

(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 3 月 23 日 (23.03.2023)



(10) 国际公布号
WO 2023/040298 A1

(51) 国际专利分类号:
G06V 10/40 (2022.01)

(21) 国际申请号: PCT/CN2022/091369

(22) 国际申请日: 2022 年 5 月 7 日 (07.05.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202111085396.0 2021 年 9 月 16 日 (16.09.2021) CN

(71) 申请人: 京东科技信息技术有限公司 (JINGDONG TECHNOLOGY INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 6 层 601, Beijing 100176 (CN)。

(72) 发明人: 张熠恒 (ZHANG, Yiheng); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 6 层 601, Beijing 100176 (CN)。邱钊凡 (QIU, Zhaofan); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 6 层 601, 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 号院 2 号楼 6 层 601, 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 号院 2 号楼 6 层 601, Beijing 100176 (CN)。

(74) 代理人: 中国贸促会专利商标事务所有限公司 (CCPIT PATENT AND TRADEMARK LAW OFFICE); 中国北京市复兴门内大街 158 号远洋大厦 F10 层, Beijing 100031 (CN)。

(54) Title: VIDEO REPRESENTATION SELF-SUPERVISED CONTRASTIVE LEARNING METHOD AND APPARATUS

(54) 发明名称: 视频表征自监督对比学习方法和装置

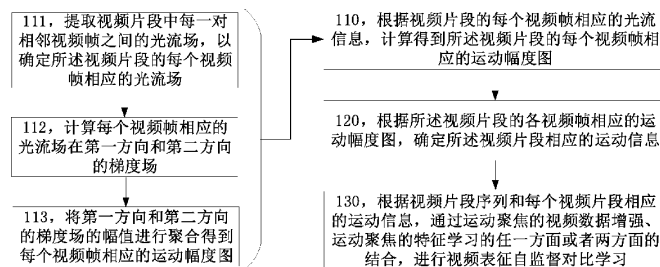


图 1

- 110 According to optical flow information corresponding to each video frame of a video clip, obtain, by means of calculation, a motion amplitude diagram which corresponds to each video frame of the video clip
- 111 Extract an optical flow field between each pair of adjacent video frames in a video clip, so as to determine an optical flow field corresponding to each video frame of the video clip
- 112 Calculate gradient fields, in a first direction and a second direction, of the optical flow field corresponding to each video frame
- 113 Aggregate the amplitudes of the gradient fields in the first direction and the second direction, so as to obtain the motion amplitude diagram corresponding to each video frame
- 120 According to the motion amplitude diagram corresponding to each video frame of the video clip, determine motion information which corresponds to the video clip
- 130 According to a video clip sequence, and the motion information corresponding to each video clip, perform video representation self-supervised contrastive learning by means of either of motion-focused video data augmentation and motion-focused feature learning or a combination of the two aspects

(57) Abstract: The present disclosure relates to the field of video learning. Provided are a video representation self-supervised contrastive learning method and apparatus. The method comprises: according to optical flow information corresponding to each video frame of a video clip, obtaining, by means of calculation, a motion amplitude diagram which corresponds to each video frame of the video clip; according to the motion amplitude diagram corresponding to each video frame of the video clip, determining motion information which corresponds to the video clip; and performing video representation self-supervised contrastive learning according to a video clip

[见续页]

WO 2023/040298 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, IQ, 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。

(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))。

sequence, and the motion information corresponding to each video clip. Therefore, a motion-focused contrastive learning solution for video representation self-supervised learning is implemented, such that widely existing and very important motion information in a video is fully utilized during a learning process, thereby improving the performance of video representation self-supervised contrastive learning.

(57) 摘要: 本公开提出一种视频表征自监督对比学习方法和装置, 涉及视频学习领域。该方法包括: 根据视频片段的每个视频帧相应的光流信息, 计算得到所述视频片段的每个视频帧相应的运动幅度图; 根据所述视频片段的各视频帧相应的运动幅度图, 确定所述视频片段相应的运动信息; 根据视频片段序列和每个视频片段相应的运动信息, 进行视频表征自监督对比学习。从而实现用于视频表征自监督学习的运动聚焦对比学习方案, 使得视频中广泛存在且十分重要的运动信息在学习过程中被充分利用, 从而提升视频表征自监督对比学习性能。