述第七损失值更新所述目标模型。

7. 根据权利要求 1 至 6 任一项所述的方法，其特征在于，所述获取多个训练样本，包括：

15 获取产品的第一目标图像、所述产品的文本标题、所述产品的第二目标图像，所述第一目标图像来自于所述产品的提供者，所述第二目标图像来自所述产品的购买者；

将所述第一目标图像作为分割模型的输入，以获取分割结果，所述分割结果指示所述产品在所述第一目标图像中的轮廓；

20 根据所述分割结果提取第三目标图像的目标区域，所述第三目标图像是通过将所述第一目标图像输入生成器中获取的，其中所述生成器在训练过程中利用了来自所述产品的提供者提供的图像，以及来自所述产品的购买者提供的图像；

对所述目标区域进行随机仿射变换，以获取变换后的所述目标区域；

将所述变换后的目标区域与预设背景进行组合，以获取第四目标图像；

25 获取第一类型的训练样本、第二类型的训练样本以及第三类型的训练样本，每个所述第一类型的训练样本包括所述第一目标图像和所述文本标题，每个所述第二类型的训练样本包括所述第二目标图像和所述文本标题，每个所述第三类型的训练样本包括所述第四目标图像和所述文本标题。

8. 根据权利要求 1 至 7 任一项所述的方法，其特征在于，所述方法还包括：

30 根据所述目标模型提取的多个图像特征，对哈希模型进行多次训练，以获取训练后的所述哈希模型，其中，所述多次训练中的任意一次训练，包括：

获取每个所述图像特征对应的哈希特征；

对多个所述哈希特征进行聚类处理，以获取多个聚类数据集；

从多个所述哈希特征中获取第一哈希特征，所述第一哈希特征是所述多个哈希特征中的任意一个；

35 根据各个所述聚类数据集的聚类中心和所述第一哈希特征的相似性，从所述多个聚类数据集中获取第一聚类数据集，所述多个聚类数据中所述第一聚类数据集的聚类中心和所述第一哈希特征的相似性最高；

根据所述第一聚类数据集的聚类中心和所述第一哈希特征的差异获取第八损失值；
根据所述第八损失值更新所述哈希模型。

9. 根据权利要求 8 所述的方法，其特征在于，所述根据所述第一损失值更新所述目标模型，包括：

5 根据所述第一损失值、所述第二损失值、所述第三损失值、所述第四损失值、所述第五损失值、所述第六损失值、所述第七损失值以及所述第八损失值更新所述目标模型。

10. 根据权利要求 8 或 9 所述的方法，其特征在于，所述多次训练中的任意一次训练，还包括：

10 对目标哈希特征施加第一约束和第二约束，以获取更新后的所述目标哈希特征，所述目标哈希特征是根据上一次更新后的所述哈希模型获取的哈希特征，所述第一约束为所述哈希特征中的第三预设值和第四预设值的数目是平均的，所述第二约束为目标矩阵的第一对角线上的元素为所述第四预设值，所述目标矩阵中除所述第一对角线之外的其他元素为所述第三预设值，所述目标矩阵中第 i 行第 j 列的元素表示多个目标元素的平均值，所述多个目标元素表示各个所述哈希特征的第 i 个维度的取值和各个所述哈希特征各自的第 j
15 个维度的取值的乘积，所述 i 和所述 j 为正整数；

对所述目标哈希特征中每一维度的取值和第二均值进行比较，所述第二均值是所述第三预设值和所述第四预设值的平均值；

所述每一维度的取值小于所述第二均值时，根据所述哈希特征的每一维度的取值和所述第三预设值之间的差异获取第九损失值；

20 所述每一维度的取值不小于所述第二均值时，根据所述哈希特征的每一维度的取值和所述第四预设值之间的差异获取第九损失值；

所述根据所述第八损失值更新所述哈希模型，包括：

根据所述第八损失值和所述第九损失值更新所述哈希模型。

11. 一种图像检索的方法，其特征在于，包括：

25 获取待检索图像；

通过目标模型提取所述待检索图像的图像特征；

根据所述待检索图像的图像特征从图像数据库中获取预测图像，所述预测图像的图像特征和所述待检索图像的图像特征的相似度达到阈值；

30 其中，所述目标模型是通过第一损失值更新初始目标模型后获取的模型，所述第一损失值是通过第一向量和第二向量之间的差异获取的，所述第一向量的维度和所述第二向量的维度相同，所述第一向量用于指示第一图像的图像特征，所述第二向量用于指示第一文本的文本特征，所述第一图像是多个训练样本中任意一个所述训练样本中包括的图像，所述第一文本用于描述所述第一图像中的目标对象，每个所述训练样本包括图像和文本，所述文本用于描述所述图像中的目标对象。

35 12. 根据权利要求 11 所述的方法，其特征在于，所述目标模型具体是通过所述第一损失值和第二损失值更新所述初始目标模型后获取的模型，所述第二损失值是根据至少一个属性和预设标签之间的差异获取的，所述至少一个属性是根据所述第一图像的图像特征，

预测的所述第一图像中的目标对象的至少一个属性。

13. 根据权利要求 12 所述的方法，其特征在于，所述目标模型具体是通过所述第一损失值、所述第二损失值以及第三损失值更新所述初始目标模型后获取的，所述第三损失值是根据第一概率和第二概率之间的差异获取的，所述第一概率是根据所述第一图像的图像特征，预测所述第一图像中的目标对象属于对象集合中各个目标对象的概率，所述第二概率是根据所述第一文本的文本特征预测所述第一文本用于描述对象集合中各个目标对象的概率。

14. 权利要求 13 所述的方法，其特征在于，所述目标模型具体是通过所述第一损失值、所述第二损失值、所述第三损失值、第四损失值以及第五损失值更新所述初始目标模型后模型的，所述第四损失值是根据所述第一概率和第一预设值之间的差异获取的，所述第一预设值指示预设的所述第一图像中的目标对象属于对象集合中各个目标对象的概率，所述第五损失值是根据所述第二概率和第二预设值之间的差异获取的，所述第二预设值指示预设的所述第一文本用于描述对象集合中各个目标对象的概率。

15. 根据权利要求 14 所述的方法，其特征在于，所述目标模型具体是根据所述第一损失值、所述第二损失值、所述第三损失值、所述第四损失值、所述第五损失值以及第六损失值更新所述初始目标模型后获取的，所述第六损失值是根据所述第一图像的图像特征和第二图像的图像特征之间的差异获取的，所述第二图像和所述第一图像从不同角度展示相同的目标对象。

16. 根据权利要求 15 所述的方法，其特征在于，所述目标模型具体是根据所述第一损失值、所述第二损失值、所述第三损失值、所述第四损失值、所述第五损失值、所述第六损失值以及第七损失值更新所述初始目标模型后获取的，所述第七损失值是通过第一距离和第二距离之间的差值与预设阈值的差异获取的，所述第一距离是根据所述第一向量和第三向量之间的距离，所述第二距离是所述第一向量和第四向量之间的距离，所述第三向量用于指示所述第二图像的图像特征，所述第四向量用于指示第三图像的图像特征，所述第三图像和所述第一图像中包括的目标对象不同。

17. 根据权利要求 11 至 16 任一项所述的方法，其特征在于，所述多个训练样本包括第一类型的训练样本、第二类型的训练样本以及第三类型的训练样本，每个所述第一类型的训练样本包括所述第一目标图像和所述文本标题，每个所述第二类型的训练样本包括所述第二目标图像和所述文本标题，每个所述第三类型的训练样本包括所述第四目标图像和所述文本标题，所述第一目标图像来自于所述产品的提供者，所述第二目标图像来自所述产品的购买者，所述文本标题是所述产品的文本标题，所述第四目标图像是将变换后的目标区域与预设背景进行组合后获取的，所述变换后的目标区域是对所述目标区域进行随机仿射变换后获取的，所述目标区域是根据分割结果提取第三目标图像后获取的，所述第三目标图像是通过将所述第一目标图像输入生成器中获取的，其中所述生成器在训练过程中利用了来自所述产品的提供者提供的图像，以及来自所述产品的购买者提供的图像，所述分割结果是将所述第一目标图像作为分割模型的输入后获取的结果，所述分割结果指示所述产品在所述第一目标图像中的轮廓。

18. 根据权利要求 11 至 17 任一项所述的方法，其特征在于，所述方法还包括：

将所述待检索图像的图像特征输入至哈希模型中，以获取所述待检索图像的哈希特征；

所述根据所述待检索图像的图像特征从图像数据库中获取预测图像，包括：

5 根据所述待检索图像的哈希特征从所述图像数据库中获取预测图像，其中所述哈希模型是通过第八损失值更新初始哈希模型后获取的，所述第八损失值是根据第一聚类数据集的聚类中心和第一哈希特征的差异获取的，所述第一聚类数据集的聚类中心是所述多个聚类数据中各个聚类数据集的聚类中心中和所述第一哈希特征的相似性最高的聚类中心，所述多个聚类数据集是对多个哈希特征进行聚类处理后获取的，所述第一哈希特征是所述多个哈希特征中的任意一个。

10 19. 根据权利要求 18 所述的方法，其特征在于，所述目标模型具体是根据所述第一损失值、所述第二损失值、所述第三损失值、所述第四损失值、所述第五损失值、所述第六损失值、所述第七损失值以及所述第八损失值更新初始目标模型后获取的。

20. 一种训练模型的装置，其特征在于，包括：

存储器，用于存储计算机可读指令；

15 与所述存储器耦合的处理器，用于执行所述存储器中的计算机可读指令从而执行如权利要求 1 至 10 任一项所描述的方法。

21. 一种图像检索的设备，其特征在于，包括：

存储器，用于存储计算机可读指令；

20 与所述存储器耦合的处理器，用于执行所述存储器中的计算机可读指令从而执行如权利要求 11 至 19 任一项所描述的方法。

22. 一种芯片系统，其特征在于，所述芯片系统包括处理器和通信接口，所述处理器通过所述通信接口获取程序指令，当所述程序指令被所述处理器执行时实现权利要求 1 至 10 中任一项所述的方法。

25 23. 一种芯片系统，其特征在于，所述芯片系统包括处理器和通信接口，所述处理器通过所述通信接口获取程序指令，当所述程序指令被所述处理器执行时实现权利要求 11 至 19 中任一项所述的方法。

24. 一种计算机可读存储介质，其特征在于，包括程序，当其被处理单元所执行时，执行如权利要求 1 至 10 中任一项所述的方法。

30 25. 一种计算机可读存储介质，其特征在于，包括程序，当其被处理单元所执行时，执行如权利要求 11 至 19 中任一项所述的方法。

26. 一种计算机程序产品，其特征在于，当所述计算机程序产品在计算机上运行时，使得计算机执行如权利要求 1 至 10 中任一项所述的方法。

27. 一种计算机程序产品，其特征在于，当所述计算机程序产品在计算机上运行时，使得计算机执行如权利要求 11 至 19 中任一项所述的方法。

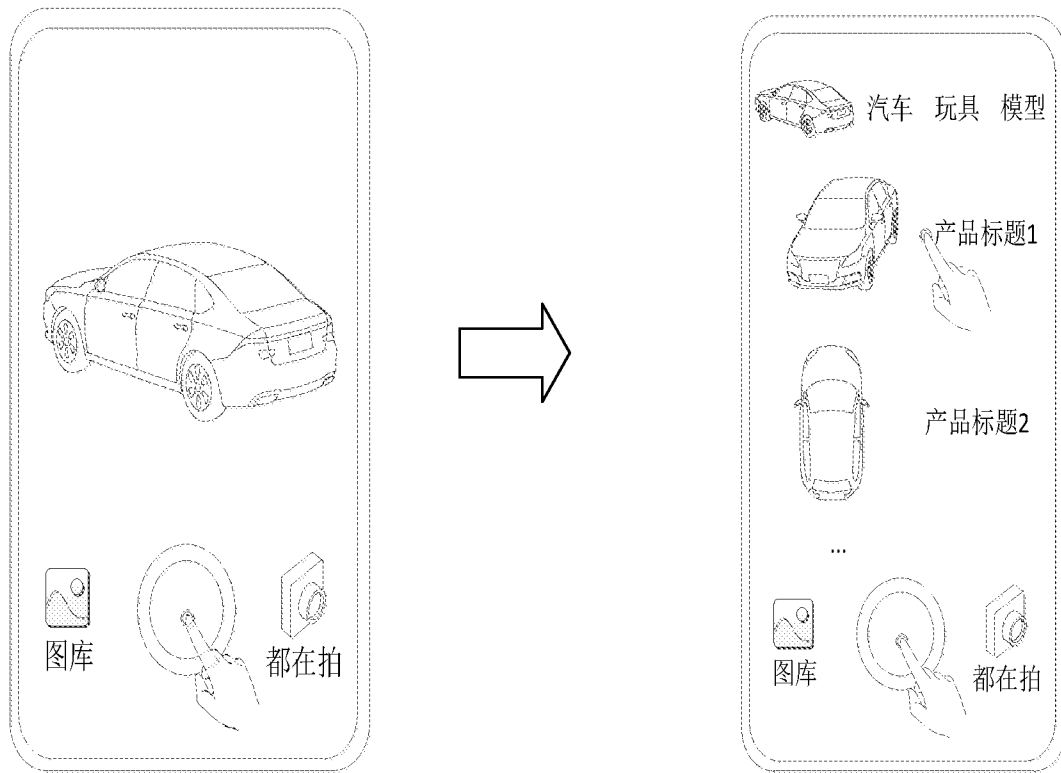
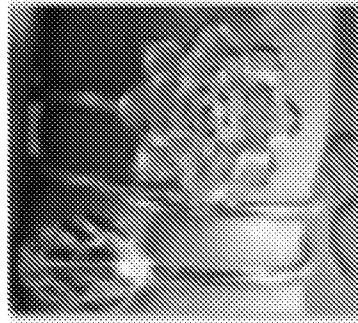
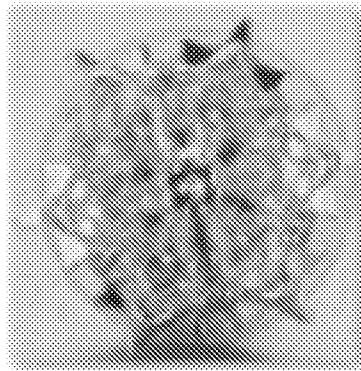


图 1

—2/11—



a



b



c

图 2

—3/11—

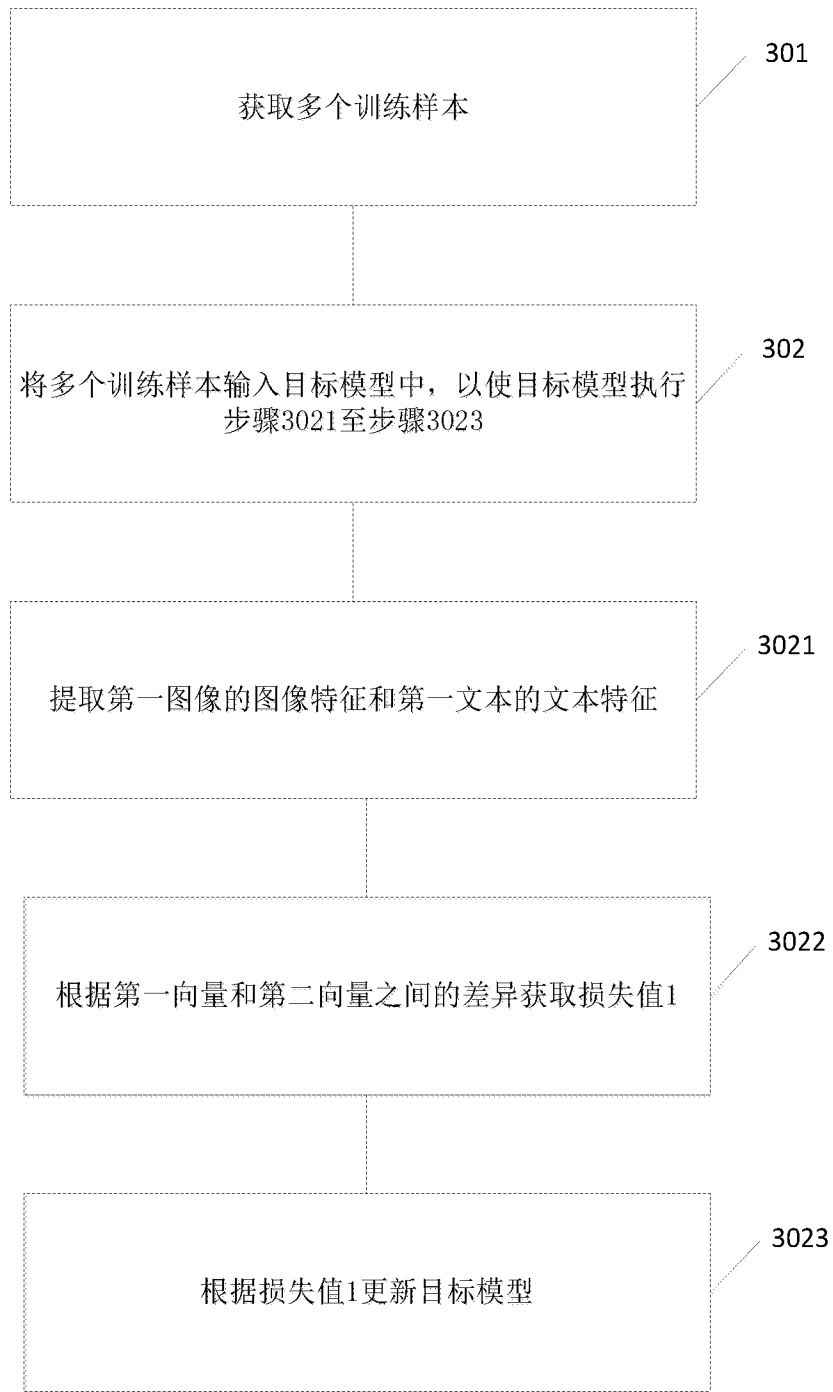


图 3

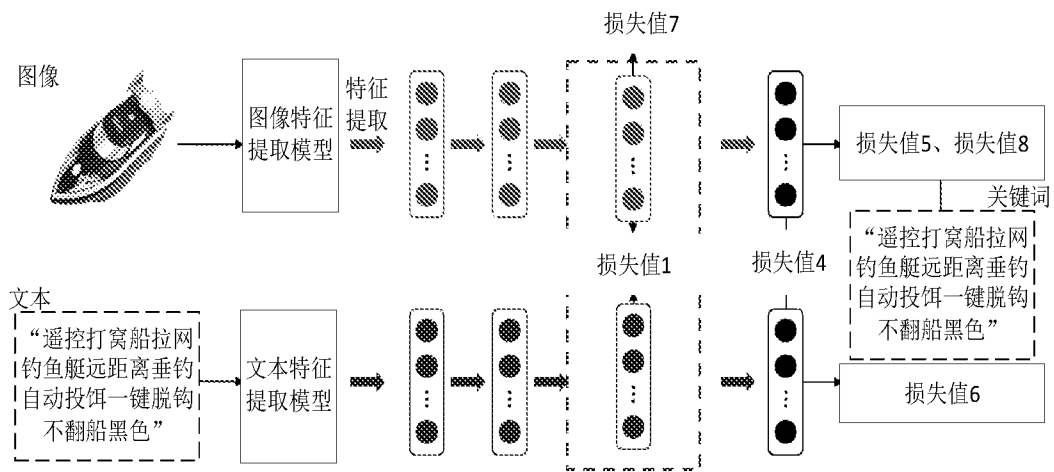


图 4

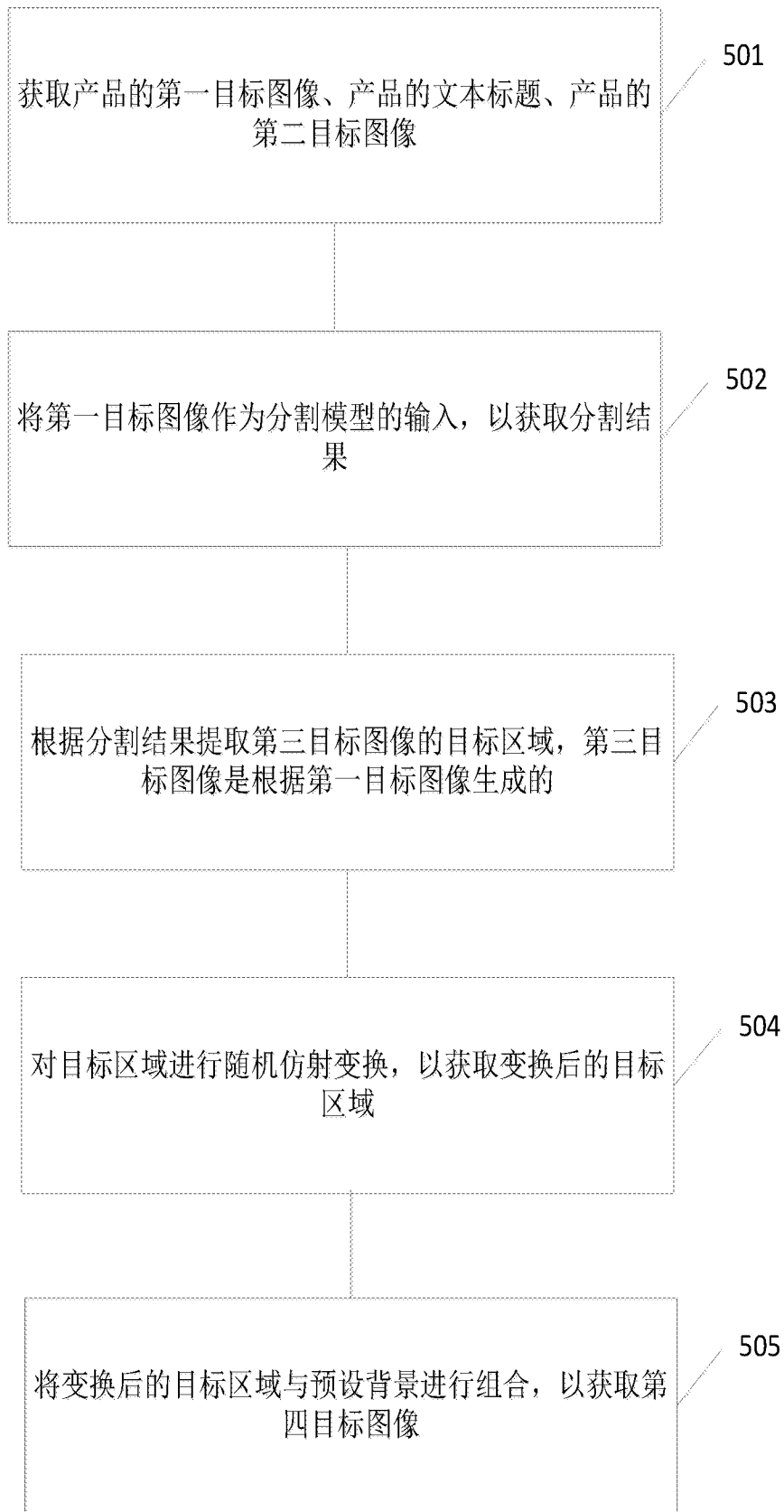


图 5

—6/11—

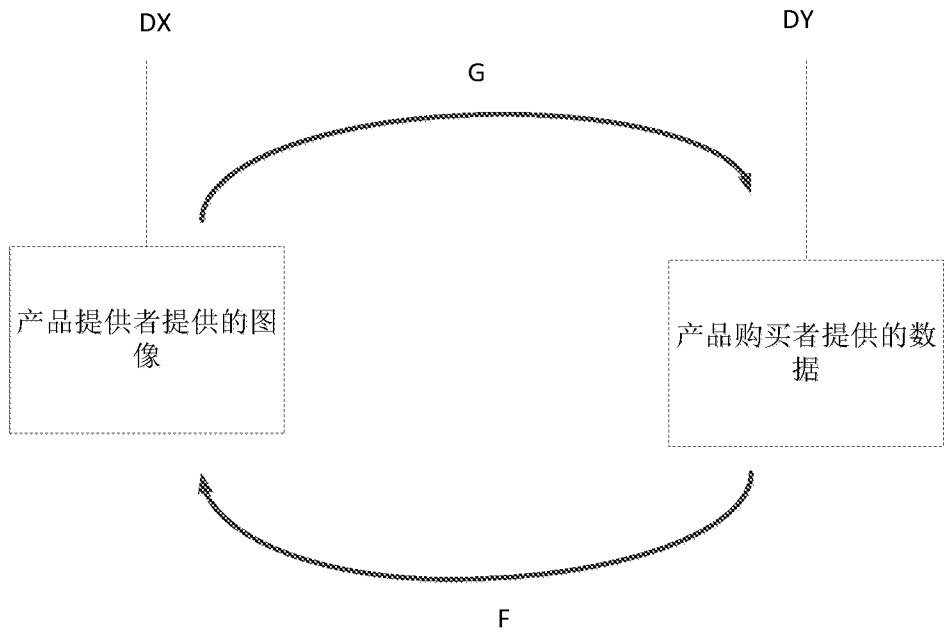


图 6

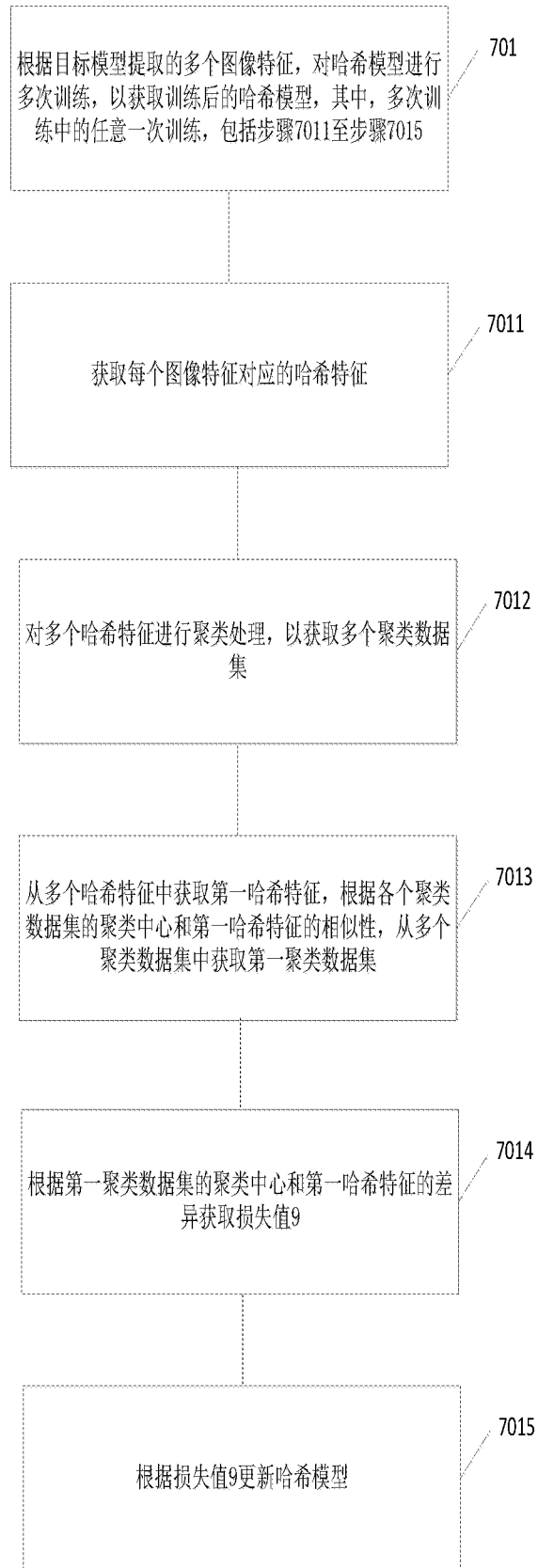


图 7

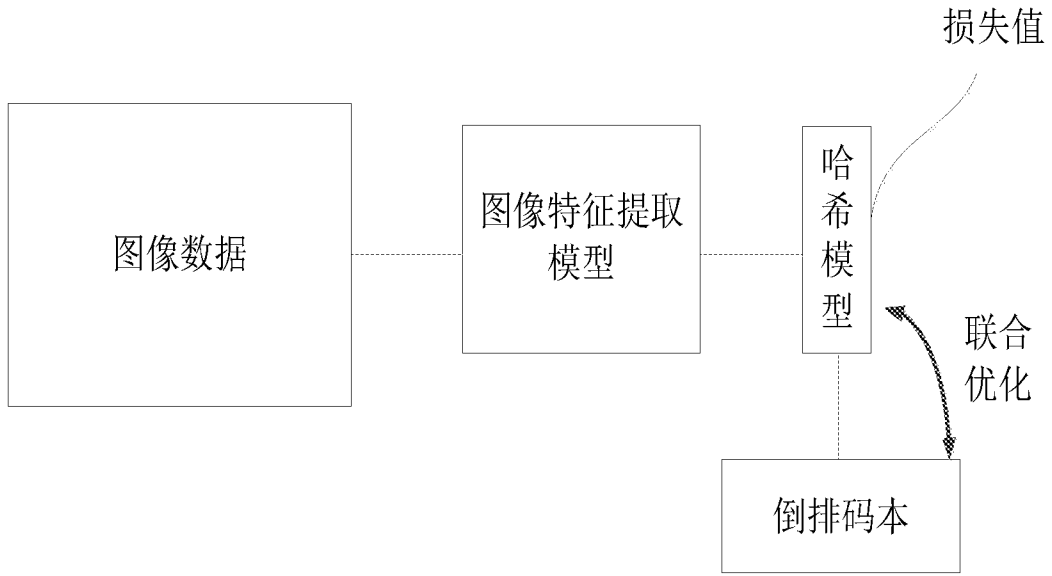


图 8

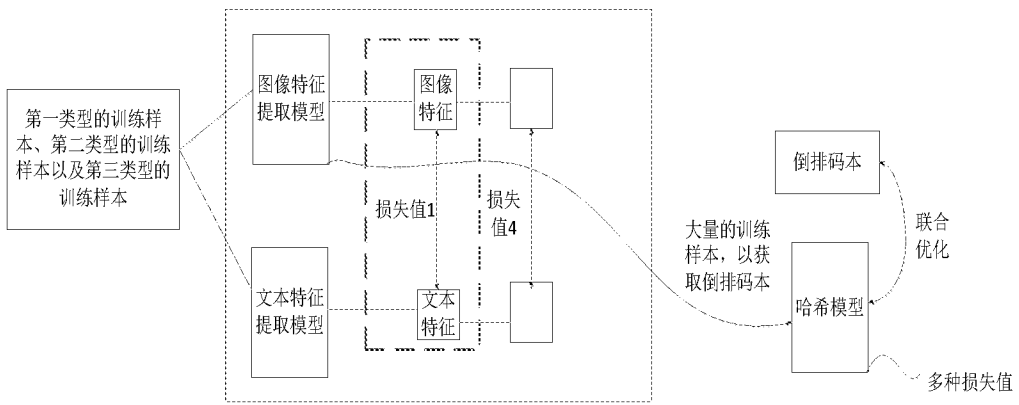


图 9



图 10

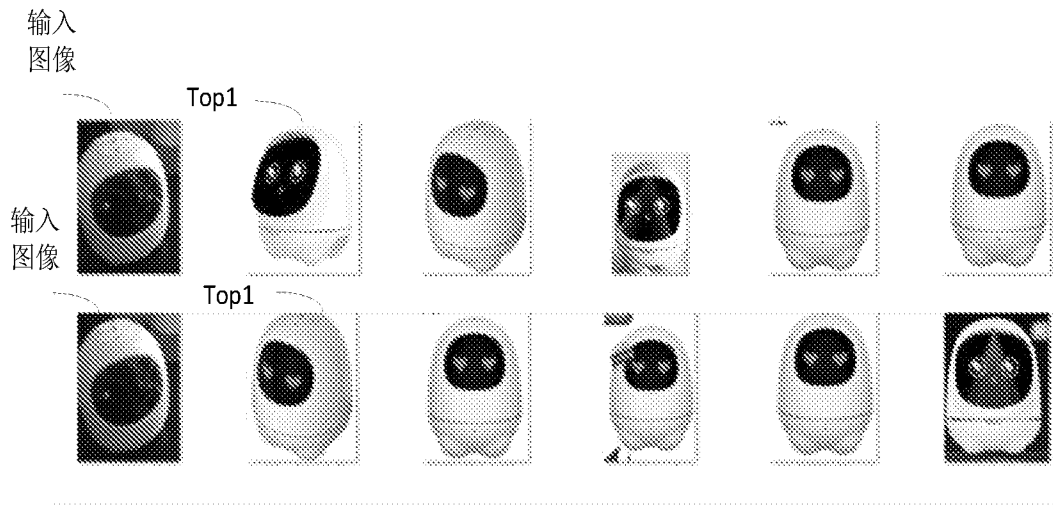


图 11

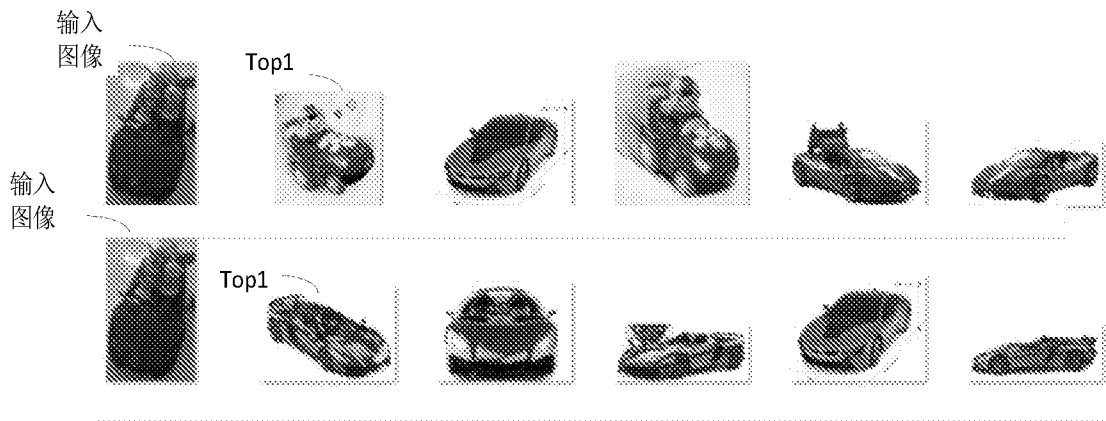


图 12

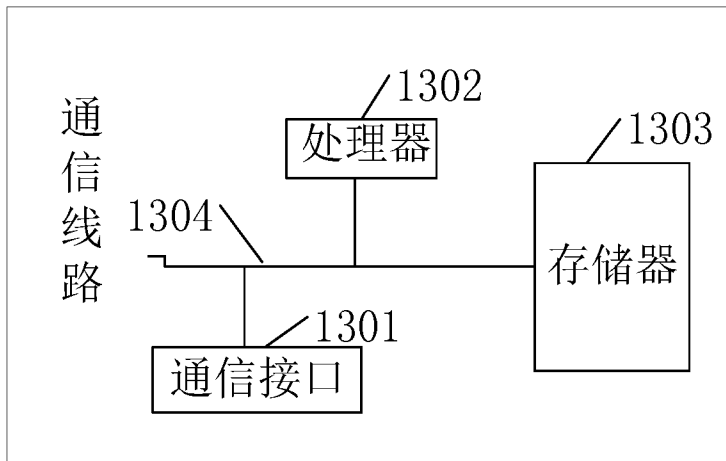


图 13

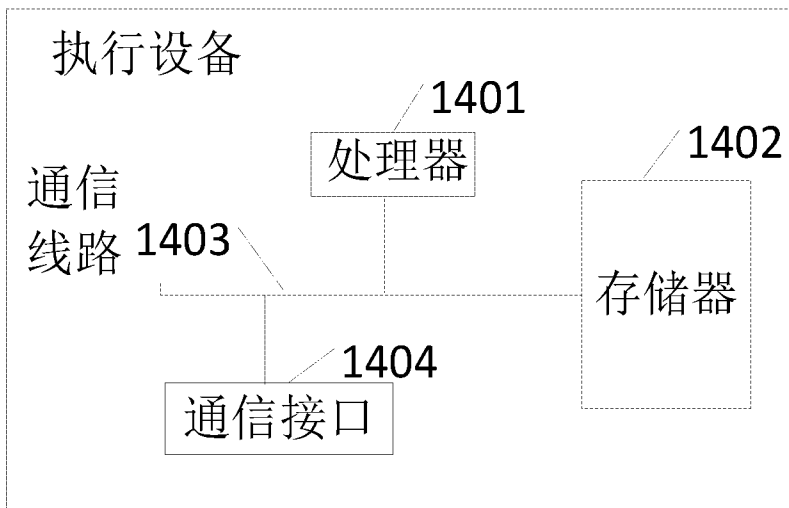


图 14

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/073923

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)		
CNPAT, CNKI, EPODOC, WPI: 模型, 样本, 训练, 图像, 文本, 差异, 差值, 损失值, 向量, 特征, 属性, 标签, 概率, model, sample, train, image, text, difference, loss, vector, feature, attribute, label, probability		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113159095 A (HUAWEI TECHNOLOGIES CO., LTD.) 23 July 2021 (2021-07-23) claims 1-27	1-27
X	CN 111209415 A (CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 29 May 2020 (2020-05-29) abstract, description, paragraphs 2-52	1, 11, 20-27
A	CN 110163267 A (XIAMEN MEITUZHIIA TECHNOLOGY CO., LTD.) 23 August 2019 (2019-08-23) entire document	1-27
A	CN 110414432 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 05 November 2019 (2019-11-05) entire document	1-27
A	US 2020160177 A1 (ROYAL BANK OF CANADA) 21 May 2020 (2020-05-21) entire document	1-27
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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		Date of mailing of the international search report
07 April 2022		29 April 2022
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China		
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/CN2022/073923

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113159095	A	23 July 2021	None			
CN	111209415	A	29 May 2020	None			
CN	110163267	A	23 August 2019	None			
CN	110414432	A	05 November 2019	None			
US	2020160177	A1	21 May 2020	CA	3061717	A1	16 May 2020

国际检索报告

国际申请号

PCT/CN2022/073923

<p>A. 主题的分类</p> <p>G06K 9/62 (2022.01) i</p> <p>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</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06K</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNPAT, CNKI, EPODOC, WPI: 模型, 样本, 训练, 图像, 文本, 差异, 差值, 损失值, 向量, 特征, 属性, 标签, 概率, model, sample, train, image, text, difference, loss, vector, feature, attribute, label, probability</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 113159095 A (华为技术有限公司) 2021年7月23日 (2021 - 07 - 23) 权利要求1-27</td> <td>1-27</td> </tr> <tr> <td>X</td> <td>CN 111209415 A (重庆邮电大学) 2020年5月29日 (2020 - 05 - 29) 摘要, 说明书第2-52段</td> <td>1, 11, 20-27</td> </tr> <tr> <td>A</td> <td>CN 110163267 A (厦门美图之家科技有限公司) 2019年8月23日 (2019 - 08 - 23) 全文</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>CN 110414432 A (腾讯科技深圳有限公司) 2019年11月5日 (2019 - 11 - 05) 全文</td> <td>1-27</td> </tr> <tr> <td>A</td> <td>US 2020160177 A1 (ROYAL BANK OF CANADA) 2020年5月21日 (2020 - 05 - 21) 全文</td> <td>1-27</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 113159095 A (华为技术有限公司) 2021年7月23日 (2021 - 07 - 23) 权利要求1-27	1-27	X	CN 111209415 A (重庆邮电大学) 2020年5月29日 (2020 - 05 - 29) 摘要, 说明书第2-52段	1, 11, 20-27	A	CN 110163267 A (厦门美图之家科技有限公司) 2019年8月23日 (2019 - 08 - 23) 全文	1-27	A	CN 110414432 A (腾讯科技深圳有限公司) 2019年11月5日 (2019 - 11 - 05) 全文	1-27	A	US 2020160177 A1 (ROYAL BANK OF CANADA) 2020年5月21日 (2020 - 05 - 21) 全文	1-27
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 113159095 A (华为技术有限公司) 2021年7月23日 (2021 - 07 - 23) 权利要求1-27	1-27																		
X	CN 111209415 A (重庆邮电大学) 2020年5月29日 (2020 - 05 - 29) 摘要, 说明书第2-52段	1, 11, 20-27																		
A	CN 110163267 A (厦门美图之家科技有限公司) 2019年8月23日 (2019 - 08 - 23) 全文	1-27																		
A	CN 110414432 A (腾讯科技深圳有限公司) 2019年11月5日 (2019 - 11 - 05) 全文	1-27																		
A	US 2020160177 A1 (ROYAL BANK OF CANADA) 2020年5月21日 (2020 - 05 - 21) 全文	1-27																		
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																				
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																				
<p>国际检索实际完成的日期</p> <p>2022年4月7日</p>		<p>国际检索报告邮寄日期</p> <p>2022年4月29日</p>																		
<p>ISA/CN的名称和邮寄地址</p> <p>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</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>姜玲玲</p> <p>电话号码 86-(10)-53961421</p>																		

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2022/073923

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	113159095	A	2021年7月23日	无			
CN	111209415	A	2020年5月29日	无			
CN	110163267	A	2019年8月23日	无			
CN	110414432	A	2019年11月5日	无			
US	2020160177	A1	2020年5月21日	CA	3061717	A1	2020年5月16日