DOI：10.13973/j.cnki.robot.2017.0889

基于深度学习的视觉 SLAM 综述

赵 洋，刘国良，田国会，罗 勇，王梓任，张 威，李军伟
（山东大学控制科学与工程学院，山东 济南 250061）

摘 要：综述了深度学习技术应用到同步定位与地图创建（SLAM）领域的最新研究进展，重点介绍和总结了深度学习与帧间估计、闭环检测和语义 SLAM 结合的突出研究成果，并对传统 SLAM 算法与基于深度学习的 SLAM 算法做了深入的对比研究．最后，展望了未来基于深度学习的 SLAM 研究发展方向．

关键词：深度学习；视觉 SLAM；帧间估计；视觉里程计；闭环检测；语义 SLAM
中图分类号：TP24 **文献标识码**：A **文章编号**：1002-0446(2017)-06-0889-08

A Survey of Visual SLAM Based on Deep Learning

ZHAO Yang，LIU Guoliang，TIAN Guohui，LUO Yong，WANG Ziren，ZHANG Wei，LI Junwei
(*School of Control Science and Engineering*, *Shandong University*, *Ji'nan* 250061, *China*)

Abstract: Latest research progresses of deep learning techniques applied to SLAM (simultaneous localization and mapping) are summarized. In addition, the prominent achievements on inter-frame motion estimation, loop closure detection and semantic SLAM incorporated with deep learning are introduced. Furthermore, the deep learning based SLAM is compared with the traditional ones in detail. Finally, the future research directions of advanced SLAM based on deep learning are discussed.

Keywords: deep learning; visual SLAM (simultaneous localization and mapping); inter-frame motion estimation; visual odometry; loop closure detection; semantic SLAM

1 引言（Introduction）

同时定位与地图构建（SLAM）是机器人搭载视觉、激光、里程计等传感器，对未知环境构建地图的同时实现自定位的过程，在机器人自主导航任务中起着关键作用[1-4]．当前 SLAM 问题的研究手段主要是通过在机器人本体上安装多类型传感器来估计机器人本体运动信息和未知环境的特征信息，利用信息融合实现对机器人位姿的精确估计以及场景的空间建模．尽管 SLAM 采用的传感器有激光和视觉等多种类型，但其处理过程一般包含 2 个部分[5]（如图 1 所示）：前端帧间估计和后端优化．前端帧间估计解决的是机器人在获取前后 2 帧传感器信息的时间间隔内的运动估计，而后端优化解决的是机器人检测到路径闭环后对历史轨迹的优化问题[6]．考虑到里程计等传感器信息积累的误差，后端优化就变得尤为重要，其关键是正确的检测闭环．相对于激光传感器单一的空间结构感知信息，视觉传感器凭借其丰富的色彩和纹理等感知信息在提高帧间估计精度和闭环检测正确率方面有着巨大的优势和潜力[7,8]．

视觉 SLAM（visual SLAM）是以图像作为主要环境感知信息源的 SLAM 系统，可应用于无人驾驶、增强现实等应用领域，是近年来的热门研究方向[9-12]．典型视觉 SLAM 算法以估计摄像机位姿为主要目标，通过多视几何理论来重构 3D 地图．为提高数据处理速度，部分视觉 SLAM 算法首先提取稀疏的图像特征，通过特征点之间的匹配实现帧间估计和闭环检测，如基于 SIFT（scale-invariant feature transform）特征的视觉 SLAM[13] 和基于 ORB（oriented FAST and rotated BRIEF）特征的视觉 SLAM[14]．SIFT 和 ORB 特征凭借其较好的鲁棒性和较优的区分能力以及快速的处理速度，在视觉 SLAM 领域受到广泛应用．但是，人工设计的稀疏图像特征当前有很多局限性，一方面如何设计

基金项目：国家自然科学基金（61603213）；中国博士后科学基金（2016M590635）；山东省优秀中青年科学家科研奖励基金（BS2014DX010）；山东大学人才引进与培养类专项基金（2015TB009）；山东省泰山学者工程．
通信作者：刘国良，liuguoliang@sdu.edu.cn 收稿/录用/修回：2017-02-15/2017-05-12/2017-06-15

稀疏图像特征最优地表示图像信息依然是计算机视觉领域未解决的重要问题[1]，另一方面稀疏图像特征在应对光照变化、动态目标运动、摄像机参数改变以及缺少纹理或纹理单一的环境等方面依然有较多挑战[1]．面对这些问题，在视觉 SLAM 领域近年出现了以深度学习技术为代表的层次化图像特征提取方法，并成功应用于 SLAM 帧间估计[15-18]和闭环检测[19-22]．深度学习算法是当前计算机视觉领域主流的识别算法，其依赖多层神经网络学习图像的层次化特征表示，与传统识别方法相比，可以实现更高的识别准确率[23-26]．同时，深度学习还可以将图像与语义进行关联，与 SLAM 技术结合生成环境的语义地图[27-28]，构建环境的语义知识库[29]，供机器人进行认知与任务推理，提高机器人服务能力和人机交互的智能性．

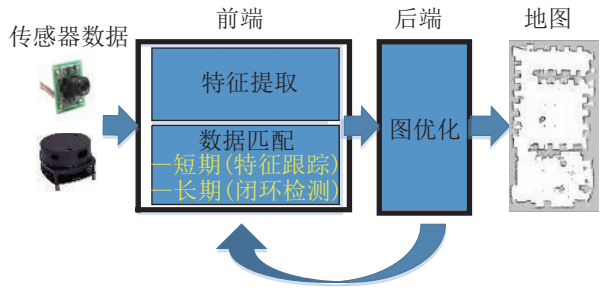

图 1 典型 SLAM 系统的前后端

Fig.1 Front-end and back-end in a typical SLAM system

通过前述分析，深度学习与 SLAM 的结合主要体现在 3 个方面，即基于深度学习的帧间估计、闭环检测和语义地图生成．因此，本文首先综述了深度学习与帧间估计的结合，主要涉及基于光流图像的深度学习帧间估计．其次，综述了深度学习与闭环检测的结合，主要涉及深度学习特征提取以及位置识别．最后，综述了深度学习与语义地图生成的结合，主要涉及利用深度学习对静态场景和动态场景进行语义分割．基于这三个方面的详尽调研，本文将深度学习 SLAM 算法和传统 SLAM 算法进行了深入对比，分析了当前算法尚存的问题与不足，并指出了未来的发展趋势和方向．

2 深度学习与帧间估计（Deep learning and inter-frame motion estimation）

帧间估计也称为视觉里程计（visual odometry），是通过分析关联摄像机图像之间的多视几何关系确定机器人位姿与朝向的过程，可作为视觉 SLAM 的前端[30]．相较于传统的基于稀疏特征或稠密特征的帧间估计方法，基于深度学习的方法无需特征提取，也无需特征匹配和复杂几何运算，使得基于深度学习的方法更加直观简洁．

Konda 和 Memisevic[15]提出基于端到端的深度神经网络架构用于预测摄像机速度和方向的改变．该方法的主要特点是利用单一类型的计算模块和学习规则提取视觉运动和深度信息以及里程计信息，主要分为 2 个步骤：首先是图像序列深度和运动信息的提取．作者利用乘性交互（multiplicative interaction）神经网络进行时序立体图像的同步检测（synchrony detection），将立体图像序列之间的空间变换估计转换为同步检测，该网络也被称为无监督同步/深度自动编码器（synchrony/depth autoencoder，SAE-D）[31]．其次是图像序列速度和方向改变估计．作者将上一层 SAE-D 提取的运动和深度信息作为卷积神经网络层（CNN）输入，用以学习图像速度和方向改变，从而执行帧间估计．实验表明，该学习算法能实现对连续帧的帧间估计，并且执行速度较快，在基于 3.20 GHz CPU，24 GB RAM 和 GTX 680 GPU 配置的机器上，平均执行速度为 0.026 秒/帧．但是在精度方面，该算法还无法达到主流的视觉里程计精度．此外，该方法采用无监督自动编码机作为 CNN 第 1 层降低了训练难度，且一定程度上缓解了网络对训练数据的过拟合．

Costante[18]等利用卷积神经网络学习图像数据的最优特征表示进行视觉里程计估计，并展示了其算法在应对图像运动模糊、光照变化方面的鲁棒性．该方法先用 Brox 算法提取连续 2 帧的稠密光流特征，以此作为 CNN 网络的输入．文中在设计深度网络时探索了 3 种不同的 CNN 架构，一是基于全局特征的 CNN-1b，一是基于局部特征的 CNN-4b，以及结合前两种架构的 P-CNN．CNN-1b 和 CNN-4b 结构相似，将浅层 CNN 和深层 CNN 并行级联入全连接网络．该方法在训练的过程中采用逐层训练的方法来解决 CNN 全局训练难的问题．为同时考虑全局特征、局部特征、浅层特征和深层特征，作者将全局特征 CNN-1b 和局部特征 CNN-4b 结合构建了 P-CNN．通过与现有的视觉里程计方法在公开数据集上比较，发现基于 P-CNN 的帧间估计算法从鲁棒性、精确性和执行速度等各方面都具有优势．但是，实验结果也说明了所提算法对训练数据的依赖，特别是当图像序列帧间速度过快时，算法误差较大，其原因是训练集缺乏高速训练样本造成估计的旋转误差较大．在机器配置为 i7-4720HQ 2.60 GHz 处理器、NVIDIA Tesla K40 GPU 的情况下，这种学习方法平均用时为 50 ms/帧，其中计算

光流用了 30 ms，执行 CNN 预测用了 20 ms.

Handa[17] 等在空间变换网络（spatial transform network）[32] 基础上进行了扩展，在设计网络时选择对经典计算机视觉方法进行回归，如端到端的视觉里程计和图像深度估计等. 作者利用神经网络构建了包含全局变换、像素变换和 M 估计器在内的 gvnn（geometric vision with neural network）软件库. 作为应用示例，作者实现了基于 RGB-D 数据的视觉里程计. 该系统的网络构架由 VGG-16 网络启发构建的 Siamse 网络层、位姿变换估计层（SE3 layer）、3 维网格生成层（3D grid generator）、投影层（projection layer）和双线性插值层（bilinear interpolation）组成. 其中，Siamse 网络的输入为 2 个连续的帧图像，输出是对摄像机 6 自由度的帧间位姿估计向量. 基于此帧间估计，作者结合深度信息，将上一帧图像投射到当前位姿，并经过双线性插值生成预测图像. 为构造损失函数进行学习，预测图像不是与当前图像进行像素级对比，而是与上一帧图像利用真实的帧间估计进行投影和双线性差值后的图像对比，从而避免了传统神经网络结构在学习过程中单方面的像素丢失和各种运动模糊、强度变化或图像噪声对匹配的影响. 并且这种方法确保在收敛时，损失函数尽可能接近 0，能够恰当地处理丢失像素. 此类回归网络在设计时各层模型意义明确，训练时也可以采用分层训练的方法.

由上可见，采用端到端的深度神经网络架构可以快速提取图像序列的帧间运动信息. 相比传统帧间估计算法，基于学习的方法替代繁琐公式计算，无需人工特征提取和匹配，显得简洁直观，并且在线运算速度较快. 然而，通过分析发现不同学习算法之间的神经网络架构设计差异性较大，对训练学习数据库有较强的依赖. 同时，由于基于深度学习神经网络的帧间估计还处于研究起步阶段，算法之间的对比性理论分析和实验还有待继续开展.

3　深度学习与闭环检测（Deep learning and loop closure detection）

闭环（loop closure）检测是指机器人在地图构建过程中，通过视觉等传感器信息检测是否发生了轨迹闭环，即判断自身是否进入历史同一地点. 闭环检测发生时可触发 SLAM 后端全局一致性算法进行地图优化，消除累积轨迹误差和地图误差. 闭环检测问题本质上是场景识别问题[33]，传统方法通过人工设计的稀疏特征或像素级别稠密特征进行匹配[34]，而深度学习则可以通过神经网络学习图像中的深层次特征，其识别率可以达到更高水平[23]. 因此，基于深度学习的场景识别可以提高闭环检测准确率[33].

基于深度学习的闭环检测技术早期主要利用预训练的 CNN 网络架构进行图像特征提取，不同层级的网络层在图像特征描述方面具有一定的差异性. Chen[33] 等首次提出了基于 CNN 模型的位置识别技术，其核心在于通过 CNN 学习图像特征表示，在其所测数据集上以 100% 的准确率提升了 75% 的召回率. 作者选择 ImageNet 大赛中用以物体识别的 OverFeat 神经网络模型进行图像描述. 为比较神经网络每层图像特征在场景识别上的性能差别，作者进一步利用各层特征构造混合矩阵：

$$M_k(i,j)=d(L_k(I_i),L_k(I_j)),\ i=1,\cdots,R,\ j=1,\cdots,T \tag{1}$$

其中 I_i 代表第 i 帧图像输入，$L_k(I_i)$ 代表与 I_i 对应的第 k 层输出，$M_k(i,j)$ 代表第 k 层训练样本 i 和测试样本 j 之间的欧氏距离，即描述了两者之间的匹配程度. 对于混合矩阵中可能的位置匹配假设，作者进一步构造空间连续性滤波器和时间连续性滤波器进行综合验证，提高匹配准确率. 同时，作者对各层网络所训练出来的特征性能进行探索，发现网络中间层特征描述对于视角相似的图像匹配效果较好，而中后层对于场景视角变化具有更强的适应性和鲁棒性.

Hou[35] 等利用 caffe 框架下的 AlexNet 模型进行特征提取，通过实验对比，发现在光照变化明显的环境下，采用深度学习的特征描述鲁棒性能优于传统特征，且特征提取更加迅速. Sünderhauf[36] 等提出利用 CNN 模型提取图像区域特征描述子，实验证明局部区域描述比全局图像描述更能有效应对图像的视角改变问题.

某一类数据预训练好的神经网络结构在应用方面具有一定局限性，如 AlexNet 预训练模型主要针对物体分类数据库 ImageNet，因此学者 Gomez-Ojeda[37] 等首次基于位置识别数据库 Places 对神经网络进行再训练，从而提高图像检索准确率. Sünderhauf[38] 等在 caffe 框架下用 ImageNet 数据库预训练好的 AlexNet 模型进行特征提取，发现经过 Places 数据库再训练后的网络在闭环检测方面更具优势. 同时为了解决匹配的实时性问题，作者提出了 2 种改进思路：一种是利用局部敏感哈希（local-sensitive hashing）搜索算法，另一种是根据语义信息对搜索区域进行分割以减少搜索对象. Bai[19] 等

在训练好的 Places CNN 网络基础上，采用局部敏感哈希算法进行图像压缩提高匹配效率．Shahid[20]等利用大型公共数据集 Nordland 对 Places-AlexNet 预训练网络进行了再训练和参数优化，并比较几种特征向量距离测度：二元欧氏距离（pairwise Euclidean）、二元余弦距离（pairwise cosine）、三元欧氏距离（triplet Euclidean）、三元余弦距离（triplet cosine）．通过实验发现余弦距离比欧氏距离训练效果更优，拥有更好的场景辨识能力．

Gao[21-22]等通过自动编码器提取图像特征来进行图像匹配．作者首先利用传统 SIFT、FAST 或 ORB 等算法提取图像特征位置，并围绕特征位置裁剪图像为不同区域子图像块．自动编码器以向量化后的区域子图像块为输入、以训练后的隐含层输出图像特征．最后构建相似性矩阵，判断是否发生闭环．在自动编码机训练部分，作者针对损失函数进行了改进．

Arandjelovic[39]等提出了一种端对端的场景识别算法．考虑到局部特征聚合描述子（vector of locally aggregated descriptors，VLAD）在场景识别中的良好效果，作者基于 CNN 和 VLAD 构造了 NetVLAD，并改进了原始 VLAD 图像表示函数，使得其可微分．实验证明基于 NetVLAD 的再训练算法能大大提高图像的匹配精度．

相较于传统闭环检测（位置识别）算法，基于深度学习的方法利用深度神经网络提取图像特征，表达图像信息更充分，对光照、季节等环境变化有更强鲁棒性．但是，如何选择合适的隐含层表示图像特征、如何设计神经网络架构和如何利用面向任务的大数据集对网络参数迁移学习优化等问题依然是未来研究的重要问题．

4 深度学习与语义 SLAM（Deep learning and semantic SLAM）

语义 SLAM 是指 SLAM 系统在建图过程中不仅获得环境中的几何结构信息，同时可以识别环境中独立个体，获取其位置、姿态和功能属性等语义信息，以应对复杂场景及完成更加智能的服务任务[27,40-41]．语义 SLAM 的优势在于[27]：(1) 传统 SLAM 方法以静态环境假设为前提，而语义 SLAM 可以预知物体（人、汽车等）的可移动属性．(2) 语义 SLAM 中的相似物体知识表示可以共享，通过维护共享知识库提高 SLAM 系统的可扩展性和存储效率．(3) 语义 SLAM 可实现智能路径规划，如机器人可以搬动路径中的可移动物体等实现路径更优．语义 SLAM 的关键在于对环境中物体目标的精准识别，而近年兴起的深度学习技术恰好是当前最具潜力和优势的物体识别方法，因此深度学习和语义 SLAM 的结合受到领域内研究者的广泛关注．

Sünderhauf[42]等提出面向物体对象的语义建图方法（如图 2 所示）：首先利用 ORB-SLAM2 算法[43]估计 RGB-D 摄像头位姿和构建环境的稀疏特征地图，并将深度图像对应的点云依据摄像头当前位姿投射到全局坐标，从而得到环境的 3D 点云地图．其次是物体检测与识别，采用 Liu 等提出的基于卷积神经网络的单次拍摄多边界框检测（SSD）方法[44]，对关键帧图像生成固定数量的物体建议边界框，并计算每个建议边界框的置信值．然后是基于超体元的 3 维目标物体点云分割，以进一步分割出前述基于图像划分得到的物体所对应点云．最后是基于最近邻方法的物体数据关联，以确定当前物体和地图中物体之间的对应性，进而添加或更新地图中目标物体的点云信息和从属类别置信值等数据．因此，利用 CNN 网络中的 SSD 方法，该语义地图最终包含有：(1) 关键帧的点云数据；(2) 地图中各物体的 3D 点云分割及其所对应关键帧关系；(3) 语义信息．

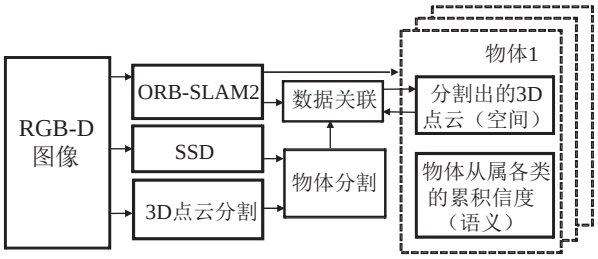

图 2 基于 SSD 方法的语义 SLAM 系统概览[42]

Fig.2 Overview of semantic SLAM system based on SSD methods

McCormac[28]等提出基于卷积神经网络的稠密 3 维语义地图构建方法 SemanticFusion，其依赖 Elastic Fusion SLAM 算法提供室内 RGB-D 视频帧间位姿估计，利用卷积神经网络预测像素级的物体类别标签，最后结合贝叶斯升级策略和条件随机场模型实现不同视角下 CNN 预测值的概率升级，最终生成包含语义信息的稠密 3 维语义地图．SemanticFusion 中的 CNN 在 caffe 框架下，选择在 Noh 等[45]提出的基于 VGG-16 网络的反卷积语义分割网络结构基础上加入深度通道，从而能够在输入四通道 RGB-D 图像后，输出稠密像素级语义概率图．

Li[46]等提出了基于 CNN 和 LSD-SLAM（large

scale direct SLAM）的单目半稠密 3 维语义建图构建方法，其过程与 McCormac 所提方法类似，但利用 LSD-SLAM 代替 Elastic Fusion 方法估计摄像机位姿，利用单目相机而非 RGB-D 深度相机和立体相机．其过程分为 3 部分：首先选取关键帧估计摄像机位姿，提高处理速度；其次利用 DeepLab-v2 中的 CNN 架构进行像素级分类，包含亚卷积和亚空间金字塔池化（ASPP）两个核心组件，以扩大滤波范围，融合多尺度特征信息；最后利用贝叶斯升级像素分类概率预测，结合条件随机场（CRF）进行地图正则化，对生成的语义分割地图进行噪声平滑．

地图的语义生成与 SLAM 过程是可以相互促进的两部分．一方面精确的地图构建有利于目标模型的学习和分类，另一方面目标的精确识别和分类有利于地图的精确构建，如精准的闭环检测等，因此两者是相辅相成的．语义信息生成的挑战在于精确的物体目标级别或像素级别的分类．相比较于传统的词袋模型等手工定义特征的物体分类方法，深度学习通过大数据学习可以使得分类过程效率更高．但是，当前基于深度学习的语义 SLAM 多是单向的，即利用传统 SLAM 改进语义分割结果，还未出现语义信息与 SLAM 相互促进的完善机制．语义 SLAM 是机器人完成高层次智能任务的前提，而深度学习的出现会对未来的语义 SLAM 起到积极的促进作用．

5　深度学习方法与传统方法对比（Comparison between the deep learning and the traditional methods）

基于图像信息特征表示的传统 SLAM 设计首先需要解决的是显著特征选取的问题[1]．鲁棒的特征应在不同视角（相似变换下）、光照强度变化和背景变化等情况下具有描述不变性．特征提取又可分为特征检测与特征描述 2 个部分．所提取的特征描述子是否具有良好不变性直接影响 SLAM 系统帧间估计精度、闭环检测检索效率与准确率和语义知识库的功能性．基于方向直方图的局部特征，如 HOG（histogram of oriented gradient）、SIFT 和 SURF（speeded up robust feature）长期占据了传统 SLAM 算法，此类特征需要具有领域专业知识的专家精心设计，又被称为特征工程（feature engineering），因此具有较大的人为性．同时手工设计特征在面临光照强度变化和物体运动等情况时性能下降，特别是在物体识别领域性能表现不理想[23,47]，成为 SLAM 系统性能提升的瓶颈．另一方面，传统 SLAM 算法特征提取与分类器设计分离，导致 SIFT 等特征的匹配准确率不高[23]．传统 SLAM 在构建语义地图时，不仅需要构建物体特征描述数据库，并且需要训练决策森林等分类器进行物体分类[27]．

近年来，SLAM 算法尝试利用深度学习技术从海量预训练图像集和实时感知图像集中直接学习高层次特征．深度学习的特征蕴含在每一层深度神经网络神经元中，因此也被称作特征学习（feature learning）．特征学习具有以下特点：(1) 需要大规模的数据库，导致训练时间长；(2) 一旦完成神经网络训练，即可同步完成匹配分类器的设计，同时端到端效率将胜于传统 SLAM 系统方案；(3) 采用 dropout 等神经元抑制算法稀疏化网络参数防止过拟合，以利用大规模的数据，具有更强的泛化能力；(4) 迁移学习使得在新应用场景的感知数据中学习特征更加便捷，易于机器人共享；(5) 缺少直观理解意义；(6) 具有乘法交互部分的深度学习网络能够对 2 帧图像之间的相似性进行计算，从而实现更加简洁的帧间估计[17]；(7) 基于深度学习技术的语义 SLAM 算法在机器人环境中识别物体的准确率更高，可有效提高语义知识库质量．

综上所述，深度学习方法在 SLAM 领域中有较大潜力，与传统方法对比展示了多个方面的优点，如表 1 所示．

表 1　传统 SLAM 算法与基于深度学习的 SLAM 算法对比

Tab.1　Comparison between the traditional SLAM algorithms and the ones based on deep learning

比较项目	传统 SLAM 算法	基于深度学习的 SLAM 算法
模型参数调整难易程度	+ 小规模数据，调参周期短	− 大规模数据，训练周期长
模型物理含义	+ 直观意义明确	− 缺少直观意义
模型泛化能力	− 信息利用不充分，参数少，泛化能力弱	+ 信息利用充分，参数多，泛化能力强
适应能力	− 迁移能力弱	+ 迁移能力强
设计流程	− 特征设计与分类器训练分离	+ 同步完成特征设计与分类器训练

6　未来展望（Future prospects）

6.1　高维传感器数据处理与融合

深度学习技术作为端到端的特征学习方法，为传感器大数据特征提取与处理提供了新思路．激光测距扫描仪是传统 SLAM 常用的传感器，具有精度高、扫描角度广和不受时间限制等优点．然而，基

于 3 维激光扫描仪的机器人位姿估计需要匹配处理海量数据，使得传统的 ICP 算法（iterative closest point algorithm）等迭代匹配算法耗时较多．针对这一问题，当前较普遍的解决方案是对大量点云数据进行筛选和提取特征[48]．Nicolai 等[49] 使用 3D 激光测距仪 VLP-16 采集 3D 点云数据，并投影到 2 维平面生成深度图像，利用 CNN 网络训练，得到端到端的帧间估计结果，其运行速度明显快于传统 ICP 匹配方法．3 维到 2 维的投影虽然降低了数据维度，但也造成信息损失，因此可考虑设计 CNN 3 维卷积和池化网络架构，从原始 3D 点云中直接提取空间特征．进一步的，在 3 维点云基础上融合 RGB 信息可同时获取环境的纹理色彩和结构特征，可提高特征对环境的信息表达力．

6.2 机器人知识库

语义信息使得机器人可获取环境中物体的属性信息，也使得高层次复杂任务的推理和执行成为可能．对语义信息进一步利用本体网络语言（OWL）加工处理，可形成标准化的机器人知识，为高层推理和任务决策奠定基础．知识的逐步积累可生成机器人知识库，为多机器人知识共享、升级和扩展提供服务[50-51]．机器人知识库框架一般有 2 个主要部分：知识描述和知识关联．知识描述（knowledge description）包括从低层传感数据和高层语义信息提取的机器人知识，如感知特征、物体类别、几何地图、语义地图、动作单元和任务等；知识关联（knowledge association）则运用单向或者双向规则的逻辑推断知识间的相关性，如逻辑推理和贝叶斯推断．在构建知识库框架的整个过程中涉及到 2 方面的转换，一方面是将机器人获取到的零散的、孤立的、异构的数据，通过本体模型转换成同构的、关联的、机器人可利用的本体数据；另一方面是将转换后的本体数据（ontology，共享的概念模型的形式化的规范说明），通过语义转换或规则转换，转换成具有实际意义的同构的、具有联系的本体语义数据．同时通过 OWL 描述语言建立起室内环境中本体语义库信息，通过相应的转换规则绑定本体数据和语义库之间的联系（语义转换），为机器人的智能决策提供思维支撑（意图转换），让机器人能够真正做到智能化决策．

6.3 云机器人

随着云技术的发展，以云平台作为服务机器人的知识库存储和分享平台成为一个热门研究方向．欧盟的 RoboEarth[52] 等项目的实施，更是在该方面实现了零的突破．另一方面，云平台提供的巨大计算能力，也为以深度学习为代表的大数据服务提供了保障[53-54]．云平台与机器人的结合产生了云机器人这一新概念，并在业界产生了广泛影响[55]．云机器人不仅可以将复杂的计算任务转移到云端，还可以处理海量数据，并分享信息和技能．单个机器人学习能力有限，学习时间长，然而通过云端的技能分享，可立刻获取其他机器人学习的技能知识，减少了学习时间，提高了机器人的服务能力．

6.4 SLAM 促进深度学习

在深度学习给人工智能领域引入新的思路、促进 SLAM 系统发展的同时，SLAM 反过来也能促进深度学习的研究进程[56]，如 Agrawal 等[57] 在特征学习过程中引入 SLAM 优化估计后的里程计信息，使深度神经网络学习到的特征效果更加显著．另外，学者们使用大尺度 SLAM 系统构建的地图和帧间位姿关系，客观上为深度学习提供了大规模图像到图像“关联”的数据集，可实现网络参数的进一步学习和调整，以适应于特定应用场景．

7 结论（Conclusion）

深度学习通过模拟人的大脑结构构造复杂的神经网络模型，利用大量数据进行训练来模拟人的学习过程，在语义分割、物体识别和动作识别等领域获得了极大成功，逐渐引起领域内研究者的广泛关注．深度学习与 SLAM 的结合在一定程度上改善了视觉里程计和场景识别等由于手工设计特征而带来的应用局限性，同时对高层语义快速准确生成以及机器人知识库构建也产生了重要影响，从而潜在提高了机器人的学习能力和智能化水平．

然而，通过深度学习技术学习的信息特征还缺少直观的意义以及清晰的理论指导，其中的训练参数很大程度依赖专家的调参经验，其训练效果较依赖数据库质量以及和当前应用场景的相似度．因此，一方面，未来 SLAM 系统性能还将长期受益于深度学习理论的深入发展；另一方面，目前深度学习多应用于 SLAM 局部的子模块，如定位模块或闭环检测模块，而如何将深度学习架构的应用贯穿于整个 SLAM 系统仍是一个巨大挑战．

参考文献（References）

[1] Fuentes-Pacheco J, Ruiz-Ascencio J, Rendón-Mancha J M , et al. Visual simultaneous localization and mapping: A survey[J]. Artificial Intelligence Review, 2015, 43(1): 55-81.

[2] 徐德．室内移动式服务机器人的感知、定位与控制 [M]. 北京：科学出版社，2008.
Xu D. Perception, localization and control of indoor mobile service robot[M]. Beijing: Science Press, 2008.

[3] Cadena C, Carlone L, Carrillo H, et al. Past, present, and future of simultaneous localization and mapping: Towards the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.

[4] 顾照鹏，刘宏．单目视觉同步定位与地图创建方法综述 [J]．智能系统学报，2015，10(4)：499-507.
Gu Z P, Liu H. A survey of monocular simultaneous localization and mapping[J]. CAAI Transactions on Intelligent Systems, 2015, 10(4): 499-507.

[5] 梁明杰，闵华清，罗荣华．基于图优化的同时定位与地图创建综述 [J]．机器人，2013，35(4)：500-512.
Liang M J, Min H Q, Luo R. Graph-based SLAM: A survey[J]. Robot, 2013, 35(4): 500-512.

[6] Kummerle R, Grisetti G, Strasdat H, et al. g^2o: A general framework for graph optimization[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2011: 3607-3613.

[7] 徐德，谭民，李原．机器人视觉测量与控制 [M]．北京：国防工业出版社，2011.
Xu D, Tan M, Li Y. Visual measurement and control for robots [M]. Beijing: National Defense Industry Press, 2011.

[8] 杨东方，王仕成，刘华平，等．基于 Kinect 系统的场景建模与机器人自主导航 [J]．机器人，2012，34(5)：581-589.
Yang D F, Wang S C, Liu H P, et al. Scene modeling and autonomous navigation for robots based on Kinect system[J]. Robot, 2012, 34(5): 581-589.

[9] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2012: 3354-3361.

[10] Klein G, Murray D. Parallel tracking and mapping on a camera phone[C]//IEEE International Symposium on Mixed and Augmented Reality. Piscataway, USA: IEEE, 2009: 83-86.

[11] 陈殿生，刘静华，殷兰兰．服务机器人辅助老年人生活的新模式与必要性 [J]．机器人技术与应用，2010，17(2)：2-4.
Chen D S, Liu J H, Yin L L. New style and necessity of service robot assisting the elderly[J]. Robot Technique and Application, 2010, 17(2): 2-4.

[12] 张建伟，张立新，胡颖，等．开源机器人操作系统——ROS [M]．北京：科学出版社，2012.
Zhang J W, Zhang L Y, Hu Y, et al. ROS: Open source robot operate system[M]. Beijing: Sicence Press, 2012.

[13] Davison A J. Real-time simultaneous localisation and mapping with a single camera[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2003: 1403.

[14] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: A versatile and accurate monocular SLAM System[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.

[15] Konda K, Memisevic R. Learning visual odometry with a convolutional network[C]//Proceedings of the 10th International Conference on Computer Vision Theory and Applications. Lisbon, Portugal: SCITCC Press, 2015: 486-490.

[16] Dosovitskiy A, Fischery P, Ilg E, et al. FlowNet: learning optical flow with convolutional networks[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2015: 2758-2766.

[17] Handa A, Bloesch M, Pătrăucean V, et al. gvnn: Neural network library for geometric computer vision[M]//Lecture Notes in Computer Science, vol. 9915. Berlin, Germany: Springer-Verlag, 2016: 67-82.

[18] Costante G, Mancini M, Valigi P, et al. Exploring representation learning with CNNs for frame-to-frame ego-motion estimation[J]. IEEE Robotics and Automation Letters, 2016, 1(1): 18-25.

[19] Bai D D, Wang C Q, Zhang B. Matching-range-constrained real-time loop closure detection with CNNs features[J]. Robotics and Biomimetics, 2016, 3(1): 70-75.

[20] Shahid M, Naseer T, Burgard W. DTLC: Deeply trained loop closure detections for lifelong visual SLAM[C/OL]// Robotics: Science and Systems. (2016-06-18) [2016-11-10]. https://roboticvision.atlassian.net/wiki/download/attachments/41320632/Shahid%20-%20DTLC.pdf?version=1&modificationDate=1466185006962&cacheVersion=1&api =v2.

[21] Gao X, Zhang T. Loop closure detection for visual slam systems using deep neural networks[C]//34th Chinese Control Conference. Piscataway, USA: IEEE, 2015: 5851-5856.

[22] Gao X, Zhang T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system[J]. Autonomous Robots, 2015, 41(1): 1-18.

[23] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2014: 580-587.

[24] Fischer P, Dosovitskiy A, Brox T. Descriptor matching with convolutional neural networks: A comparison to SIFT[EB/OL]. (2015-06-24) [2016-11-10]. https://arxiv.org/pdf/1405.5769.pdf.

[25] 伍锡如，黄国明，孙立宁．基于深度学习的工业分拣机器人快速视觉识别与定位算法 [J]．机器人，2016，38(6)：711-719.
Wu X R, Huang G M, Sun L N. Fast visual identification and location algorithm for industrial sorting robots based on deep learning[J]. Robot, 2016, 38(6): 711-719.

[26] 牛杰，卜雄洙，钱堃，等．一种融合全局及显著性区域特征的室内场景识别方法 [J]．机器人，2015，37(1)：122-128.
Niu J, Bu X Z, Qian K, et al. An indoor scene recognition method combining global and saliency region features[J]. Robot, 2015, 37(1): 122-128.

[27] Salas-Moreno, Renato F. Dense semantic SLAM[D]. London, UK: Imperial College, 2014.

[28] McCormac J, Handa A, Davison A, et al. SemanticFusion: Dense 3D semantic mapping with convolutional neural networks[EB/OL]. (2016-08-28) [2016-11-10]. https://arxiv.org/pdf/1609.05130.pdf.

[29] Tenorth M, Kunze L, Jain D, et al. Knowrob-map-knowledge-linked semantic object maps[C]//IEEE/RAS International Conference on Humanoid Robots. Piscataway, USA: IEEE, 2010: 430-435.

[30] Kitt B, Geiger A, Lategahn H. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme[C]//Intelligent Vehicles Symposium. Piscataway, USA: IEEE, 2010: 486-492.

[31] Konda K, Memisevic R. Unsupervised learning of depth

and motion[EB/OL]. (2013-12-16) [2016-11-10]. https://arxiv.org/pdf/1312.3429.pdf.

[32] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[C]//Advances in Neural Information Processing Systems. San Francisco, USA: Morgan Kaufmann, 2015: 2017-2025.

[33] Chen Z T, Lam O, Jacobson A, et al. Convolutional neural network-based place recognition[EB/OL]. (2014-09-06) [2016-11-10]. https://arxiv.org/pdf/ 1411.1509.pdf.

[34] Cummins M, Newman P. FAB-MAP: Probabilistic localization and mapping in the space of appearance[J]. International Journal of Robotics Research, 2008, 27(6): 647-665.

[35] Hou Y, Zhang H, Zhou S L. Convolutional neural network-based image representation for visual loop closure detection[C]//IEEE International Conference on Information and Automation. Piscataway, USA: IEEE, 2015: 2238-2245.

[36] Sünderhauf N, Shirazi S, Jacobson A, et al. Place recognition with ConvNet landmarks: Viewpoint-robust, condition-robust, training-free[C/OL]//Robotics: Science and Systems. (2016-06-27) [2016-11-10]. http://www.roboticsproceedings. org/rss11/p22.pdf.

[37] Gomez-Ojeda R, Lopez-Antequera M, Petkov N, et al. Training a convolutional neural network for appearance-invariant place recognition[EB/OL]. (2015-3-27) [2016-11-10]. https://arxiv.org/pdf/1505.07428.pdf.

[38] Sunderhauf N, Shirazi S, Dayoub F, et al. On the performance of ConvNet features for place recognition[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2015: 4297-4304.

[39] Arandjelovic R, Gronat P, Torii A, et al. NetVLAD: CNN architecture for weakly supervised place recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016: 5297-5307.

[40] 于金山，吴皓，田国会，等. 基于云的语义库设计及机器人语义地图构建 [J]. 机器人，2016，38(4)：410-419.
Yu J S, Wu H, Tian G H, et al. Semantic database design and semantic map construction of robots based on the cloud[J]. Robot, 2016, 38(4): 410-419.

[41] Tenorth M, Kunze L, Jain D, et al. Knowrob-map-knowledge-linked semantic object maps[C]//IEEE/RAS International Conference on Humanoid Robots. Piscataway, USA: IEEE, 2010: 430-435.

[42] Sünderhauf N, Pham T, Latif Y, et al. Meaningful maps – Object-oriented semantic mapping[EB/OL]. (2016-9-26) [2016-11-10]. https://arxiv.org/pdf/ 1609.07849.pdf.

[43] Mur-Artal R, Tardos J D. Probabilistic semi-dense mapping from highly accurate feature-based monocular SLAM[C]//Robotics: Science and Systems. Cambridge, USA: MIT Press, 2015.

[44] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[M]//Lecture Notes in Computer Science, vol.9905. Berlin, Germany: Springer-Verlag, 2016: 21-37.

[45] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//IEEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2015: 1520-1528.

[46] Li X, Belaroussi R. Semi-dense 3D semantic mapping from monocular SLAM[EB/OL]. (2016-11-13) [2016-12-10]. https://arxiv.org/padf/1611.04144. pdf.

[47] Masci J, Meier U, An D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[M]//Lecture Notes in Computer Science, vol.6791. Berlin, Germany: Springer-Verlag, 2011: 52-59.

[48] Tong C H, Anderson S, Dong H, et al. Pose interpolation for laser-based visual odometry[J]. Journal of Field Robotics, 2014, 31(5): 787-813.

[49] Nicolai A, Skeele R, Eriksen C, et al. Deep learning for laser based odometry estimation[EB/OL]. (2016-6-17) [2016-11-10]. http://juxi.net/workshop/deep-learning-rss-2016/papers/Nicolai%20-%20Deep%20Learning%20Lidar%20 Odometry.pdf.

[50] Lim G H, Suh I H, Suh H. Ontology-based unified robot knowledge for service robots in indoor environments[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 2011, 41(3): 492-509.

[51] Li C C, Tian G H, Chen H Z. The introduction of ontology model based on SSO design pattern to the intelligent space for home service robots[C]//IEEE International Conference on Robotics and Biomimetics. Piscataway, USA: IEEE, 2016.

[52] Waibel M, Beetz M, Civera J, et al. RoboEarth[J]. IEEE Robotics & Automation Magazine, 2011, 18(2): 69-82.

[53] Kehoe B, Patil S, Abbeel P, et al. A survey of research on cloud robotics and automation[J]. IEEE Transactions on Automation Science and Engineering, 2015, 12(2): 1-12.

[54] Tian G H, Chen H Z, Lu F. Cloud computing platform based on intelligent space for service robot[C]//IEEE International Conference on Information and Automation. Piscataway, USA: IEEE, 2015: 1562-1566.

[55] 田国会，许亚雄. 云机器人：概念、架构与关键技术研究综述 [J]. 山东大学学报：工学版，2014，44(6)：47-54.
Tian G H, Xu Y X. Cloud robotics: Concept, architectures and key technologies[J]. Journal of Shandong University: Engineering Science, 2014, 44(6): 47-54.

[56] 林辉灿，吕强，张洋，等. 稀疏和稠密的 VSLAM 的研究进展 [J]. 机器人，2016，38(5)：621-631.
Lin H C, Lü Q, Zhang Y, et al. The sparse and dense VSLAM: A survey[J]. Robot, 2016, 38(5): 621-631.

[57] Agrawal P, Carreira J, Malik J. Learning to see by moving[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2015: 37-45.

作者简介：

赵 洋（1994 –），男，硕士生. 研究领域：SLAM，机器人智能感知与自主导航，机器学习.

刘国良（1983 –），男，博士，副研究员. 研究领域：机器人智能感知与自主导航，机器学习，信息融合.

田国会（1969 –），男，博士，教授. 研究领域：服务机器人，智能空间，云机器人系统.