

(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
2019 年 5 月 9 日 (09.05.2019)



(10) 国际公布号
WO 2019/085905 A1

(51) 国际专利分类号:
G06F 16/50 (2019.01)

中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。

(21) 国际申请号: PCT/CN2018/112735

(22) 国际申请日: 2018 年 10 月 30 日 (30.10.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711049294.7 2017年10月31日 (31.10.2017) CN

(71) 申请人: 北京市商汤科技开发有限公司(BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD) [CN/CN]; 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。

(72) 发明人: 鲁盼(LU, Pan); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。 李鸿升(LI, Hongsheng); 中国北京市海淀区中关村东路1号院3号楼7层710-712房间, Beijing 100084 (CN)。 王晓刚(WANG, Xiaogang);

(74) 代理人: 北京思源智汇知识产权代理有限公司 (BEIJING SIYUANZHUI INTELLECTUAL PROPERTY AGENCY); 中国北京市海淀区中关村东路18号财智国际大厦B座502室, Beijing 100083 (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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(54) Title: IMAGE QUESTION ANSWERING METHOD, DEVICE AND SYSTEM, AND STORAGE MEDIUM

(54) 发明名称: 图像问答方法、装置、系统和存储介质

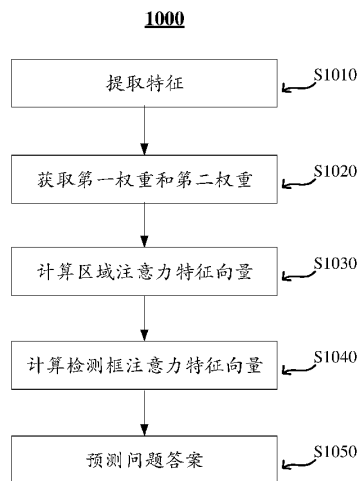


图 1

S1010 Extract features
S1020 Obtain a first weight and a second weight
S1030 Calculate a regional attention feature vector
S1040 Calculate a detection box attention feature vector
S1050 Predict an answer to a question

(57) Abstract: The present application provides an image question answering method, device and system, and a storage medium. The image question answering method comprises: extracting a question feature representing semantics of a question, a global feature of an image, and a detection box feature of a detection box surrounding an object in the image; obtaining, on the basis of the question feature, the global feature, and the detection box feature, a first weight of each of at least one region of the image and a second weight of each of at least one detection box of the image; performing weighing processing on the global feature using the first weight to obtain a regional attention feature of the image; performing weighing processing on the detection box using the second weight to obtain a detection box attention feature of the image; and predicting an answer to the question on the basis of the question feature, the regional attention feature, and the detection box attention feature. The technical solution provided by the present application can comprehensively take a global feature and a detection box feature comprising an object detection box of an image into consideration, thereby improving the efficiency and precision of image question answering.

WO 2019/085905 A1

[见续页]

(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))。

(57) 摘要: 本申请提供了一种图像问答方法、装置、系统和存储介质。图像问答方法包括: 提取表示问题的语义的问题特征、图像的全局特征、以及包围图像中的对象的检测框的检测框特征; 基于问题特征、全局特征和检测框特征获取图像的至少一个区域中每个区域的第一权重以及图像的至少一个检测框中每个检测框的第二权重; 利用第一权重对全局特征加权处理, 得到图像的区域注意力特征; 利用第二权重对检测框特征加权处理, 得到图像的检测框注意力特征; 基于问题特征、区域注意力特征和检测框注意力特征预测问题的答案。本申请提供的技术方案可综合考虑图像的全局特征和包含有对象检测框的检测框特征, 从而提高图像问答的效率和精度。