

(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
2024 年 1 月 25 日 (25.01.2024)



(10) 国际公布号
WO 2024/016464 A1

- (51) 国际专利分类号:
G06T 7/50 (2017.01) **G06N 3/04** (2006.01) [CN/CN]; 中国浙江省台州市椒江区工人东路 293 号 1001 室, Zhejiang 200433 (CN)。
- (21) 国际申请号: PCT/CN2022/120231 (72) 发明人: 郝霖(HAO, Lin); 中国上海市浦东新区祖冲之路 2277 弄 11 号楼, Shanghai 201203 (CN)。王
(22) 国际申请日: 2022 年 9 月 21 日 (21.09.2022) 国权(WANG, Guoquan); 中国上海市浦东新区祖
(25) 申请语言: 中文 冲之路 2277 弄 11 号楼, Shanghai 201203 (CN)。叶
(26) 公布语言: 中文 德建(YE, Dejian); 中国上海市浦东新区祖冲之路
(30) 优先权: 2277 弄 11 号楼, Shanghai 201203 (CN)。
202210864352.6 2022 年 7 月 21 日 (21.07.2022) CN (74) 代理人: 上海弼兴律师事务所 (SHANGHAI
(71) 申请人: 上海清鹤科技股份有限公司(CLEARTV BESHINING LAW OFFICE); 中国上海市小木桥路
CO., LTD.) [CN/CN]; 中国上海市浦东新区祖冲之 681 号外经大厦 21 楼, Shanghai 200032 (CN)。
路 2277 弄 11 号楼, Shanghai 201203 (CN)。浙江清 (81) 指定国(除另有指明, 要求每一种可提供的国家
鹤科技有限公司(ZHEJIANG CLEARTV CO., LTD.) 保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,

(54) Title: THREE-DIMENSIONAL AUTO-ENCODER AND TRAINING METHOD THEREFOR, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 三维自编码器及其训练方法、电子设备及存储介质

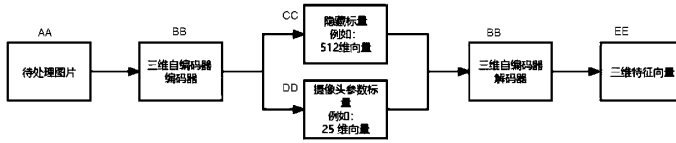


图 1

AA Picture to be processed
BB Three-dimensional auto-encoder and decoder
CC Hidden scalar, for example: 512-dimensional vector
DD Camera parameter scalar, for example: 25-dimensional vector
EE Three-dimensional feature vector

(57) Abstract: Disclosed are a three-dimensional auto-encoder and a training method therefor, a training method for a three-dimensional vision module, an electronic device, and a storage medium. The three-dimensional auto-encoder comprises an encoder and a decoder; the encoder is used for extracting spatial feature parameters of an input picture to be processed and outputting the spatial feature parameters to the decoder; the decoder is used for outputting, according to the spatial feature parameters, a target picture comprising depth information of the picture to be processed; and the picture to be processed and the target picture are used for determining a loss function of the three-dimensional auto-encoder. According to the present invention, perception of three-dimensional scenario features of an image is implemented by means of self-supervision. Moreover, training can be performed by using massive training sets and using a large amount of unlabeled original data, which saves the labor and cost, and a trained model can be conveniently transplanted to other scenarios. Inductive bias of a three-dimensional spatial relationship is added to a neural network design, so that interactive information with an environment comprised in an action can be perceived and the accuracy of recognition can be improved.

(57) 摘要: 本发明公开了一种三维自编码器及其训练方法、三维视觉模块的训练方法、电子设备及存储介质。三维自编码器包括编码器和解码器; 所述编码器用于提取输入的待处理图片的空间特征参数并输出至所述解码器; 所述解码器用于根据所述空间特征参数输出包括所述待处理图片深度信息的目标图片, 所述待处理图片和所述目标图片用于确定所述三维自编码器的损失函数。本发明通过自监督方式实现对图像的三维场景特征的感知; 且能够采用海量训练集, 利用大量的未标注原始数据进行训练, 节省人工和成本, 训练好的模型还能方便地移植到其他场景。通过加入三维空间关系的归纳偏置的神经网络设计可以感知到动作中包含的与环境的交互信息, 提高识别的准确率。

BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, 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, SC, 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)。

根据细则4.17的声明:

- 关于申请人有权申请并被授予专利(细则4.17(ii))
- 发明人资格(细则4.17(iv))

本国际公布:

- 包括国际检索报告(条约第21条(3))。