

(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

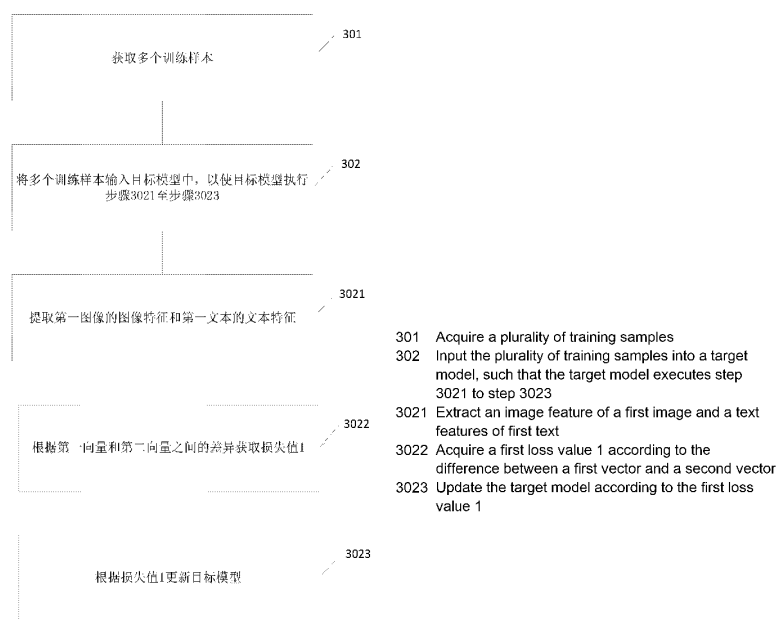


图 3

(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

[见续页]

(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。