计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190900034

基于非监督深度学习的闭环检测方法

汪 丹[1] 石朝侠[1] 王燕清[2]
1 南京理工大学计算机科学与工程学院 南京 210094
2 南京晓庄学院信息工程学院 南京 210038
(117106010726@njust.edu.cn)

摘 要 闭环检测是同时定位与建图(Simultaneous localization and mapping,SLAM)的重要组成部分,能够有效减小SLAM系统中的累积误差,并且如果在定位与建图过程中跟踪丢失,还可以利用闭环检测进行重定位。与传统的手动设计的特征(hand-crafted feature)相比,从神经网络中学习到的图像特征具有更好的环境不变性和语义识别能力。考虑到基于陆标(landmark)的卷积特征能够克服整个图像特征对视点变化敏感的缺陷,文中提出了一种新的闭环检测算法。其首先通过卷积神经网络的卷积层直接识别出图像的感兴趣区域生成陆标,然后对图像中识别出的每个陆标提取卷积特征,生成图像的最终表示以检测闭环。为了验证算法的有效性,在典型的数据集上进行了对比实验,结果表明所提算法具有优异的性能,且即使是在极端的视点和外观变化的情况下仍然具有高鲁棒性。

关键词:同时定位与建图;闭环检测;人为设计特征;卷积特征;深度学习
中图法分类号 TP391

Loop Closure Detection Method Based on Unsupervised Deep Learning

WANG Dan[1],SHI Chao-xia[1] and WANG Yan-qing[2]
1 School of Computer Science and Engineering,Nanjing University of Science and Technology,Nanjing 210094,China
2 School of Information Engineering,Nanjing Xiaozhuang University,Nanjing 210038,China

Abstract Loop closure detection is one of the most critical parts for simultaneous localization and mapping (SLAM) systems. It can reduce the accumulative error in SLAM system. If the tracking is lost during localization and mapping,it can also use the loop closure detection for relocation. Image features learned from neural networks have better environmental invariance and semantic recognition capabilities compared to traditional hand-crafted features. Considering that the landmark-based convolution features can overcome the defect that the whole image features are sensitive to viewpoint changes,this paper proposes a new loop closure detection algorithm. Firstly,it directly identifies the saliency region of the image through the convolutional layer of the convolutional neural network to generate a landmark. And then,it extracts the ConvNet features from the landmarks to generate the final image representations. In order to verify the effectiveness of the algorithm,some comparative experiments were performed on some typical datasets. The rusults show that the proposed algorithm has superior performance,and has highly robust even in drastic viewpoints and appearance changes.

Keywords SLAM,Loop closure detection,Hand-crafted feature,ConvNet feature,Deep learning

1 引言

SLAM系统旨在在一个未知的环境中实时重建环境的三维结构,同时对机器人自身进行定位[1]。在一个SLAM系统中,前端用于估算相邻图像间相机的运动以及构造局部地图;后端的任务是对这些数据进行优化。如果仅考虑相邻帧之间的关联,误差将会随着时间的推移而累积,最终得到具有累积误差的不可靠的SLAM系统,我们也就无法构建全局一致的轨迹和地图[2]。

闭环检测对SLAM系统的意义非常重大,其本质是检测相机当前经过的位置是否是之前已经到过的地方[3],从而构建起间隔更久远的帧之间的约束关系。它能消除系统的累积误差,保证估计的轨迹和地图在长时间下的正确性[4]。此外,在跟踪丢失以后,由于闭环检测提供了当前数据与所有历史数据的关联,因此可以利用闭环检测做重定位[5]。因此,一个好的闭环检测算法能很大程度地提升SLAM系统的精度和鲁棒性。

然而,闭环检测算法一直存在两个巨大的挑战:1)由于天

到稿日期:2019-09-04 返修日期:2020-01-09 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金项目(61371040)
This work was supported by the National Natural Science Foundation of China (61371040).
通信作者:石朝侠(scx@njust.edu.cn)

气、遮挡和动态物体造成的外观变化；2）由于相机拍摄位置等造成的视点变化。虽然现有的一些算法[6-8]已经表现出了很好的性能，但是这些方法存在不能兼顾这两个挑战、计算量比较大或者只适用于特定的场景等缺点。

传统的方法是直接利用图像的原始像素[9]或者从图像中提取的人为设计的特征[6]来生成图像表示，然后根据图像像素直接进行匹配或者通过词袋模型加速图像描述符匹配。通常有两类手动设计特征：1）局部的人为设计特征，如SIFT[10]，SURF[11]和ORB[8]；2）全局的人为设计特征，如GIST[12]和HOG[13]。基于局部特征的描述符对于视点变化是鲁棒的，但是不适合处理外观变化。而全局特征描述符在环境变化中性能表现良好，但是当环境中存在视点变化和遮挡时，其表现不佳。因此，在照明变化、视点变化、遮挡和其他因素的组合变化的情况下，局部特征和全局特征都不能提供令人满意的性能。

深度学习在机器人和计算机视觉领域的成功应用表明，在闭环检测中利用基于卷积特征的方法比利用手动设计的特征具有更明显的优势，尤其是在光照变化的环境中。与手动设计的局部特征相比，卷积特征具有更好的环境不变性；与手动设计的全局特征相比，卷积特征具有更好的语义识别能力。

文献[14]利用Overfeat网络[15]提取卷积特征作为图像全局描述符，但描述符太大，无法实时检测闭环。文献[16]提出了使用先进的深度学习技术来提取鲁棒特征，用于替换SeqSLAM中原来的特征。但是在这些方法中，卷积特征都是提取自通用神经网络，而非针对闭环检测的专用网络。为此，文献[17]设计了具有针对性的卷积神经网络，文献[18]进一步在足够大的且变化的数据集中训练这种专有网络，其中的训练数据集拍摄于数千个不同的场景，拥有大量的外观变化。然而，这些网络架构依赖于监督学习，需要带标签的图像作为训练数据集。最近，文献[19]设计了一种专门用于闭环检测的无监督深度神经网络架构，其关键思想是从网络中提取的卷积特征比上述所有卷积特征更轻巧、紧凑。但卷积特征仍然无法很好地解决视点不变性，其原因与描述整个图像的全局特征相似。文献[20-21]提出，基于陆标的卷积特征生成的图像表示能够显著提高闭环检测在视点变化情况下的鲁棒性。然而这些方法都要求特殊的陆标检测器（landmark detectors）来识别闭环检测任务中的感兴趣区域（Region-of-interest，ROI）。文献[22]证明了卷积神经网络（Convolutional Neural Networks，CNNs）（如vgg16网络[23]、ResNet[24]、GoogleNet[25]、AlexNet[26]等）的最后几层卷积层通常嵌入了非常丰富的语义信息，这些信息对应于一些对闭环检测任务有意义的图像区域。基于这项理论，本文提出一种全新的陆标生成机制，即直接根据卷积层的激活值识别图像中的ROI，而无需任何陆标检测器。然后对每一个生成的陆标，利用文献[19]设计的深度神经网络提取其卷积特征。因此，相比于现有的大多数闭环检测算法[5-6,9,19,21]，本文提出的基于陆标-卷积特征的闭环检测算法能方便、直接地生成陆标，并且采用了专为闭环检测设计的网络结构来提取图像的卷积特征。

2　鲁棒的闭环检测系统

基于场景识别的闭环检测可以描述为：给定一个查询帧I_q和一组具有F幅图像的数据库图像$I_d=\{I_d^F, f=1,2,\cdots,F\}$，闭环检测的目的是在数据库图像中找到与$I_q$拍摄于同一场景下的参考帧$I_d^q$。

本文提出的鲁棒的闭环检测系统（系统的组成结构如图1所示）主要由3个部分组成.

1）输入层：首先将查询帧和数据库图输入到预先训练的卷积神经网络中，直接从卷积神经网络的卷积层识别出感兴趣区域，用矩形框的形式可视化在图示中。

2）陆标检测层：对每一幅查询帧和数据库图像，用识别出的感兴趣区域生成M（M＝100）个陆标，其中，两幅图像帧中相互匹配的陆标用同一种颜色标记。

3）卷积特征提取层：利用一个无监督的深度神经网络对从图像中生成的每一个陆标提取卷积特征描述子，得到对应的特征向量。

4）陆标匹配层：通过计算查询帧陆标向量和每一幅数据库图像的陆标向量之间的余弦距离，来交叉匹配两帧的陆标区域，保留相互匹配的陆标。

5）场景检索：根据匹配的陆标对，计算查询帧和每一幅数据库图像之间的整体相似度，以确定每个查询帧的最佳匹配参考帧。

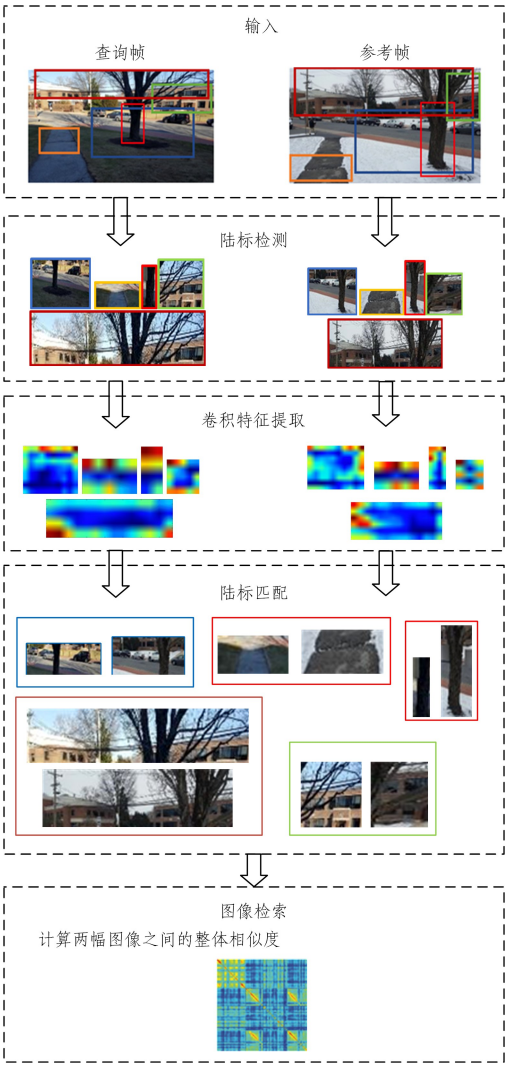

图1　系统组成结构图

Fig. 1　Structure diagram of system

3 识别图像的感兴趣区域作为陆标

对于每一帧图像，首先利用 AlexNet[26] 卷积神经网络的特定卷积层获得每一个通道对应的特征图，那么总的特征图数量即为该层的通道数；然后对于所有的特征图，将所有非零的激活值和其周围 8 个相邻的激活值分别聚为一类，记为 $C_i(i\in\{1,2,\cdots,N\})$，$N(N\geqslant 100)$ 表示一幅图像中的聚类数目。每个聚类 C_i 的能量值 E_i 的计算式为：

$$E_i=\frac{1}{|C_i|}\sum_j a_i^j,\ \forall a_i^j\in C_i,j\in\{1,2,\cdots,9\} \tag{1}$$

其中，$|c_i|$ 表示第 i 个聚类的大小，a_i^j 表示 C_i 中的第 j 个激活值。在得到 N 个聚类的能量值之后，取其中 M 个具有最大能量值的聚类映射回原图像中，作为最终生成的陆标，记为 $L_s,s\in[1,2,\cdots,M]$。图 2 给出了用本文方法在图像上识别感兴趣区域生成陆标的例子，其中用矩形框的形式将生成的陆标可视化在图像中。

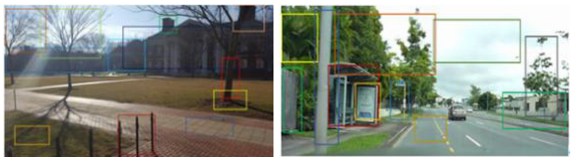

图 2 识别感兴趣区域并生成陆标

Fig. 2 ROI identifing to generate landmarks

4 卷积特征提取

对于检测到的每个陆标，利用文献[19]中构建的无监督卷积自动编码器网络来提取卷积特征描述符，该网络结构是专门为闭环检测任务设计的。图 3 给出了该网络的框架，其中 conv，relu，pool，norm，FC 分别表示卷积层、激活层、池化层、归一化层和全连接层。将 M 个路标作为输入，X 表示 HOG 特征的维度，$\hat{X}$ 表示重建的特征描述子的维度。在该自编码模型中，对 3 个卷积层使用线性整流函数(ReLU)激活，而对全连接层使用 sigmoid 激活，以便网络能更好地重建 HOG 特征。当训练结束时，网络即具备学习重建 HOG 特征的能力，该方法仅保留 3 个卷积层及对应的池化层，丢弃除此之外所有的网络层来提取图像的卷积特征。此外，由于对相同大小的输入提取的 HOG 特征的维度是相同的，因此可以利用欧几里得距离作为 HOG 描述子的距离度量，在损失层利用 l_2 损失函数来比较 X 和其重建的 $\hat{X}$ 的大小，即 $\|X-\hat{X}\|_2^2$。网络的参数设置如表 1 所列。

表 1 参数设置

Table 1 Parameter settings

layer	Kernel size	Output number	pad	stride
conv1	5	64	4	2
pool1	3	64	0	2
conv2	4	128	2	1
pool2	3	128	0	2
conv3	3	4	0	1

该网络已经被证明是快速可靠的，无须降低提取到的卷积特征的维数即可实现实时检测闭环。文献[19]中的一系列实验表明，用此网络学习到的 HOG 特征检测闭环的能力要明显优于原始的 HOG 特征，并且能够取代基于卷积特征的闭环检测系统中的通用神经网络。此外，由于该网络不需要特定环境的训练，因此可以直接应用其预训练模型来提取本文实验中所用的数据集图像的特征。记图像 I 生成的任一陆标提取的特征向量为 $L_{C_i}^I$，特征维度为 1 064，则对于任意一幅图像，总的特征维度为 $M\times 1064$。

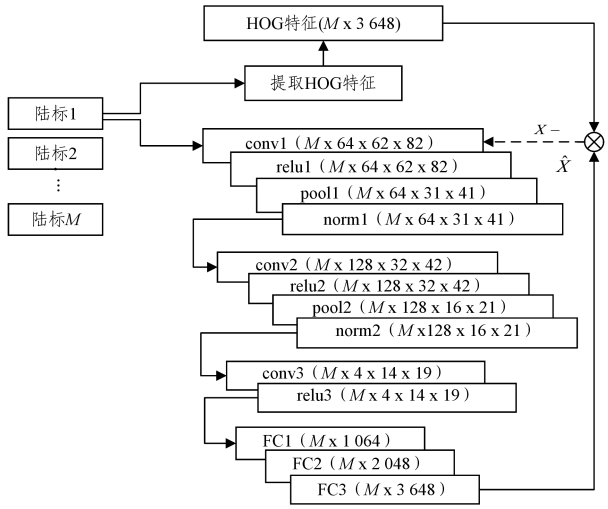

图 3 无监督卷积自编码网络的框架

Fig. 3 Framework of unsupervised convolutional autoencoder network

5 相似性计算

为了计算 I_q 和 I_d^q 之间的相似性得分，交叉匹配从这两幅图像中提取的所有陆标。利用余弦距离度量 I_q 的一个陆标 $u(u\in i)$ 与 I_d^q 的一个陆标 $v(v\in i)$ 之间的相似性。

$$d_{u,v}=\frac{\boldsymbol{L}_u^{I_q^T}\boldsymbol{L}_v^{I_d^q}}{\|\boldsymbol{L}_u^{I_q}\|\ \|\boldsymbol{L}_v^{I_d^q}\|} \tag{2}$$

其中，$d_{u,v}$ 即为 u 和 v 的余弦距离。其中 $\boldsymbol{L}_u^{I_q}$ 和 $\boldsymbol{L}_v^{I_d^q}$ 分别表示对 I_q 中的陆标 u 和 I_d^q 中的陆标 v 提取的卷积特征向量，$\|\cdot\|$ 表示向量的长度。

使用简单的线性搜索来确定 I_q 和 I_d^q 所有陆标之间的匹配。

对于每一个匹配的陆标对 (u,v)，根据它们的区域大小确定其权重，权重记为 $W_{u,v}$：

$$W_{u,v}=\exp\left(\frac{|h_u-h_v|}{\max(h_u,h_v)}+\frac{|w_u-w_v|}{\max(w_u,w)}\right) \tag{3}$$

其中，h_u,h_v,w_u,w_v 分别为 (u,v) 区域的高和宽，$|h_u-h_v|$ 和 $|w_u-w_v|$ 分别表示两个区域的高的差值的绝对值和宽的差值的绝对值。

最终，I_q 和 I_d^q 的整体相似性得分 S_{I_q,I_d^q} 为：

$$S_{I_q,I_d^q}=\frac{1}{M}\sum_{u,v}W_{u,v}d_{u,v} \tag{4}$$

对于每一帧查询图像 I_q，遍历并计算其与数据库中所有图像 I_d^f 的相似性得分，其中得分最高的图像即为 I_q 的最佳匹配：

$$Z=\arg\max_{I_q},S_{I_q,I_d^f},f\in[1,2,\cdots,F] \tag{5}$$

其中，Z 表示与 I_q 具有最高相似性得分的参考帧。

6　实验结果

在闭环检测算法中，通常利用准确率-召回率（Precision-Recall，P-R）曲线作为评估算法好坏的标准。准确率指真实闭环占所有被检测到的回环的概率；召回率指所有真实闭环被检测出来的概率。在 SLAM 中，通常对准确率的要求更为严格，因为如果准确率较低，将导致算法误检测为闭环，使得优化算法给出完全错误的结果，从而导致建出的地图失效。而如果召回率较低，意味着将会有部分闭环未被检测到，从而使得构建的地图受到累计误差的影响，但只需要一两次闭环就可以完全消除由此造成的误差。本文利用 R-R 曲线的曲线下面积（AUC），在 100％的精确率下的最大召回率以及当有较高的召回率时的精确率作为实验的评估标准。

为了证明本文算法优越的性能，将其与以下几个具有代表性的算法进行对比。

1）SeqSlam[6]：该算法是一种基于序列的开源方法，经证明是一种成功的闭环检测算法。其对由光照、天气和一天中的时间变化引起的显著的外观变化具有高鲁棒性。

2）DBoW[8]：该算法将特征描述符离散化以构造词汇树，然后使用构造的词汇树快速找到几何验证阶段的点的对应关系。其作为闭环检测算法被应用于 ORB-SLAM 中并表现出了良好的性能。

3）AlexNet[26]：文献[27]证明了 AlexNet 的第三层卷积层对场景的识别是最鲁棒的，但该层产生的特征向量为 64 896 维，如此大的维度将导致无法实现实时的闭环检测。因此本文先利用高斯随机投影（GRP）[28-29]将该层产生的特征向量维度降低到 1 024 维。

4）文献[19]提出的算法是最新的闭环检测的开源方案，该算法设计了一个专门用于闭环检测任务的深度神经网络，并用 Places dataset[30]数据集进行训练。由于已经证明了应用此网络不需要用与环境有关的训练集来训练，因此本文直接使用其预训练模型在测试集上进行实验。把用该算法生成的 P-R 曲线记为 calc。

实验将利用几个具有挑战性的公共数据集，从外观不变性和视点不变性两个方面评估闭环检测系统的鲁棒性。

6.1　视点不变性评估-The Gardens Point Dataset

The Gardens Point Dataset[1)] 包括 3 个遍历轨迹。其中的一个轨迹记录拍摄于夜间；其他的两个轨迹拍摄于白天，分别沿着人行道的左侧和右侧拍摄，表现出在路径左侧和右侧行走时出现的视点变化以及主要由行人等动态物体引起的轻微外观变化。因此在实验中，选择白天的这两个轨迹序列作为评估视点不变性的测试集。

图 4 展示了本文方法及对比算法在该数据集上生成的 P-R 曲线。可以看出，seqslam，DBoW，AlexNet 和 calc 这 4 个算法表现相当，calc 和 AlexNet 算法有相同的 AUC，当精确率为 100％时，calc 和 seqslam 算法具有次小的召回率，而本文算法能得到最大的召回率；此外，本文提出的算法还具有最大的 AUC。由此可见，当环境中存在视点变化时，本文提出的算法能正确匹配更多的陆标，明显优于其他 4 个算法。

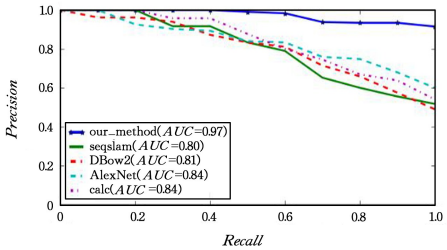

图 4　在 Gardens Point Dataset 上的准确率-召回率曲线

Fig. 4　Precision-recall curves on Gardens Point Dataset

6.2　视点和外观不变性评估-The Campus Loop Dataset

The Campus Loop Dataset[19] 是最具挑战性的公共场景识别数据集之一。该数据集由两个序列组成，每个序列包含 100 帧的图像，包含室内和室外环境。第一个序列是在下雪天拍摄的，地面在室外环境中被雪覆盖；第二个序列是在晴天拍摄的。比较这两个序列的图像可以看到，该数据集提供了由天气变化造成的极端的天气变化以及视点变化。

对比实验如图 5 所示。当环境中既存在视点变化也存在外观变化时，其他 4 个算法的性能下降很明显，而本文算法仍然具有最大的召回率和最大的 AUC。因此，即使环境中存在视点变化和外观变化，本文算法仍然具有很高的鲁棒性。

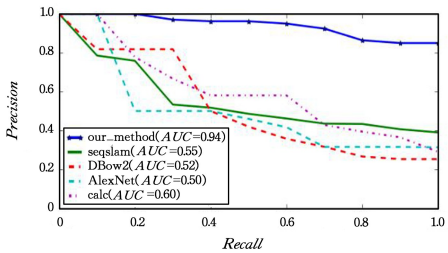

图 5　在 Campus Loop Dataset 上的准确率-召回率曲线

Fig. 5　Precision-recall curves on Campus Loop Dataset

结束语　本文提出一种对外观和视点变化均鲁棒的闭环检测算法，该算法不同于其他基于陆标的相关算法，不要求任何额外的陆标检测器，而是直接由卷积神经网络的深层卷积层识别显著性区域，生成陆标。此外，该算法利用专为闭环检测设计的神经网络提取图像特征，而非从通用神经网络中提取特征，进一步提升了算法的性能。实验结果表明，无论是环境中存在剧烈的视点变化还是极端的外观变化，所提算法均具有很高的鲁棒性。虽然作为单独的闭环检测系统其性能不错，但是改进闭环检测是为了提升 SLAM 的鲁棒性，将闭环检测算法与 SLAM 系统相结合会涉及关键帧的选择、闭环验证机制等问题，因此接下来的工作是将本文提出的闭环检测算法与主流 SLAM 系统相结合，用于改进 SLAM 系统。

参 考 文 献

[1]　VAKHITOV A，LEMPITSKY V. Learnable Line Segment Descriptor for Visual SLAM[J]. IEEE Access，2019（7）：39923-39934.

1) http://tinyurl.com/gardenspointdataset

[2] GAO X, ZHANG T, LIU Y, et al. Visual SLAM XIV: From Theory to Practice[M]. Beijing: Electronic Industry Press, 2017.

[3] LOWRY, STEPHANIE M, et al. Visual Place Recognition: A Survey[J]. IEEE Transactions on Robotics, 2016, 32 (1): 1-19.

[4] ROTTMANN N, BRUDER R, SCHWEIKARD A, et al. Loop Closure Detection in Closed Environments[J]. arXiv: 1908. 04558.

[5] CHEN J, LI J, XUY, et al. A compact loop closure detection based on spatialpartitioning[C]// International Conference on Image. IEEE, 2017.

[6] MILFORD M J, WYETH G F. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights [C]// 2012 IEEE International Conference on Robotics and Automation. IEEE, 2012.

[7] BAMPIS L, AMANATIADIS A, GASTERATOS A. Fast loop-closure detection using visual-word-vectors from image sequences[J]. The International Journal of Robotics Research, 2017, 37(1): 62-82.

[8] MUR-ARTALR, TARDOSJ D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.

[9] CUMMINS M, NEWMANP. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance[J]. The International Journal of Robotics Research, 2008, 27(6): 647-665.

[10] LOWED G. Object Recognition from Local Scale-Invariant Features[C]// Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. IEEE, 1999.

[11] XU A, NAMITG. SURF: Speeded-Up Robust Features[J]. Computer Vision & Image Understanding, 2008, 110(3): 404-417.

[12] OLIVA, AUDE, TORRALBA, et al. Building the gist of a scene: the role of global image features in recognition[J]. Progress in brain research, 2006, 155: 23-36.

[13] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005.

[14] CHEN Z, LAM O, JACOBSON A, et al. Convolutional Neural Network-based Place Recognition[C]// Proceedings of the 16th Australasian Conference on Robotics and Automation. 2014.

[15] SERMANET P, EIGEN D, ZHANG X. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. arXiv: 1312. 6229.

[16] BAI D, WANG C, ZHANG B, et al. Sequence Searching with CNN Features for Robust and Fast Visual Place Recognition [J]. Computers & Graphics, 2018, 70: 270-280.

[17] GOMEZ-OJEDA R, LOPEZ-ANTEQUERA M, PETKOV N, et al. Training a Convolutional Neural Network for Appearance-Invariant Place Recognition[J]. arXiv: 1505. 07428.

[18] CHEN Z, JACOBSON A, SUNDERHUF N, et al. Deep learning features at scale for visual place recognition[C]// IEEE International Conference on Robotics and Automation (ICRA), 2017: 3223-3230.

[19] MERRILL N, HUANGG. Lightweight Unsupervised Deep Loop Closure[J]. arXiv: 1805. 07703v01.

[20] SÜNDERHAUF, NIKO, SHIRAZI S, Jacobson A, et al. Place recognition with ConvNet landmarks: Viewpoint-robust, condition-robust, training-free[M]// Proceedings of the 2010 Academy of Marketing Science (AMS) Annual Conference. Springer International Publishing, 2015.

[21] HOU Y, ZHANG H, ZHOU S. BoCNF: efficient image matching with Bag of ConvNet features for scalable and robust visual place recognition[J]. Autonomous Robots, 2017, 42(9): 1-17.

[22] MATTHEW D Z, FERGUS R. Visualizing and understanding convolutional networks[J]. arXiv: 1311. 2901.

[23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.

[24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[25] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[26] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.

[27] SÜNDERHAUF N, DAYOUB F, SHIRAZI S, et al. On the Performance of ConvNet Features for Place Recognition[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2015: 4297-4304.

[28] DASGUPTA S. Experiments with Random Projection[C]// Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. 2013: 143-151.

[29] BRIGHAM E, MANNILA H. Random projection in dimensionality reduction: Applications to image and text data[C]// Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2001: 245-250.

[30] ZHOU B, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 million Image Database for Scene Recognition[J]. IEEE Transactions on Pattern Analysis and machine Intelligence, 2017, pp(99): 1-1.

WANG Dan, born in 1995, postgraduate. Her main research interests include visual SLAM and so on.

SHI Chao-xia, born in 1972, Ph. D, professor, postgraduate supervisor. His main research interests include unmanned vehicle autonomous navigation, SLAM and multi-robot collaboration.