

(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



(10) 国际公布号
WO 2022/206094 A1

(43) 国际公布日
2022 年 10 月 6 日 (06.10.2022)

(51) 国际专利分类号:
G06K 9/00 (2022.01) *G06F 40/30* (2020.01)
G06K 9/62 (2022.01) *G06N 3/04* (2006.01)

(21) 国际申请号: PCT/CN2022/070476

(22) 国际申请日: 2022 年 1 月 6 日 (06.01.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110338045.X 2021 年 3 月 30 日 (30.03.2021) CN

(71) 申请人: 京东科技控股股份有限公司
(JINGDONG TECHNOLOGY HOLDING CO., LTD.)
[CN/CN]; 中国北京市北京经济技术开发区科创十一街 18 号 C 座 2 层 221 室, Beijing 100176 (CN)。

(72) 发明人: 潘滢炜(PAN, Yingwei); 中国北京市北京经济技术开发区科创十一街 18 号 C 座 2 层 221 室, Beijing 100176 (CN)。 李业豪(LI, Yehao); 中国北京市北京经济技术开发区科创十一街 18 号 C 座 2 层 221 室, Beijing 100176 (CN)。 姚霆(YAO, Ting); 中国北京市北京经济技术开发区科创十一街 18 号 C 座 2 层 221 室, Beijing 100176 (CN)。 梅涛(MEI, Tao); 中国北京市北京经济技术开发区科创十一街 18 号 C 座 2 层 221 室, Beijing 100176 (CN)。

(74) 代理人: 北京英赛嘉华知识产权代理有限公司(INSIGHT INTELLECTUAL PROPERTY LIMITED); 中国北京市朝阳区建外大街光华东里 8 号中海广场中楼 7 层, Beijing 100020 (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,

(54) Title: METHOD AND APPARATUS FOR GENERATING CAPTIONING DEVICE, AND METHOD AND APPARATUS FOR OUTPUTTING CAPTION

(54) 发明名称: 用于生成字幕器以及输出字幕的方法和装置

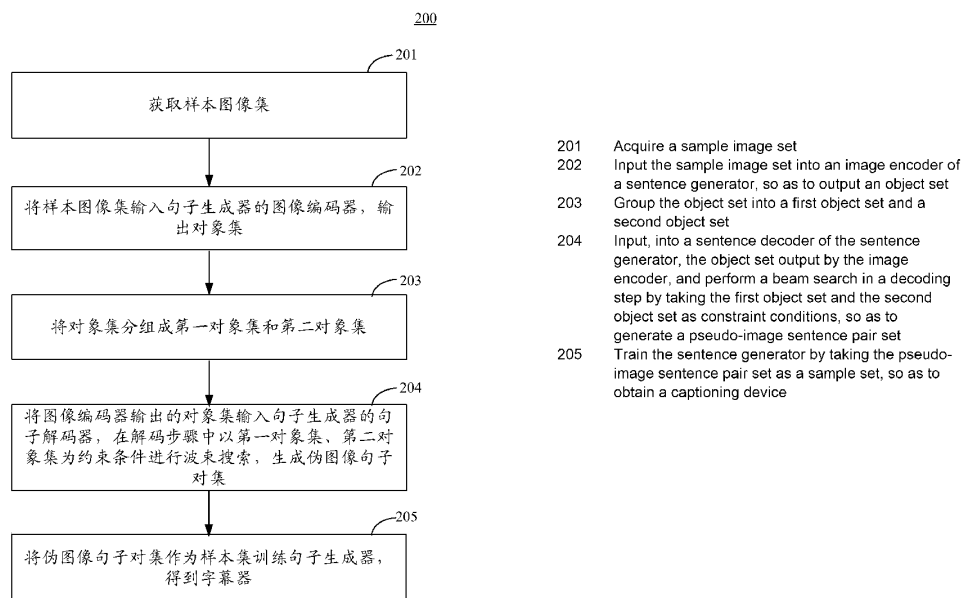


图 2

(57) Abstract: A method and apparatus for generating a captioning device, and a method and apparatus for outputting a caption. The method for generating a captioning device comprises: acquiring a sample image set; inputting the sample image set into an image encoder of a sentence generator, so as to output an object set; grouping the object set into a first object set and a second object set, wherein the first object set is an object set that is included within a predetermined object set, and the second object set is an object set that is excluded from the predetermined object set; inputting, into a sentence decoder of the sentence generator, the object set output by

[见续页]

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))。

the image encoder, and performing a beam search in a decoding step by taking the first object set and the second object set as constraint conditions, so as to generate a pseudo-image sentence pair set; and training the sentence generator by taking the pseudo-image sentence pair set as a sample set, so as to obtain a captioning device.

(57) 摘要: 一种用于生成字幕器的方法、装置以及用于输出字幕的方法、装置。用于生成字幕器的方法包括: 获取样本图像集; 将样本图像集输入句子生成器的图像编码器, 输出对象集; 将对象集分组成第一对象集和第二对象集, 其中, 第一对象集为被包含在预定对象集内的对象集, 第二对象集为被排除在预定对象集外的对象集; 将图像编码器输出的对象集输入句子生成器的句子解码器, 在解码步骤中以第一对象集、第二对象集为约束条件进行波束搜索, 生成伪图像句子对集; 将伪图像句子对集作为样本集训练句子生成器, 得到字幕器。