

深度学习在视觉 SLAM 研究中的应用综述

敬学良, 王晨升, 杨光, 李彦江

(北京邮电大学自动化学院, 北京, 100876)

摘要: 目前采用视觉传感器的地图同步定位与地图构建 (SLAM) 逐渐成为 SLAM 研究中的重点方向, 首先介绍了视觉 SLAM 的基本结构, 并分析了传统中特征点法和直接法的局限性; 然后重点对视觉 SLAM 中采用深度学习方法的视觉里程计和回环检测的最新研究成果进行了综述, 并把深度学习方法和传统方法进行了简要对比; 最后对视觉 SLAM 的发展趋势进行展望。

关键词: 深度学习; 视觉 SLAM; 视觉里程计; 回环检测

中图分类号: TP24

A Survey of Visual SLAM Application of Deep Learning in Visual SLAM Research

JING Xueliang, WANG Chensheng, YANG Guang, LI Yangjiang

(Beijing university of posts and telecommunications automation, Beijing 100876)

Abstract: At present, Simultaneous Localization and Mapping (SLAM) using visual sensors has gradually become a hot research topic of SLAM. Firstly, the basic structure of visual SLAM is introduced, and the limitations of traditional feature-based method and direct method are analyzed. Then the latest research results of visual odometry and loop closure detection using depth learning method in visual SLAM are reviewed, and the deep learning method and traditional method are compared briefly. Finally, the development trend of visual SLAM is proposed.

Key words: deep learning; visual SLAM((simultaneous localization and mapping); visual odometry; loop closure detection

0 引言

同时定位与地图构建 (SLAM) 的出现, 是为了使机器人及其它无人驾驶设备具备在未知环境中, 对自身定位的同时可以构建地图的能力, 从而实现真正的自主移动^[1]。在 SLAM 的众多解决方案中, 包含有各种不同类别的传感器。目前, 主流的传感器有激光雷达、声呐、摄像机^[2]。其中, 使用摄像机的 SLAM 被称为视觉 SLAM^[3]。其基本框架及如图 1 所示, 主要包括视觉里程计、后端优化、回环检测和地图构建。一般的运行流程是从摄像机获取原始图像, 由视觉里程计进行相机的位姿估计^[4], 由于此部分的工作不可避免地会产生误差, 因此需要在后端部分对视觉里程计的输出进行优化, 同时由回环检测部分判断相机是否出现在之前出现过的位姿。在地图构建部分, 依据图像和优化后的相机位姿信息, 建立应用所需的环境地图。

文献调研表明, 传统视觉 SLAM 在动态及光照变化大等场合下, 运行效果较差。而近年来, 深度学习在计算机视觉诸多领域取得了优秀的成果。利用深度学习技术, 可以更加准确的完成目标识别, 跟踪以及分割任务。因此, 使用深度学习技术, 来解决传统视觉 SLAM

基金项目: 江苏省重点研发计划---产业前瞻与共性关键技术 (BE2016032); 2016 年无锡市科技发展 (产业前瞻与共性技术) 项目 (CGE02G1609)

作者简介: 敬学良 (1994-), 男, 北京邮电大学硕士生, 主要研究方向: 智能装备及系统设计

通信联系人: 王晨升 (1964-), 男, 博士, 北京邮电大学副教授、硕导, 主要研究方向: 智慧控制, 虚拟现实. E-mail: cwang@bupt.edu.cn

40 中存在的一些问题，成为了当前视觉 SLAM 的一个发展趋势。

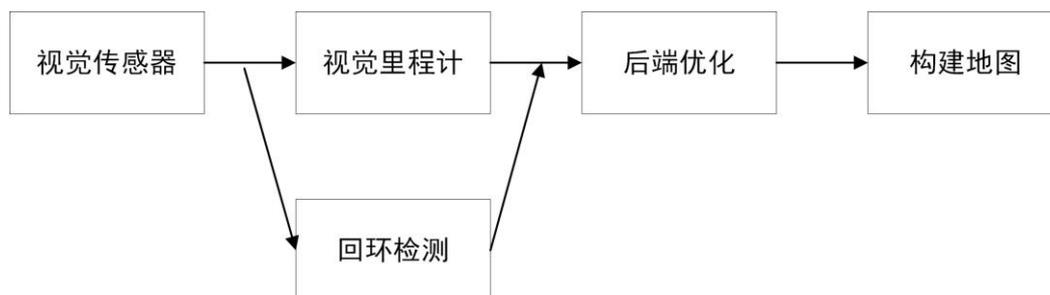


图 1 视觉 SLAM 基本框架

Fig. 1 Visual SLAM basic framework

45 本文对近年来深度学习方法在视觉 SLAM 研究中的应用研究进行了调研，把深度学习的应用分为视觉里程计和回环检测两部分展开综述，并对深度学习方法和传统方法进行了简要对比，分析了未来视觉 SLAM 的发展方向。

1 深度学习在视觉 SLAM 的应用分类

传统视觉 SLAM 根据对数据处理的方式不同主要分为两种方法，分别为特征点法和直接法。特征点法需要先对图像提取特征点并计算描述子，描述子用来衡量特征点的相似度，文献[5]介绍了一些主流的特征提取的方法，常见的描述子有 BRIEF^[6]、SURF^[7]、SIFT^[8]、HoG^[9]和 ORB^[10]，使用特征点方法的 SLAM 框架主要有主要代表有 PTAM^[11]和 ORB-SLAM^[12]。直接法根据使用的像素数量，可以进一步细分为稠密、半稠密和稀疏方法。直接法的典型代表有：DTAM^[13]、LSD-SLAM^[14]和 SVO^[15]。

55 特征点法对于较大的运动变化误差与直接法相比不敏感，鲁棒性更好，但是特征过多或过少均不能正常运行，提取特征点和计算描述子需要耗费大量时间，并且提取特征点后丢失了许多环境信息，只能构建出稀疏地图。直接法无需提取特征点和计算描述子，减少了时间消耗，在计算能力受限的场合中比较适用，但是该方法基于灰度不变假设理论^[16]，因此不能在光照变化大的情况下工作。

60 SLAM 的研究者针对传统方法的不足，把深度学习方法应用在视觉 SLAM 中的视觉里程计和回环检测部分，来改善对原始图像数据的处理效果。

2 结合深度学习的视觉里程计

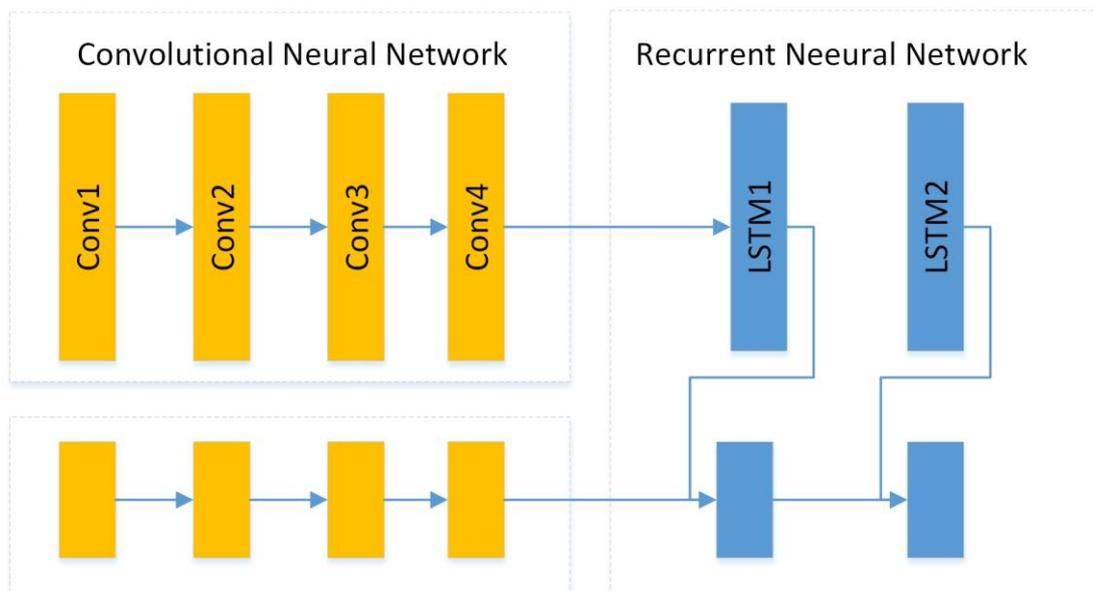
视觉里程计 (Visual Odometry, VO) 也被称作帧间估计，它通过观察当前环境的图像序列，估计机器人的运动情况，即相对于参考系的平移和旋转^[17]。和车轮里程计相似，视觉里程计的预估也会受不断累积的误差影响。但是，视觉里程计已被证明在长时间运行时，其准确性和可靠性都强于车轮里程计，并且通常不会受不平坦地形造成的车轮滑动的影响^[18]。传统的视觉里程计算法，一般采用基于特征点的方法或直接法。但是基于特征点需要复杂的特征提取和运算，使用直接法则依赖于像素强度值。如果结合深度学习方法，则可以充分利用输入图像的信息，CNN 等神经网络在图像处理方面的优势已得到充分验证，研究人员也尝试利用各种神经网络来实现视觉里程计。

70 文献[19]提出了一种使用端到端的深度学习架构，从视觉信息预测速度和方向变化的方法。该架构使用单一类型的计算模块和学习规则来从原始数据中提取视觉运动，深度和最终

的测距信息。在这项工作中，通过训练 CNN 把局部深度和运动的表示与速度和方向的局部变化相关联，从而学习执行视觉测距。为了避免训练 CNN 导致噪声滤波器和训练数据过度拟合，作者使用了同步/深度自动编码器（SAE-D）^[20]学习的特征初始化深度卷积网络。

75 Mohanty 等人^[21]采用 CNN 架构，部分基于 AlexNet^[22]，将视觉里程计预测作为回归问题。作者通过不同的数据集划分方式，做了已知环境和未知环境的对比实验，并观察了加入 FAST 特征或进行预训练对未知环境下实验效果的影响。最终表明该方法能够学习类似于 FAST 的特征，但是添加这些特征是没有必要的。当部署在已知环境中时，网络架构能够实时学习实际尺度，而使用几何方法的单目视觉里程计需要先进行人工标定。

80 Wang 等人^[23]利用深度递归卷积神经网络（RCNNs），实现了端到端的单目视觉里程计。作者认为视觉里程计本质是几何问题，而诸如 VGGNet^[24]和 GoogLeNet^[25]等 DNN 架构，其中大多数设计都考虑到了识别，分类和检测问题，这意味着它们的训练需要从外观和图像环境中学习知识。因此不能直接用于视觉里程计。该系统使用原始 RGB 图像作为 CNN 的输入，学习得到降维的卷积特征 Conv6 传递给 RNN 进行顺序建模，为了更好的获取长轨迹之间图片的联系，在 RNN 中加入了 LSTM 单元，构成如图 2 所示的结构。由于该系统不依赖于传统 VO 算法中的任何模块以及摄像机标定来进行姿态估计，且以端到端的方式进行训练，因此不需要细致地调整 VO 系统的参数，并且具有较好的泛化能力，可以在陌生场景中良好运行。



90 图 2 带有 LSTM 单元的 RCNN

Fig. 2 RCNN with LSTM

95 文献[26]提出了一种名为 UnDeepVO 的新型单目视觉里程计系统。UnDeepVO 能够通过使用深度神经网络估计单目相机的 6-DoF 姿势及其视野的深度。该系统的姿势估计部分采用基于 VGG 的 CNN 架构，使用连续的相邻的单眼图像作为输入，并预测它们之间的 6-DoF 变换。深度估计部分主要基于编码器解码器架构来生成稠密深度图，与文献[27]、[28]中的深度估计方法不同，它们从网络产生视差图像（深度的倒数）。和文献[23]相比该方法是无监督的，可以减少标注工作。

高翔在[29]中提出的 PoseNet，是首次利用卷积神经网络实现了实时的 6-DOF 单目相机位姿回归系统。PoseNet 在 GoogLeNet 的基础上进行了修改，用仿射变换回归器替换了全部

100 三个 softmax 分类器，在最后的回归器前插入了一个全连接层来提升泛化效果。该系统可以在室内和室外环境下实时运行，并可在大型户外场景中获得约 2 米和 3 度的精度。

3 结合深度学习的回环检测

105 回环检测也被称为闭环检测，通过非相邻帧来检测是否出现相同场景并添加约束。假设机器人在移动并连续观察环境特征，而由于路线出现环路，机器人回到了初始位置^[30]。此时由于存在累积误差，其预测距离已经偏离了实际位置，但是通过观察当前帧机器人可以得知它已回到初始位置，因此它可以在当前帧和初始帧之间添加约束，减少了位置与地图估计的总体漂移。通常进行回环检测，需要使用特征匹配技术来对当前帧和之前的所有帧中抽取的关键帧进行匹配。传统的描述图像的特征有词袋 (bag of visual words)^[31]、本地增强描述向量 (Vector of Locally Augmented Descriptors)^[32]等。随着深度学习的发展，使用深度神经网络对图像的识别匹配的准确率更高，对于提升回环检测的准确率也有帮助。

110 文献[33]较早的利用深度学习进行回环检测，作者基于 Places 数据集生成 CNN 预训练模型，CNN 的每一层输出都作为图像的特征向量，并采用公式(1)的方法进行正规化，其中表示 d 维的特征向量 X ，所有层的输出组成图像的描述符，用于回环检测。实验证明该方法在光照变化的情况下明显优于基于手工描述的传统方法。

$$\left(x_1, \dots, x_d\right) \leftarrow \left(\frac{x_1}{\sqrt{\sum_{i=1}^d x_i^2}}, \dots, \frac{x_d}{\sqrt{\sum_{i=1}^d x_i^2}} \right) \quad (1)$$

115 随后的回环检测研究中，人们对使用预训练模型的神经网络结构进行了各种改进。文献[34]的方法改变了使用全部图像特征的思路，仅使用 CNN 提取局部特征来进行图像的相似性对比，实验结果证明了该方法可以有效处理视角改变或有局部遮挡的情况。Ruben^[35]则是在预训练的基础上，再次从数据集选取图像三元组进行训练，三元组的两个来自相同位置，而另一个来自不同的位置，优化了在天气或光照变化引起的外观变化场景下的回环检测效果。文献[36]使用 AlexNet 在 ImageNet 的物体识别数据库进行预训练，并经过 Places 数据库再训练后提升了效果。同时为了提高匹配的速度，作者利用局部敏感哈希算法进行搜索，并基于语义信息分割搜索区域来减少搜索对象。

120 部分研究把传统的特征提取方法与深度学习方法结合使用。高翔等人^[37]依旧采用了传统的 SIFT 和 ORB 等方法获取图像特征的位置，并依据特征未知裁剪出子图像块，然后采用自动编码器，把子图像块作为输入，计算出整幅图像的特征，来进行回环检测。该方法相比传统方法鲁棒性更强，但是准确性提升效果一般。文献[38]把本地增强描述向量 (VLAD) 与 CNN 特征相结合设计了 NetVLAD，该方法对 VLAD 算法进行了改进，使其计算更适合模型进行端到端学习。同时由于 NetVLAD 中加入了图像的位置信息，为了优化地点的识别性能，作者基于谷歌街景数据集，使用随机梯度下降方法进行弱监督训练，使匹配度相同的图像，GPS 越接近得到的分值越低。作者使用弱监督端到端的方法，在位置识别的排序任务上取得了最先进的效果，我们可以从中得到启发，在其他排序类任务上，我们也可以使用大量的弱标注数据集，比如使用自然语言描述的图像。

130

4 深度学习方法和传统方法的对比

135 从现有的研究成果来看,深度学习方法在视觉里程计和回环检测任务中均取得了一定的
改进成果。主要原因是深度学习方法在提取图像特征方面具有很强的优势。传统特征提取方
法依赖于人工设计的特征点,例如 SIFT、SURF、ORB 等。这些方法是计算机视觉领域的
研究者经过多年研究设计的,由于包含较大的人为因素,因此当光照强度和环物运动变
化较多时,这些人工设计的特征提取算法表现不够理想。而由以上的研究可知,卷积神经
140 网络等神经网络可以代替传统人工设计的特征点方法,通过学习获取图像特征。学习需要使
用大量的图像数据库,且耗费的时间较长。但通过使用预训练模型,多神经网络结合,调整
网络参数等方法,相比使用传统方法,可以更容易优化特征提取效果。综合多个方面,将深
度学习方法与传统方法对比如表 1 所示:

表 1 传统方法与深度学习方法对比

Tab. 1 Comparison of Traditional Methods And Deep Learning Methods

比较内容	传统方法	深度学习方法
适应性	迁移能力弱,不易在多变环境下使 用	迁移能力强,深度利用图像信息
需要的数据量	少	大量
训练时间	短时间调参即可	长时间训练
硬件要求	较低,可以达到 cpu 实时	较高,通常需要 GPU 的帮助

145 5 总结与展望

深度学习赋予了计算机“自主”从数据中学习的能力,其在对象识别、物体追踪和语义分
割方面表现出良好效果,受到广泛关注,视觉 SLAM 的研究者也将深度学习与视觉 SLAM
相结合,减少了手工设计特征在视觉里程计和回环检测中使用的局限性,并且深度学习强大
的学习能力,可以使视觉 SLAM 系统构建出更丰富的知识库,提升机器人的智能水平。但
150 是,目前的深度学习方法缺少严谨的数学理论支撑,许多参数的调整依赖研究者的经验和实
验,在特殊情况下,也可能会受到硬件能力不足的制约。

5.1 融合语义信息

SLAM 的基本任务是帮助机器人定位与构建地图,但是为了完成更高阶的任务,仅知道
位置信息是不够的,还需要在 SLAM 的各个环节融入高维的语义信息,供机器人使用。在
155 文献[39]中,作者尝试利用 SSD 目标检测方法,实现语义级别的物体检测,从而提升了 SLAM
在动态环境下的鲁棒性。同样在地图构建环节,也可以加入语义信息,使机器人具备实例级
别的识别能力。

5.2 构建机器人云平台

160 逐渐面向商用的 5G 极大的提高了移动终端的网速,使实时视频流传输变得更加稳定。
由于机器人等设备内嵌芯片运算能力有限,通过视频传输,将复杂的运算任务交由云平台处
理,可以使当前使用深度学习的视觉 SLAM 方法实时性得到保证,同时在云端可以共享海
量环境数据,进一步优化训练模型。

5.3 传感器升级与融合

从目前主要研究成果来看,视觉 SLAM 所使用的传感器大部分为单目相机、双目相机、RGBD 相机,根据环境与任务要求不同,选用的传感器也不同。而文献[40]使用了一种名为事件相机的新型视觉传感器完成了位姿估计与三维重建任务,该相机具有高时间分辨率、高动态范围、无运动模糊等优点,其主要模拟人类视觉系统,能够输出像素级亮度变化。研究新型传感器或进行多传感器融合,是优化视觉 SLAM 一个不错的思路。

[参考文献] (References)

- 170 [1] Raja Chatila, Jean-Paul Laumond. Position referencing and consistent world modeling for mobile robots[A]. Proceedings of IEEE International Conference on Robotics and Automation. 1985
- [2] T.J.Chong, X.J.Tang, C.H.Leng, M.Yogeswaran, O.E.Ng, Y.Z.Chong. Sensor Technologies and Simultaneous Localization and Mapping (SLAM)[J]. Procedia Computer Science, 2015, 76: 174-179.
- 175 [3] Fuentes-Pacheco J, Ruiz-Ascencio J, Rendon-Mancha J M. Visual simultaneous localization and mapping: a survey [J]. Artificial Intelligence Review, 2015, 43 (1): 55-81.
- [4] Davide Scaramuzza, Friedrich Fraundorfer. Visual Odometry [Tutorial][J]. IEEE Robotics & Automation Magazine, 2011, 18(4):80-92.
- [5] Klippenstein, J, Zhang H. Quantitative Evaluation of Feature Extractors for Visual SLAM[A]. Fourth Canadian Conference on Computer and Robot Vision[C]. 2007.
- 180 [6] Calonder M., Lepetit V., Strecha C., Fua P. BRIEF: Binary Robust Independent Elementary Features[A]. Lecture Notes in Computer Science[C]. Berlin, Heidelberg, 2010. 778-792.
- [7] H. Bay, T. Tuytelaars, L. Van Gool. Surf: Speeded up robust features[A]. Lecture Notes in Computer Science[C]. Berlin, Heidelberg, 2006, 404-417.
- [8] Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60: 91.
- 185 [9] Navneet Dalal, Bill Triggs. Histograms of Oriented Gradients for Human Detection[A]. International Conference on Computer Vision & Pattern Recognition (CVPR '05)[C]. San Diego, United States, 2005. 886--893
- [10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf[A]. Computer Vision (ICCV), 2011 IEEE International Conference on, 2011. 2564-2571.
- 190 [11] Klein G, Murray DW. Parallel tracking and mapping for small AR workspaces[A]. Proceedings of International Symposium on Mixed and Augmented Reality[C]. 2007, 225-234.
- [12] Mur-Artal R, Montiel JMM, Tardós JD. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Trans Robot, 2015, 31(5): 1147-1163.
- [13] Newcombe RA, Lovegrove SJ, Davison AJ. DTAM: dense tracking and mapping in real-time[A]. Proceedings of International Conference on Computer Vision[C], 2011. 2320-2327.
- 195 [14] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM[A]. Proceedings of European Conference on Computer Vision[C]. 2014. 834-849.
- [15] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry[A]. Proceedings of International Conference on Robotics and Automation[C]. 2014. 15-22.
- 200 [16] 高翔, 张涛, 颜沁睿, 刘毅. 视觉 SLAM 十四讲: 从理论到实践[M]. 北京:电子工业出版社, 2017.
- [17] Nister, D., Naroditsky, O., Bergen, J. (n.d.). Visual odometry[A]. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004[C]. Washington, DC, USA, 2004.
- 205 [18] Khalid Yousif, Alireza Bab-Hadiashar, Reza Hoseinnezhad. An Overview to Visual Odometry and Visual SLAM: Applications to Mobile Robotics[J]. Intelligent Industrial Systems, 2015, 1(4).
- [19] Konda, Kishore Reddy and Roland Memisevic. Learning Visual Odometry with a Convolutional Network[A]. VISAPP[C]. 2015.
- [20] Konda K, Memisevic R. Unsupervised learning of depth and motion[OL]. [2016-11-10]. <https://arxiv.org/pdf/1312.3429.pdf>
- 210 [21] Mohanty V, Agrawal S, Datta S, et al. DeepVO: a deep learning approach for monocular visual odometry [OL]. [2016-11-18]. <https://arxiv.org/pdf/1611.06069.pdf>
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks[A]. Advances in neural information processing systems[C]. 2012. 1097-1105.
- 215 [23] S. Wang, R. Clark, H. Wen and N. Trigoni. DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks[A]. 2017 IEEE International Conference on Robotics and Automation (ICRA)[C]. Singapore, 2017. 2043-2050.
- [24] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition[OL]. [2014-9-4]. <https://arxiv.org/pdf/1409.1556.pdf>
- 220 [25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. 2015. 1-9.
- [26] Ruihao Li, Sen Wang, Zhiqiang Long, Dongbing Gu. UnDeepVO: Monocular Visual Odometry through

- Unsupervised Deep Learning[OL]. [2017-9-20]. <https://arxiv.org/pdf/1709.06841.pdf>
- 225 [27] Clement Godard, Oisín Mac Aodha, Gabriel J. Brostow. Unsupervised Monocular Depth Estimation With Left-Right Consistency[A]. The IEEE Conference on Computer Vision and Pattern Recognition[C]. 2017. 270-279.
- [28] Tinghui Zhou, Matthew Brown, Noah Snavely, David G. Lowe. Unsupervised Learning of Depth and Ego-Motion From Video[A]. The IEEE Conference on Computer Vision and Pattern Recognition[C]. 2017. 1851-1858.
- 230 [29] Gao, X., Zhang, T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system[J]. Autonomous Robots, 2005, 41(1): 1-18.
- [30] Adrien Angeli, David Filliat, Stéphane Doncieux, Jean-Arcady Meyer. A Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words[J]. IEEE Transactions on Robotics, IEEE, 2008, 1027 -1037.
- [31] G. Csurka, C. Dance, L. Fan, J. Williamowski, C. Bray. Visual categorization with bags of keypoints[A]. ECCV04 workshop on Statistical Learning in Computer Vision[C]. 2004. 59-74.
- 235 [32] Arandjelovic, R., Zisserman, A. All about VLAD[A]. The IEEE Conference on Computer Vision and Pattern Recognition[C]. 2013. 1578-1585.
- [33] Hou, Y., Zhang, H., Zhou, S. Convolutional neural network-based image representation for visual loop closure detection[A]. 2015 IEEE International Conference on Information and Automation[C]. Lijiang: 2015. 2238-2245.
- 240 [34] Niko Sünderhauf, Sareh Shirazi, Adam Jacobson, Feras Dayoub, et al. Place recognition with ConvNet landmarks: Viewpoint-robust, condition-robust, training-free[A]. Proceedings of Robotics: Science and Systems XII, Auditorium Antonianum[C]. Rome: 2015.
- [35] Ruben Gomez-Ojeda, Manuel Lopez-Antequera, Nicolai Petkov, Javier Gonzalez-Jimenez. Training a Convolutional Neural Network for Appearance-Invariant Place Recognition[OL]. [2015-5-27]. <https://arxiv.org/pdf/1505.07428.pdf>
- 245 [36] Niko Sünderhauf, Feras Dayoub, Sareh Shirazi, Ben Ugcroft, Michael Milford. On the Performance of ConvNet Features for Place Recognition[OL]. [2015-1-17]. <https://arxiv.org/pdf/1501.04158.pdf>
- [37] X. Gao, T. Zhang. Loop closure detection for visual SLAM systems using deep neural networks[A]. 2015 34th Chinese Control Conference[C]. Hangzhou: 2015. 5851-5856.
- 250 [38] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, Josef Sivic. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition[A]. The IEEE Conference on Computer Vision and Pattern Recognition[C]. 2016. 5297-5307
- [39] Linhui Xiao, Jinge Wang, Xiaosong Qiu, Zheng Rong, Xudong Zou. Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment[J]. Robotics and Autonomous Systems, 2019, 117: 1-16.
- 255 [40] Kim, H., Leutenegger, S., Davison, A. J. Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera[A]. Lecture Notes in Computer Science[C]. 2016. 349-364